



BIENÁLE ČESKÉ LINGVISTIKY 2024  
BIENNIAL OF CZECH LINGUISTICS 2024

Praha 17.–20. 9. 2024  
<https://bcl2024.ff.cuni.cz>

# Sborník abstraktů Book of abstracts

Univerzita Karlova – Charles University  
Filozofická fakulta – Faculty of Arts  
nám. Jana Palacha 1/2  
116 38 Praha 1  
Česká republika – Czech Republic

První ročník Bienále české lingvistiky poskytne platformu pro setkání (nejen) české lingvistické obce v celé její šíři – napříč lingvistickými subdisciplínami, filologiemi i teoretickými a metodologickými paradigmaty. Jádrem Bienále tvoří odborné tematické panely organizované lingvisty z českých akademických a vědeckých pracovišť.

Bienále proběhne díky finanční a materiální podpoře [Filozofické fakulty Univerzity Karlovy](#) a [Programu Cooperatio](#) (oblast Lingvistika). Bienále se koná pod záštitou [Jazykovědného sdružení České republiky](#), rektorky Univerzity Karlovy [prof. MUDr. Mileny Králíčkové, Ph.D.](#), a děkanky Filozofické fakulty Univerzity Karlovy [Mgr. Evy Lehečkové, Ph.D.](#)

---

The first edition of the Biennial of Czech Linguistics provides a platform for the meeting of the Czech linguistic community (and beyond) in its whole breadth – across linguistic subdisciplines, philologies, as well as theoretical and methodological paradigms. The core of the Biennial is constituted by thematic workshops organized by linguists affiliated with Czech academic or scientific institutions.

The Biennial will take place thanks to the financial and material support from [Faculty of Arts, Charles University](#) and the [Cooperatio program](#) (area Linguistics). The Biennial is held under the auspices of [Jazykovědné sdružení České republiky](#), the Rector of Charles University [prof. MUDr. Milena Králíčková, Ph.D.](#), and the Dean of the Faculty of Arts, Charles University, [Mgr. Eva Lehečková, Ph.D.](#)

© 2024 Autoři abstraktů / Authors of abstracts  
Redaktoři / Editors: Alexandr Rosen & Radek Šimík

**17. 9. 2024**

**Doprovodné akce**

***Satellite events***

- Diskusní panel o postojích českých lingvistek a lingvistů k jazykové regulaci  
*Discussion panel on the attitudes of Czech linguists towards language regulation*
- Vyhledávání v paralelním korpusu InterCorp za použití anotace Universal Dependencies (workshop)  
*Searching the InterCorp parallel corpus using the Universal Dependencies annotation (workshop)*
- Multilingvismus v českém prostředí (popularizační panel)  
*Multilingualism in the Czech context (popularization panel)*

**18.–20. 9. 2024**

**Hlavní program**

***Main program***

Tematické panely, plenární přednášky, posterová sekce

*Thematic workshops, plenary talks, a poster session*

# Obsah / Contents

## Plenární přednášky / Plenary talks

1. Klitika: jejich fonologie, morfologie a syntax / Clitics: Phonology, morphology, syntax (CZE, ENG)
2. Production and comprehension in adults from a psycholinguistic perspective (ENG)
3. Metody korpusové analýzy diskurzu v češtině / Methods of Corpus-based Discourse Studies in the Czech Context (CZE, ENG)
4. Onomastika v době digital humanities (CZE)
5. Sémantika, pragmatika a logika / Semantics, pragmatics, and logic (CZE, ENG)
6. Conversion/Zero-derivation: Aspects of the theory (ENG)
7. Aktuální výzvy (nejen) české dialektologie (CZE)
8. Empirický výzkum jazyka v sociální interakci / Empirical research on language in social interaction (CZE, ENG)
9. Frazeologie češtiny: strukturní, kognitivní, kulturní a didaktické aspekty (CZE)
10. Manifestations of sound symbolism in language: phonaesthemes, ideophones and onomatopoeia (ENG)
11. Nové empiricky podložené pohledy na mnohojazyčnost / New Empirically Based Perspectives on Multilingualism (CZE, ENG)
12. Lingvistické aspekty překladu: nové impulsy i návrat ke kořenům (CZE)
13. Jazykový management ve vícejazyčných podnicích a jejich okolí (CZE)
14. Aktuální otázky české lexikologie a lexikografie (CZE)
15. Data-based research in word formation (ENG)
16. Polar (yes/no) questions: form, meaning and more (ENG)
17. Mezi jazyky: metody kontrastivního výzkumu s využitím korpusů / Between Languages: Methods of Contrastive Research Using Corpora (CZE, ENG)
18. Lingvistická anotace korpusů (CZE)
19. Speech and language development in children acquiring Czech (ENG)
20. L2 learning – a path to bilingualism / Osvojování dalšího jazyka – cesty k bilingvistice (CZE, ENG)
21. Metafora semper viva (CZE, ENG)
22. Poster session (CZE, ENG)

# Plenární přednášky

## *Plenary talks*

Bienále české lingvistiky 2024  
Biennial of Czech Linguistics 2024

# Obsah / Table of Contents

*Czech morphology as a testing ground for morphological theory*

Olivier Bonami (Université Paris Cité, Laboratoire de linguistique formelle, CNRS)

*From philology through computational linguistics to phonological typology and back:*

*Development of syllabic consonants in Czech*

Pavel Kosek (Masaryk University, Brno)

*Multilingualism from a psycholinguistics point of view*

Barbara Mertins (TU Dortmund)

## Czech morphology as a testing ground for morphological theory

Olivier Bonami (Université Paris Cité, Laboratoire de linguistique formelle, CNRS)

Recent research in theoretical morphology frequently uses large scale quantitative studies for empirical validation (see, among many others, Baayen et al. 2019; Bonami & Paperno, 2018; Bonami & Strnadová, 2019; Cotterell *et al.*, 2019; Malouf, 2017; Marelli & Baroni, 2015). While other methods, from basic philological work to behavioral experiments and computational simulations, are of course highly valuable, the distinct value of this line of work is twofold. First, computational methods applied to large datasets allow one to directly address how the overall morphological system is organized, rather than focusing on convenience samples of data whose representativeness is always disputable. Second, implemented methods provide a useful check on observer bias: while they are by no means bias-free, computational operationalizations have a systematicity that human observers hardly attain, making it easier for sources of bias to be identified and addressed.

A distinct characteristic of this line of research is its dependence on the size and quality of datasets. The statistical properties of word distributions (Baayen 2001, Blevins *et al.* 2017) are such that small samples are underinformative when working on morphology and other types of lexical issues. In addition, while in the age of the internet, vast datasets are readily available for many languages, the quality of these datasets, from lexical databases extracted from crowdsourced dictionaries (e.g. Batsuren 2022) to web corpora (e.g. Baroni *et al.* 2009; Schäfer and Bildhauer 2012), is relatively poor, leading to potential errors or noise in research that build on them (see e.g. Malouf *et al.* 2020).

In this context, Czech stands out as a compelling target language for empirical validation of theoretical research in morphology, thanks to a suite of linguistic resources whose scope and quality is unparalleled for any other language: the SYN collection of corpora, totaling more than 5 billion words of edited text, carefully tagged and lemmatized (Hnátková *et al.*, 2014; Jelínek *et al.*, 2021); the MorFlexCZ inflectional lexicon (Hajič *et al.*, 2020); and the DeriNet derivational database (Vidra *et al.*, 2019). In parallel, the Czech morphological system has a typologically average level of complexity, making it a good fit for the study of many morphological phenomena, some of which are poorly exemplified by e.g. Germanic or Romance languages.

In my talk I will illustrate this point by giving a broad overview of recent research in inflectional morphology based on the Czech declension system, addressing *inter alia* the following research questions: the structure of inflection class systems (Beniamine & Bonami, 2023), the place of overabundance (multiple forms with the same function) in the description of inflection (Guzmán Naranjo & Bonami, 2021), and the role of features in structuring inflectional paradigms (Bonami, Kyjánek & Wauquier, 2023).

## Bibliography

- Baayen, Harald R. 2001. *Word frequency distributions*. Dordrecht: Springer.
- Baayen, R. Harald, Yu-Ying Chuang, Elnaz Shafaei-Bajestan, and James P. Blevins. 2019. The Discriminative Lexicon: A Unified Computational Model for the Lexicon and Lexical Processing in Comprehension and Production Grounded Not in (De)Composition but in Linear Discriminative Learning. *Complexity*: 4895891.
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation* 43: 209–226.
- Batsuren, Khuyagbaatar, Omer Goldman, Salam Khalifa, Nizar Habash, Witold Kieraś, Gábor Bella, Brian Leonard, Garrett Nicolai, Kyle Gorman, Yustinus Ghanggo Ate, Maria Ryskina, Sabrina Mielke, Elena Budianskaya, Charbel El-Khaissi, Tiago Pimentel, Michael Gasser, William Abbott Lane, Mohit Raj, Matt Coler, et al. 2022. Unimorph 4.0: Universal Morphology. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 840–855
- Beniamine, Sacha, and Olivier Bonami. 2023. *Inflection class systems*. In Peter Ackema, Sabrina Bendjaballah, Eulàlia Bonet and Antonio Fábregas (eds.), *The Wiley-Blackwell Companion to Morphology*.
- Blevins, James P., Petar Milin and Michael Ramscar. 2017. The Zipfian Paradigm Cell Filling Problem. In Ferenc Kiefer, James P. Blevins, and Huba Bartos (eds.), *Morphological paradigms and functions*. Leiden: Brill.
- Bonami, Olivier, Lukáš Kyjánek, and Marine Wauquier. 2023. Assessing the Featural Organisation of Paradigms with Distributional Methods. *Proceedings of the 6<sup>th</sup> meeting of the Society for Computation in Linguistics*.
- Bonami, Olivier, and Denis Paperno. 2018. Inflection vs. derivation in a distributional vector space. *Lingue e linguaggio* 17: 173–195.
- Bonami, Olivier, and Jana Strnadová. 2019. Paradigm structure and predictability in derivational morphology. *Morphology* 29: 167–197.
- Cotterell, Ryan, Christo Kirov, Mans Hulden, and Jason Eisner. 2019. On the Complexity and Typology of Inflectional Morphological Systems. *Transactions of the Association for Computational Linguistics* 7: 327–342.
- Guzmán Naranjo, Matías, and Olivier Bonami. 2021. Overabundance and inflectional classification: Quantitative evidence from Czech. *Glossa* 6.Hnátková, Milena, Michal Křen, Pavel Procházka, and Hana Skoumalová. 2014. The SYN-series corpora of written Czech. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 160–164.
- Hajič, Jan, Jaroslava Hlaváčová, Marie Mikulová, Milan Straka, Barbora Štěpánková. 2020. *MorfFlex CZ 2.0*. Data/software, LINDAT-CLARIAH, URL: <http://hdl.handle.net/11234/1-3186>, 2020.
- Hnátková, Milena, Michal Křen, M., Pavel Procházka, and Hana Skoumalová. 2014. The SYN-series corpora of written Czech. *Proceedings of the Ninth International Conference on Language Resources and Evaluation*, pp. 160–164.
- Jelínek, Tomáš, Jan Křivan, Vladimír Petkevič, Hana Skoumalová, and Jana Šindlerová, J. (2021): SYN2020: A new corpus of Czech with an innovated annotation. In: Kamil Ekštejn, František Pártl, Miloslav Konopík (eds.) *Text, Speech, and Dialogue. TSD 2021*, pp. 48–59. Cham: Springer.
- Malouf, Robert. 2017. Abstractive morphological learning with a recurrent neural network. *Morphology* 27: 431–458.
- Malouf, Robert, Farrell Ackerman, and Arturs Semenuks. 2020. Lexical databases for computational analyses: A linguistic perspective. *Proceedings of the Society for Computation in Linguistics*. Vol. 3, Article 29.
- Marelli, Marco, and Marco Baroni. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological Review* 122: 485–515.
- Schäfer, Roland, and Felix Bildhauer. 2012. Building Large Corpora from the Web Using a New Efficient Tool Chain. *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, pp. 486–493.
- Vidra, Jonáš, Zdeněk Žabokrtský, Magda Ševčíková and Lukáš Kyjánek. 2019. DeriNet 2.0: Towards an All-in-One Word-Formation Resource. *Proceedings of the Second Workshop on Resources and Tools for Derivational Morphology (DeriMo 2019)*, pp. 81–89.



## From philology through computational linguistics to phonological typology and back: Development of syllabic consonants in Czech

Pavel Kosek (Masaryk University, Brno)

The talk addresses the historical phonology of Czech. It demonstrates how the phonological system can be seen through the lens of versology.

The evolution of the phonological system has been studied extensively from a philological perspective by Gebauer (1894) and analyzed within the functionalist framework by Trávníček (1935), Komárek (1969), and Lamprecht et al. (1986). Our goal is to present new research perspectives based on a computational approach to historical linguistic data and their interpretation within phonological typology. Specifically, we focus on the development of syllabic consonants.

First, we demonstrate how syllabic consonants can be reliably identified using an automatic parser that can be applied to texts from various historical periods of Czech (Ziková et al. 2023). Then, we will show that the data obtained from the parser reveal new trends in the development of syllabic consonants in Czech, which align with general linguistic principles such as the Sonority Sequencing Principle (Selkirk 1984, Parker 2011) and licensing of consonant clusters (Charette 1990, Scheer 2007, Cyran 2010). Finally, the paper turns back to philological aspects of the research on historical phonology. We discuss grapheme-phoneme correspondences in historical Czech orthography and compare different editorial strategies used in modern editions.

The talk is a collaborative work that is part of the GAČR project (GA23-04719S), which combines traditional philological research with computational linguistics and theoretical phonology.

### Bibliography

- Charette, Monik. 1990. License to govern. *Phonology* 7, 233–253.
- Cyran, Eugeniusz. 2010. *Complexity scales and licensing in phonology*. Berlin, New York.
- Gebauer, Jan. 1894. *Historická mluvnice jazyka českého I. Hláskosloví*. Praha.
- Komárek, Miroslav. 1969. *Historická mluvnice česká I. Hláskosloví*. Praha.
- Lamprecht, Arnošt – Šlosar, Dušan – Bauer, Jaroslav. 1986. *Historická mluvnice češtiny*. Praha.
- Parker, Steve. 2011. Sonority. In van Oostendorp, M. – Colin, E. J. – Hume, E. – Rice, K. (eds.), *The Blackwell Companion to Phonology. Vol. II: Suprasegmental and prosodic phonology*. 1160–1184. Oxford.
- Selkirk, Elisabeth. 1984. On the major class features and syllable theory. In Aronoff, M. & Oehrle, R. T. (eds.), *Language sound structure*. 107–136. Cambridge, MA.
- Trávníček, František. 1935. *Historická mluvnice československá: hláskosloví a tvarosloví*. Praha.
- Ziková, Markéta – Březina, Martin – Čech, Radek – Kosek, Pavel. 2023. Syllabic Consonants in Historical Czech and How to Identify Them. *Jazykovedný časopis* 74 (1), 391–400.

## Multilingualism from a psycholinguistics point of view

Barbara Mertins (TU Dortmund)

The perception of multilingualism in academic and non-academic contexts is still based on a number of myths and prejudices. These are related, for example, to code-switching or the so-called “perfect command” bilinguals are supposed to have in both languages (cf. de Houwer, 2009; Mertins, 2023; Delucchi Danhier & Mertins, 2018). Additionally, subjective prestige people assign to a particular language and the differentiation made between migration languages and educational languages greatly affect not only the perception of multilingualism but also the bilingual language use and the development of identity in bilingual speakers (Mertins, 2023; Mertins, Odermann & Alipaß, 2024).

Psycholinguistic data from at least the two last decades show that multilingualism brings cognitive, linguistic and metalinguistic advantages that are of a central importance, especially in educational contexts when it comes to the acquisition of literacy (cf. Odermann, 2023; Bialystok, 2012; Habertzettl & Wulff, 2014; Wermelinger, Gampe & Daum, 2017). In this lecture, I will present findings from a series of my own studies as well as data from other relevant research investigating psycholinguistic foundations of multilingualism and show that multilingualism offers a great potential for individuals as well as for the whole society (Pavlenko, 2014). Furthermore, I will explore on the issue why the discourse on multilingualism is often based on the dichotomy *advantages* versus *disadvantages* and provide a new empirically-based perspective on this debate. In conclusion, I will argue for seeing multilingualism as a natural state of mind of more than a half of the world’s population and discuss some effects multilingualism brings in terms of linguistic relativity and diversity (e.g. Andresen, Delucchi Danhier & Mertins, 2024; Mertins & Delucchi, 2023).

### Bibliography

- Andresen, H., Delucchi Danhier, R., & Mertins, B. (2024). Dominant-while-speaking: How bilingual Norwegian–English children conceptualize goal-oriented motion events. *International Journal of Bilingualism*, 13670069241251781. <https://doi.org/10.1177/13670069241251781>
- Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, 16(4), 240–250. <https://doi.org/10.1016/j.tics.2012.03.001>
- De Houwer, A. (2009). *Bilingual First Language Acquisition*. Multilingual Matters. <https://doi.org/10.21832/9781847691507>
- Delucchi Danhier, R., & Mertins, B. (2018). Psycholinguistische Grundlagen der Inklusion: Schwerpunkt Bilingualismus. In S. Hußmann & B. Welzel (Eds.), *DoProfiL – Das Dortmunder Profil für inklusionsorientierte Lehrerinnen- und Lehrerbildung* (p. 161–178). Waxmann.
- Habertzettl, S., & Wulff, N. (2014). In kurzer Zeit zum Erfolg: Der kompetente L2-Lerner. In I. Petersen & B. Lütke (Hrsg.), *Deutsch als Zweitsprache—Erwerben, lernen und lehren. Beiträge aus dem 9. Workshop Kinder mit Migrationshintergrund 2013* (S. 15–33). Klett-Fillibach.
- Mertins, B. (2023). Sprache und Identität. In H. Höfer, M. Delere, & T. Vogel-Lefèbre (Hrsg.), *Bildung, Diversität und Medien. Erkundungen entlang aktueller Diskurslinien* (S. 167–177). Kopaed.
- Mertins, B. & Delucchi Danhier, R. (ed.) (2023). *Diversity in Cognition*. Peter Lang Verlag.
- Mertins, B., Odermann, K., & Alipaß, J. (in print 2024). Mehrsprachigkeit in der deutschen Schule: Der Einfluss der eigenen Vorurteile von Lehrkräften auf den Schulerfolg der Schüler:innen mit dem sogenannten „Migrationshintergrund“. [Manuscript submitted for publication]. *Journal of Elementary Education*.
- Pavlenko, A. (2014). *The bilingual mind: And what it tells us about language and thought*. Cambridge University Press.

Wermelinger, S., Gampe, A., & Daum, M. M. (2017). Bilingual toddlers have advanced abilities to repair communication failure. *Journal of Experimental Child Psychology*, *155*, 84–94. <https://doi.org/10.1016/j.jecp.2016.11.005>

1

*Klitika: jejich fonologie,  
morfologie a syntax*

*Clitics: Phonology,  
morphology, syntax*

organizuje / organizer

Pavel Caha

(Masarykova Univerzita)

(CZE, ENG)

Bienále české lingvistiky 2024  
Biennial of Czech Linguistics 2024

## Popis panelu / Workshop description

**(CZE)** Klitiky jsou typicky definovány na základě fonologie: jsou to nepřízvučná slova, která se přimykají k sousednímu slovu, s nímž pak tvoří jednu prozodickou jednotku, tj. prozodické slovo. Stejnou vlastnost mají také afixy. Další vlastnosti klitik, které sdílí s afixy, přesahují hranice fonologie: klitiky nelze kontrastivně zdůraznit, nelze je koordinovat ani modifikovat (viz např. Zwicky 1985, Cardinaletti & Starke 1999).

Oproti afixům ale klitiky nemusí mít stálého hostitele: jejich pozice závisí na syntaktickém kontextu. Pro češtinu je tato pozice obecně charakterizována jako pozice za prvním syntaktickým konstituentem, což naznačuje, že umístění klitik je věcí syntaxe (viz Fried 1994, Toman 1996, Avgustinova & Oliva 1997, Lenertová 2004, Hana 2007, Tabachnick, v tisku).

Klitiky se tedy chovají jako afixy, ale zároveň se chovají i jako objekty ovlivněné syntaktickými procesy. Z tohoto důvodu jsou klitiky dobrým východiskem pro zkoumání obecnělingvistických otázek týkajících se hranic syntaxe, morfologie a fonologie. Klitiky se navíc nacházejí ve většině slovanských jazycích, což umožňuje jejich kontrastivní zkoumání (Marušič a kol., v tisku).

V panelu uvítáme příspěvky, které se zaměřují na fonologické, morfologické, syntaktické a sémantické chování klitik se zvláštním zaměřením na slovanské jazyky. Mezi možná témata patří (mimo jiné):

- Jaké segmentální a prozodické fonologické procesy probíhají na hranicích mezi klitiky a neklitiky? Chovají se klitiky jako slova nebo afixy? Nebo se liší od obou?
- Jaký je vztah mezi umístěním klitik v řetězci (2. pozice) a syntaktickými procesy, o kterých je známo, že jsou citlivé na lineární efekty, jako jsou vztahy koreference (vazání)?
- Vztahují se na klitiky obecná omezení týkající se syntaktického posunu (jako je např. coordinate-structure constraint)?
- Do jaké míry mohou být klitiky (ho, nám) synkretická se samostatnými slovy (nám) na straně jedné a afixy (dobr-é-ho) na straně druhé?
- Jaké morfologické vlastnosti lze považovat za charakteristické pro klitiky jako gramatickou třídu?

**(ENG)** One of the defining characteristics of clitics is phonological: they are accentless. Because they lack stress, they need to rely on a neighboring word in order to be pronounced within the same minimal prosodic unit: the prosodic word. This property also holds for affixes. Additional affix-like properties of clitics go beyond phonology: clitics cannot be focused, coordinated or modified (Zwicky 1985, Cardinaletti & Starke 1999, among others). In all these respects clitics behave like affixes.

But on the other hand, they need not have a permanent predetermined host. The exact location of their pronunciation depends on the syntactic context they are used in. Specifically, the position that Czech clitics occupy is generally defined as the position after the first syntactic constituent of the clause (for relevant discussion, see Fried 1994, Toman 1996, Avgustinova & Oliva 1997, Lenertová 2004, Hana 2007, Tabachnick to appear), suggesting that clitic positioning is a matter of syntax.

These descriptive observations lead to a tension. On the one side, clitics pattern with affixes, and on the other side, they behave like objects affected by syntactic processes. Given this tension, clitics allow linguists to explore some general questions concerning the boundaries of syntax, morphology and phonology. Moreover, clitics are found across a number Slavic languages and beyond, allowing for a fruitful cross-linguistic investigation (Marušič et al. to appear).

We welcome contributions that focus on phonological, morphological, syntactic and semantic behaviour of clitics with special focus on Slavic. Possible topics include (but are not limited to):

- What kind of segmental and prosodic processes take place at the boundaries between clitics and non-clitics? Do clitics behave like words or affixes? Or are they unlike either of those?
- Does the special placement of clitics in the string (2nd position) interact with syntactic processes that are known to be sensitive to linear effects, such as coreference relations (binding)?
- Do general constraints on filler-gap dependencies (such as the Coordinate Structure Constraint) apply to clitics?
- To what extent can clitics be syncretic with (i.e., identical to) independent words on the one hand, and affixes on the other hand?
- Which morphological properties may be potentially characteristic of clitics as a grammatical class?

### **Literatura / References**

Avgustinova, T. & K. Oliva. 1997. On the Nature of the Wackernagel Position in Czech. In Junghanns, U. & G. Zybatow (eds.), *Formale Slawistik*. Frankfurt am Main: Vervuert. 25–47.

Cardinaletti, A. & Starke, M. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In H. Riemsdijk (ed.), *Clitics in the Languages of Europe*. Berlin, New York: De Gruyter Mouton. 145-234.

Fried, M. 1994. Second-position clitics in Czech: Syntactic or phonological? *Lingua* 94, 155–175.

Lenertová, D. 2004. Czech Pronominal Clitics. *Journal of Slavic Linguistics* 12, 135–171.

Marušič, F., P. Mišmaš & R. Žaucer. to appear. Placement and Ordering of the (En)clitics. D. Šipka & W. Browne (eds.), *The Cambridge Handbook of Slavic Linguistics*. Cambridge: CUP.

Tabachnick, G. to appear. Intervention effects and phase boundaries in Czech clitic climbing. <https://s18798.pcdn.co/guyt/wp-content/uploads/sites/18383/2022/10/Tabachnick-FASL-proceedings-2022.pdf>.

Toman, J. 1996. A note on clitics and prosody. In A. L. Halpern & A. M. Zwicky, *Approaching second. Second position clitics and related phenomena*. Stanford, CA: CSLI Publications, 505–510.

Zwicky, A. M. 1985. Clitics and particles. *Language* 61(2), 283–305.

# Obsah / Table of contents

*Czech auxiliary clitics and their non-clitic counterparts*

Michaela Čakányová (Palacký University in Olomouc)

*Optimal clitic placement in Czech and Slovak*

Nina Adam (University of Potsdam)

*Syllabicity matters: phonological and prosodic factors influencing clausal position of the Old Czech be-auxiliary*

Martin Březina (Masaryk University)

*The perception of attachment direction of Slovenian second-position clitics*

Franc Marušič, Guy Tabachnick, Rok Žaucer (University of Nova Gorica)

*Phonology mirrors syntax: Czech pronominal clitics*

Anna Pořomská, Markéta Ziková (Masaryk University)

*Clitic-like pronouns in Burgenland Croatian PPs*

Luka Szucsich (Humboldt University Berlin)

*The order of genitive and accusative clitics in Czech and Slovenian*

Guy Tabachnick (University of Nova Gorica)

## Czech auxiliary clitics and their non-clitic counterparts

Michaela Čakányová (Palacký University in Olomouc)

This paper deals with the properties of Czech auxiliary clitics, such as the AUX *být* “be” used for past tense, present conditional mood and past conditional mood. It follows the studies of Lenertová (2004), Hana (2007), Dotlačil (2007), Veselovská (2008) and Kulik (2023) for the description of their basic properties and puts them into comparison with their non-clitic counterparts, namely the complementizers *aby*, *kdyby* and the lexical (copular) and non-clitic AUX verb *být*. The paper uses examples from the Czech National Corpus (2017, tit. ORAL) and (2020, tit. SYN2020) and runs morphosyntactic tests to show if and how these categories differ, including syntactic operations such as gapping in coordinate structures, their ability to appear in chains and their reduced form. The aim is to clearly describe the properties of AUX clitics in Czech and see how this description fits under the umbrella term of clitics in general, including their usually listed properties such as the inability to carry stress, the inability to appear in isolation, the fixed order, the morphological simplicity, the lack of selection demands on the host, and the ability to be combined with other members of the clitic class (Uhlířová, Kosta, and Veselovská 2017).

The AUX clitic *by* (in T position) used for conditional mood in Czech and the conjunctions *aby* and *kdyby* (in C position) have the same morphological paradigm. Their order with respect to other clitics is also identical. Because the complementizers *aby* and *kdyby* include the *-by* merged with the words *a* and *kdy* respectively, these conjunctions are always sentence initial. The *-by* in *aby* and *kdyby* is an affix. As such, it cannot stand independently, it does not take stress and its position is fixed. It can be followed by other clitics and the canonical order of the clitic chain is preserved (1-2). The gapping coordination involving movement out of conjuncts is also observed for the conditional and complementizer (3-4).

- |                                                                                                                                                |                                                                                                                                                             |
|------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1) Řekl jsem mu, že bych si to půjčil.<br>told am him that COND REFL it borrow<br>‘I told him that I would borrow it.’                         | 2) Řekl jsem mu, aby si to půjčil.<br>told am him COMP REFL it borrow<br>‘I told him to borrow it.’                                                         |
| 3) Já bych mu to dal a řekl (bych mu), ať jde.<br>I COND him it gave and told COND him to go<br>‘I would give it to him and tell him to come.’ | 4) Chce, abych mu to dal a (abych mu) řekl, ať jde.<br>want COMP him it gave and COMP him told to go<br>‘He wants me to give it to him and tell him to go.’ |

Unlike the complementizer, the conditional clitic does not have to always assume the second position.

- 5) Věděl jsem, že jejího otce by popadl vztek. (Meyer 2010, 370)  
knew<sub>M.SG</sub> aux<sub>1.SG</sub> that her father<sub>ACC</sub> COND seize furor  
‘I knew that her father would be infuriated.’

Another part of the analysis is focused on the past AUX clitic *být* “be” in opposition to the copular and passive AUX use. Both past and passive AUXs are sometimes labelled as clitics, but they have different syntactic properties and only the past AUX is a clitic. A special attention is given to the *-s* variation of the 2nd person singular form. If the form is reduced, it needs to attach to some other element (6) but not necessarily the first word or even the second one in the clause, but rather a phrase (7), thus behaving more like a clitic than a bound affix.

- |                                                                                                                                                                  |  |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| 6) Ty jsi onemocněl = Tys onemocněl.<br>you are fallen-ill you <sub>2.SG</sub> fallen-ill<br>‘You have fallen ill.’                                              |  |
| 7) Byls u naší tety. = U naší tetys byl. (Sgall 2011, 16)<br>were <sub>2.SG</sub> at our aunt at our aunt <sub>2.SG</sub> were<br>‘You have been to our aunt’s.’ |  |



By carrying out the tests, and by pointing out some key properties of the analysed expressions, the study tries to present a comprehensive overview of the verb *být* “be” clitics (mood and tense). And on the other part of the scale, it lists the non-clitic uses of this verb (voice and lexical V) and their properties.

## Bibliography

- Dotlačil, Jakub. 2007. “Why Clitics Cannot Climb Out of CP: A Discourse Approach.” In *Formal Approaches to Slavic Linguistics: The Toronto Meeting 2006*, edited by Richard Compton, Magdalena Goledzinowska, and Ulyana Savchenko, 76–93. Ann Arbor, MI: Michigan Slavic Publications.
- Hana, Jiri. 2007. “Czech Clitics in Higher Order Grammar.” PhD Thesis, Columbus, OH: The Ohio State University.
- Kopřivová, Marie, David Lukeš, Zuzana Komrsková, Zuzana Poukarová, Martina Waclawičová, Lucie Benešová, and Michal Křen. 2017. *ORAL: Korpus Neformální Mluvené Češtiny, Verze 1 z 2. 6. 2017*. Prague: Institute of the Czech National Corpus, Faculty of Arts, Charles University in Prague. <http://www.korpus.cz>.
- Křen, Michal, Václav Cvrček, J. Henyš, M. Hnátková, and T. Jelínek. 2020. *SYN2020: Reprezentativní Korpus Psané Češtiny*. Prague: Institute of the Czech National Corpus, Faculty of Arts, Charles University in Prague. <http://www.korpus.cz>.
- Kulik, Irenäus. 2023. “Clitic Climbing without Restructuring in Czech and Polish.” In *Advances in Formal Slavic Linguistics 2021*, edited by Petr Biskup, Marcel Börner, Olav Mueller-Reichau, and Iuliia Shcherbina. Open Slavic Linguistics. Berlin: Language Science Press.
- Lenertová, Denisa. 2004. “Czech Pronominal Clitics.” *Journal of Slavic Linguistics* 12 (1–2): 135–71.
- Meyer, Roland. 2010. “Mood in Czech and Slovak.” In *Mood in the Languages of Europe*, edited by Björn Rothstein and Rolf Thieroff, 358–75. Amsterdam: John Benjamins.
- Sgall, Petr. 2011. *Jazyk, Mluvení, Psaní*. Praha: Karolinum.
- Uhlířová, Ludmila, Peter Kosta, and Ludmila Veselovská. 2017. “Klitikon.” In *CzechEncy - Nový Encyklopedický Slovník Češtiny*, edited by Petr Karlík, Marek Nekula, and Jana Pleskalová. <https://www.czechency.org/slovník/KLITIKON>.
- Veselovská, Ludmila. 2008. “The Extended Verbal Projection in Czech: Three Variants of the Verb Be.” In *Formal Description of Slavic Languages*, edited by Gerhild Zybatow and Luka Szucsich, 555–69. Berlin: Peter Lang.

## Optimal clitic placement in Czech and Slovak

Nina Adam (University of Potsdam)

The phonological deficiency of clitics, in the sense that they prosodically attach to a neighbouring word or phrase, has been a core feature of their definition for decades (e.g., Zwicky, 1977), but is also central in more recent analyses (e.g., Franks, 2016). On the other hand, it has long been recognised that the term “clitic” comprises a very diverse set of lexical entities (e.g., Spencer & Luís, 2012). The proposed talk focuses on clitics in Czech and Slovak. I will argue that phonology has no role in the placement of these elements in the two languages, and that it is instead syntax, and indirectly information structure, that determines their positioning in their clause as well as their climbing behaviour. I will show how this can be modelled with competing, violable, and ranked constraints.

Czech and Slovak clitics (marked by bold print) typically appear after the first constituent of their clause; this can be a phrase (1a; 2a), but also the initial verb, especially in clauses with a dropped subject (1b; 2b). In embedded clauses, clitics appear directly after the complementiser (1c; 2c), or after an information-structurally marked phrase that has been moved to the left periphery (1d; 2d).

## (1) Czech

- a. Svému bratrovi **by** pomohla.  
 POSS.DAT brother.DAT COND.3 help.PTCP
- b. Pomohla **by** svému bratrovi.  
 help.PTCP COND.3 POSS.DAT brother.DAT
- c. ...že **by** pomohla svému bratrovi.  
 COMP COND.3 help.PTCP POSS.DAT brother.DAT
- d. ...že svému bratrovi **by** pomohla.  
 COMP POSS.DAT brother.DAT COND.3 help.PTCP  
 ‘(that) she would help her brother.’

## (2) Slovak

- a. Svoimu bratrovi **by** pomohla.  
 POSS.DAT brother.dat COND help.PTCP
- b. Pomohla **by** svojemu bratrovi.  
 help.PTCP COND POSS.DAT brother.DAT
- c. ...že **by** pomohla svojemu bratrovi.  
 COMP COND help.PTCP POSS.DAT brother.DAT
- d. ...že svojemu bratrovi **by** pomohla.  
 COMP POSS.DAT brother.DAT COND help.PTCP  
 ‘(that) she would help her brother.’

For Czech, Junghanns (2003) shows that the second position is not defined phonologically: if clitics appeared there because they require a prosodic host, prosodic boundaries should affect placement. This is the case in BCS, where clitics cannot follow a prosodic break and instead appear further to the right. In Czech, however, the second position is optimal for clitics even if they are framed by prosodic breaks there and pronounced in isolation. The same applies in Slovak, as (3) shows.

## (3) Slovak

- Dôkazy, ktoré predložila, **ho**, ako je vidieť, nepresvedčujú.  
 evidence.PL REL present.PTCP him.ACC as is see.INF NEG:convince.3PL  
 ‘Obviously, the evidence that she has presented does not convince him.’

If clitics do not require a host, altruistic movement of an element into the pre-clitic position becomes less plausible. Also, we must distinguish between information-structurally fronted phrases, as in (1a), (2a), (1d), and (2d), and elements like initial subjects or verbs, as in (1b) and (2b), which may well appear in their canonical position. I therefore argue that clitics appear in varying positions (as also shown for BCS by Bošković, 2001) because they are subject to two conflicting positional constraints, leading to different outcomes depending on the syntactic environment. Such an approach was proposed for Czech by Richardson (1997), elaborated by Adam (2024), and can be extended to Slovak.

Regarding clitic climbing, I will present a syntactic OT analysis based on Dotlačil’s (2007) finding that Czech clitics cannot move out of CP because they cannot fulfil salient information-structural roles. I propose that due to this deficiency, clitics cannot appear at the left edge of CP – and therefore, a representational analogue of the phase impenetrability condition (Chomsky, 1999) prevents their climbing to the matrix clause.



## Bibliography

- Adam, Nina (2024). *Optimal Clitic Positioning in Czech and Beyond. Clitic syntax, morphology, and prosody from a constraint-based perspective*. PhD thesis, University of Göttingen. <http://dx.doi.org/10.53846/goediss-10310>.
- Bošković, Željko (2001). *On the nature of the syntax-phonology interface: Cliticization and related phenomena*. Amsterdam: North Holland.
- Chomsky, Noam (1999). *Derivation by phase*. Cambridge: MIT Press.
- Dotlačil, Jakub (2007). Why clitics cannot climb out of CP: A discourse approach. In M. G. Ulyana Savchenko & R. Compton (Eds.), *Formal Approaches to Slavic Linguistics (FASL): The Toronto meeting 2006* (pp. 76–93). Ann Arbor: Michigan Slavic Publications.
- Franks, Steven (2016). Clitics are/become minimal(ist). In F. Marušič & R. Žaucer (Eds.), *Formal studies in Slovenian syntax*. Amsterdam: John Benjamins.
- Junghanns, Uwe (2003). *Non-cliticizing clitics?* Paper presented at the Fifth Conference on Formal Description of Slavic Languages (FDSL-5), University of Leipzig, 26-28 November.
- Richardson, Matthew (1997). Czech clitics: 2P or not 2P, that is the question. In L. Kaiser (Ed.), *Studies in the morphosyntax of clitics* (pp. 131–50). New Haven: Yale University.
- Spencer, Andrew & Luís, Ana (2012). *Clitics: An introduction*. Cambridge: Cambridge University Press.

**Syllabicity matters: phonological and prosodic factors influencing clausal position  
of the Old Czech *be*-auxiliary**

Martin Březina (Masaryk University)

**Aim.** This paper investigates phonological and prosodic factors influencing the O(ld) Cz(ech) grammatical auxiliary's position within the clause. Specifically, I analyze various present tense forms of the verb *býti* ('to be') in the 2SG and 3SG. The 2SG and 3SG stand out among other Person&Number combinations of the *be*-auxiliary as they adopt more than one distinct form: the syllabic (*j*)*si* (2SG, initial *j*- is often omitted) and *je(st)* (3SG, final cluster may be omitted), and the non-syllabic *s* (2SG) and *j* (3SG). This provides an excellent opportunity to explore the influence of phonological factors on word order.

**Background and data.** Previous studies (KOSEK 2015, 2017; BŘEZINA 2023a,b) have characterized the OCz *be*-auxiliary as an enclitic. As such, its position within the clause is highly restricted, that is, it mostly occupies the second position in the clause (2P), as in C(ontemporary) Cz(ech). Unlike CCz, however, it can also occur anywhere between the 2P and the position immediately after the *l*-participle (DP), even in main clauses (cf. CCz in PMČ: 648ff; FRIED 1994):

- (1) a. *Proto s<sub>2P</sub> toho křížě nezazřel,*  
therefore 2SG that cross NEG.saw.PTCP  
'Therefore, you did not see that cross.' (PASMUZA: 133)
- b. *neb nikdy jest<sub>DP</sub> otec bez syna nebyl.*  
because never 3SG father without son NEG.was.PTCP  
'Because (there) never was a father without a son.' (ŠTÍTKLEM: 11bisv)
- c. *chválú a čscí korunoval jsi<sub>DP</sub> mě*  
praise and honor crowned.PTCP 2SG me  
'You crowned me with praise and honor.' (ŘÁDKORA: 57r)

In this paper, I aim to attribute the *be*-auxiliary's variation in the clausal position to the following phonological and prosodic factors: (1) the form of the *be*-auxiliary itself (in terms of syllabicity), and (2) the form of the preceding element, i.e., (a) the final phoneme of the hosting phrase, V(owel)# or C(onsonant)#, and (b) the length of the initial phrase (in syllables).

**Material.** The sample used for this analysis comprises over 1,300 instances of the *be*-auxiliary and its context in the 14th and 15th c. texts, extracted from the OCz diachronic corpus, *Staročeská textová banka*.

**Results.** The analysis revealed that both the 2SG and the 3SG non-syllabic forms occupy the 2P significantly more often than the syllabic forms. Additionally, they proved to be more tolerant of the relatively long initial phrases. The most peculiar behavior is exhibited by the 3SG non-syllabic form *j*. This form is especially selective towards the ending of its host as it only follows V# hosts, even to the extent that it often affects its position within the clitic cluster, which is highly unusual:

- (2) a. *ješt<sub>o</sub> j mi můj otec vložil v srdce,*  
that 3SG me.DAT my father placed.PTCP in heart  
'That my father placed in my heart.' (ŠTÍTKLEM: 145v)
- b. *což mi j b<sup>ó</sup>h popřál,*  
what me.DAT 3SG God allowed.PTCP  
'What God has allowed me.' (ŠTÍTKLEM: 116v)

Although the 3SG form is covert in CCz, phonological effects similar to those shown in (2) are now posited on the non-syllabic 2SG form *s* (e. g., SEDLÁČEK 1994). Moreover, the presented OCz data corroborate various claims regarding the phonology's impact on clitic behavior found in the literature: phonologically lighter forms of the same category can move leftwards within the cluster, even in the same language (cf. ZIMMERLING & KOSTA 2013); formal reduction must occur before the movement (JUNG

2020); the root *-e-* keeps the 3SG forms in the rightmost position of the cluster (see discussion in FRANKS & KING 2000: 211ff), not their person and number features (cf. MIGDALSKI 2020).

## Bibliography

- BŘEZINA, M. 2023a. The Distribution of (Non-)Syllabic Present Tense Forms of the Verb *býti* in the Second-Person Singular in Old Czech. *Listy filologické*, 146(1–2), pp. 79–110.
- BŘEZINA, M. 2023b. Distribution of (non-)syllabic present tense forms of the verb *býti* in the 3rd-person singular in Old Czech. *Linguistica Brunensia*, 71(2), pp. 59–81.
- FRANKS, S., KING, T. H. 2000. *A Handbook of Slavic Clitics*. New York: Oxford University Press.
- FRIED, M. 1994. Second-position clitics in Czech: Syntactic or phonological? *Lingua*, 94, pp. 155–175.
- JUNG, H. 2020. The be-auxiliary's categorial status in Old Russian. *Studia Linguistica*, 74(1), pp. 1–32.
- KOSEK, P. 2015. Development of Word Order of Preterit Auxiliary Clitics in the Old Czech Bibles. In: *Slavic Languages in the Perspective of Formal Grammar: proceedings of FDSL 10.5, Brno 2014*, Frankfurt am Main: Peter Lang Edition, pp. 177–198.
- KOSEK, P. 2017. Die Wortstellung des Präteritum-Auxiliars in den ältesten tschechischen Prosatexten. *Zeitschrift für Slawistik*, Akademie-Verlag, 62(4), pp. 621–646.
- MIGDALSKI, K. 2020. The auxiliary 'be' as a bundle of phi-features. In: IONIN, T., MACDONALD, J. E. (eds.), *Annual Workshop on Formal Approaches to Slavic Linguistics: The Urbana-Champaign Meeting 2017*, Ann Arbor: Michigan Slavic Publications, pp. 246–263.
- PMC – GREPL, M. et al. 2012. *Průruční mluvnice češtiny*. Praha: NLN.
- SEDLACEK, M. 1994. Ty jsi se (učil), ty jsi si (pamatoval)? *Naše řeč*, 76(1), pp. 27–43.
- Staročeská textová banka*. Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka, v1.1.15. Available at: <<https://korpus.vokabular.ujc.cas.cz>>. All Old Czech data are extracted from this source and cited using its abbreviations.
- ZIMMERLING, A., KOSTA, P. 2013. Slavic clitics: a typology. *STUF – Language Typology and Universals*, 66(2), pp. 178–214.

## The perception of attachment direction of Slovenian second-position clitics

Franc Marušič, Guy Tabachnick, Rok Žaucer (University of Nova Gorica)

**Background** Slovenian has “second-position” pronominal and auxiliary clitics that cluster after the first syntactic constituent, as in (1). As these are clitics, they need to prosodically attach to an adjacent word, which they can do in either direction: when preceded by a prosodic break, as in (2), they attach to the following word; when followed by a prosodic break, as in (3), they attach to the preceding word. Golden and Milojević Sheppard (2000) and Toporišič (2000) assume that these clitics attach by default to the left, as predicted if they are in C (Golden & Milojević Sheppard, 2000): the complement of C, a phase head, should be spelled out in one phase, while C and its specifier (i.e. the clitics and the constituent preceding them) should be spelled out together in another phase (Chomsky, 2001). However, Orešnik (1984) argues empirically that Slovenian clitics usually attach to the right. We address this issue with an experiment testing clitic attachment direction through the perception of prosodic boundaries.

**Predictions** Since clitics cannot attach across a clause boundary, they must attach to the right when directly preceded by a subordinate clause, and to the left when followed by a subordinate clause. Following Orešnik (1984), we predict that clitics attach by default to the right, so neutral sentences without a subordinate clause should behave experimentally like sentences with a pre-clitic subordinate clause and differently from sentences with a post-clitic subordinate clause.

**Experiment** Our novel experimental paradigm tests perception of prosodic boundaries that are smaller than those tested in previous experiments (e.g. Krivokapić & Byrd, 2012) and are not detectable by differences in meaning (cf. Petrone et al., 2017). Participants heard Slovenian sentences with beeps inserted at points in or around clusters comprising two monosyllabic clitics, and identified the syllable or point between two syllables where they heard the beep (Figure 1). We used 30 stimulus sentences, including (1)–(3), divided into 5 experimental conditions, each with 6 target and 3 filler beep locations. We hypothesized that beeps are more likely to be perceived at prosodic boundaries than within a prosodic word. That is, beeps should be attracted to the left edge of the clitic cluster in condition 4 (which directly follows a subordinate clause) and to its right edge in condition 5 (which directly precedes one). Given our hypothesis that Slovenian clitics prefer to attach to the right, we predict that conditions 1–3 will pattern similarly to condition 4 and differently from condition 5. Each of the 49 native Slovenian speaker participants completed 90 trials: they heard each sentence in two target trials and one filler trial, and heard two sentences from each condition with beeps in each of the nine locations twice.

**Results** Figure 2 shows the distribution of perceived beeps by experimental condition. There is a clear difference between conditions 1–4, where responses were centered around the left edge of the clitic cluster (at point  $-1$ ), and condition 5, where responses were more balanced (the asymmetry is due to an overall leftward bias across all responses). Condition 4 had more responses at the left edge of the cluster itself, while conditions 1–3 also had many responses on the preceding syllable ( $-1.5$ ). These patterns are largely confirmed by a model predicting a response’s distance from the left edge of the cluster by condition (dummy-coded, with condition 4 as baseline) and actual beep location with random intercepts for participant and sentence. Responses were significantly closer to the edge in condition 4 than condition 5 ( $t = 5.31$ , estimated  $p < .001$ ); differences were much smaller and not significant between condition 4 and conditions 1–3 ( $|t| \leq 1.59$ ,  $p \geq .125$ ). Comparison of estimated marginal means found that responses were also significantly closer in conditions 1–3 than condition 5 ( $p < .001$  for all).

**Conclusion** These results are in line with our predictions: Large prosodic boundaries at subordinate clause edges attract the perception of beeps (as in condition 4). In neutral contexts (conditions 1–3), beeps are attracted to the left edge of the cluster, suggesting that Slovenian speakers perceive a prosodic boundary there—that is, that they attach clitics to the right. Our experiment thus supports the claim of Orešnik (1984), suggesting that either the particular phase-based account of clitic spell-out is incorrect or the phonology forces a prosody–syntax mismatch through readjustment in the last phase.





## Phonology mirrors syntax: Czech pronominal clitics

Anna Polómská, Markéta Ziková (Masaryk University)

This paper explores the syntax-phonology interface, with a focus on pronominal clitics in Czech, which are argued to be bound elements both syntactically and phonologically ([R5], [R6], [R9], [R14]). Syntactically, they are Wakernagel's clitics, positioned after the first syntactic constituent. Phonologically, they are enclitics, that integrate prosodically with the preceding P(honological) W(word). Typically, these aspects of cliticness are studied independently, but our aim is to investigate their interaction.

**Background.** The standard assumption is that the ordering of the first constituent and the clitic like Czech *ho* 'him' is the result of movement ([R1], [R15]), which involves different mechanisms for various constituent types. Verbal constituents and non-verbal constituents are differentiated. This distinction is based on the fact, that verbal pre-clitic constituents cannot be modified (1aa), unlike the non-V ones (1bb). This cross-linguistic observation leads some scholars to consider the V-hosts of clitics as heads ([R8], [R13], [R15]). According to [R11], these heads form a single syntactic constituent with the clitic. By contrast, non-V constituents are syntactic phrases, which are syntactically separated from the clitic.

**Hypothesis.** We hypothesize a lower degree of enclisis with phrases compared to complex heads, where the clitic is adjacent to the verb. To test this hypothesis, we analyze 3 phonological phenomena observed in spoken Czech: (i) degemination, (ii) regressive voicing assimilation in obstruent clusters, and (iii) obstruent devoicing. We focus on these processes since they are applied within the PW in Czech [R12], which is the prosodic domain where enclitics are typically integrated [R4]. While the first two processes occur within the PW, the third one indicates the PW boundary. In sum, the prosodic integration of the enclitic *ho* is expected to be more salient with the Vhost compared to the non-V-host. If a geminate arises due to the encliticization of *ho*, it is much more likely to be preserved at the non-V boundary than at the V-boundary, where degemination is expected to occur more frequently. Furthermore, we anticipate that voicing assimilation triggered by *ho* will be more productively applied to final obstruents of V-hosts. Conversely, final obstruents of non-V hosts are expected more likely to undergo devoicing. These hypotheses were tested experimentally.

**Experiments.** We conducted 3 experiments using recordings from 20 Czech native speakers. In total, we tested 420 items, a sample of which is provided in (2). Exp1 assessed the degree of (de)gemination, Exp2 examined the degree of regressive voicing assimilation, and Exp3 investigated the degree of obstruent devoicing before the clitic. We analyzed the phonetic characteristics of the /**bolded**/ strings using the *Praat* program ([R3], [R10]). In Exp1, we measured the degree of degemination in terms of the duration of the string [R7]. In Exp2-3, we measured pitch of the first obstruent to determine the fraction of locally (un)voiced frames in it [R2].

**Discussion.** The results in (3) confirm that the clitic *ho* is more prosodically integrated with a V-host than with a non-V-host. In Exp1, the greater degree of enclisis is manifested by the geminate having an average duration of 5ms shorter at the V-boundary than at the non-V-boundary. In Exp2-3, the deeper enclisis with a V-host is indicated by two observations: (i) the verb-final obstruent assimilates more frequently with the clitic (91% vs 83%), (ii) the verb-final obstruent undergoes devoicing less commonly before the clitic (8% vs 20%). Although enclisis is more pronounced with V-hosts across all observed parameters, the prosodic integration of the clitic still does not reach the level we observed with genuine suffixes. This suggests that pronominal clitics and suffixes are two distinct phonological categories in Czech.

**In future research,** we aim to investigate whether the prosodic integration hierarchy, i.e., stem+suffix > V-head+CL > non-V-phrase+CL, also applies to other phenomena, such as stress assignment, and to other types of clitics besides pronominal ones.

- (1) a. [Kousl] ho pes. aa. \*[Trochu kousl] ho pes.  
bit CL.3.SG.ACC dog a\_bit bit CL.3.SG.ACC dog
- b. [Pes] ho kousl. bb. [Zuřivý pes z útulku] ho kousl.  
dog CL.3.SG.ACC bit angry dog from shelter CL.3.SG.ACC bit

(2)	host	EXP1: (de)gemination	EXP2: assimilation	EXP3: devoicing
	V	Přisti/ <b>h</b> <b>h</b> /o během podvodu.	Přine/ <b>s</b> <b>h</b> /o do kanceláře.	Přive/ <b>z</b> <b>h</b> /o do soboty.
	non-V	Vra/ <b>h</b> <b>h</b> /o zranil.	Pe/ <b>s</b> <b>h</b> /o pokousal.	Mrá/ <b>z</b> <b>h</b> /o zranil.

(3)	clitic host	EXP1	EXP2	EXP3
	V-head	120 ms	91%	8%
	non-V-phrase	125 ms	83%	20%

### Bibliography.

- [R1] Anderson, S. 2005. *Aspects of the Theory of Clitics*.
- [R2] Bjorndahl, C. 2022. Voicing and frication at the phonetics-phonology interface: An acoustic study of Greek, Serbian, Russian, and English. *Journal of Phonetics* 92.
- [R3] Boersma, P. & Weenink, D. 1992–2024. *Praat: doing phonetics by computer*.
- [R4] Booij, G. 1996. Cliticization as prosodic integration: The case of Dutch. *The Linguistic Review* 13.
- [R5] Franks, S. et al. 2004. Pronominal Clitics in Slavic. *Journal of Slavic Linguistics* 12.
- [R6] Fried, M. 1999. Inherent vs. derived clisis: Evidence from Czech. *Journal of Linguistics* 35.
- [R7] Hedia, S. B. 2019. *Gemination and degemination in English affixation*.
- [R8] Kayne, R. 1991. Romance Clitics, Verb Movement, and pro. *Linguistic Inquiry* 22.
- [R9] Lenertová, D. 2004. Czech Pronominal Clitics. *Journal of Slavic Linguistics* 12.
- [R10] Machač, P. & Skarnitzl, R. 2009. *Principles of Phonetic Segmentation*.
- [R11] Matushansky, O. 2006. Head Movement in Linguistic Theory. *Linguistic Inquiry* 37.
- [R12] Palková, Z. 1997. *Fonetika a fonologie češtiny*.
- [R13] Rivero, M. L. 1991. Long Head Movement and Negation: Serbo-Croatian vs. Slovak and Czech. *The Linguistic Review* 4.
- [R14] Toman, J. 1996. A note on clitics and prosody. In Halpern, A. & Zwicky, A. (eds.), *Approaching second. Second position clitics and related phenomena*.
- [R15] Wilder, Ch. & Čavar, D. 1994. Long Head Movement? Verb movement and cliticization in Croatian. *Lingua* 93.
- [R16] Zwicky, A. 1985. Clitics and Particles. *Language* 61.

## Clitic-like pronouns in Burgenland Croatian PPs

Luka Szucsich (Humboldt University Berlin)

For Slavic languages, it has been claimed that clitics may not appear under P (\*[P clitic], cf. [1,2,7]). Similar to Slovenian [6] and Western varieties of Croatian [5], Burgenland Croatian (BCr) seems to contradict this claim by allowing for pronominal forms governed by Ps which at first sight look like clitics (1a,b). Still, I argue that the respective forms are actually strong, but truncated pronouns accompanied by stress and tone shift to the preposition.

Truncation of pronouns in BCr is possible only with 1.SG and 2.SG (and reflexive) pronouns in the ACC, (1a,b). With 3.SG/PL ACC, truncation is not licit, but stress and tone are retracted to the P (associated stress shift), (1c). With 1.PL and 2.PL only stress is retracted (dissociated stress shift), (1d). This shows that PP-internal clitic-like pronouns indeed cannot be analyzed as clitics. With Ps taking the GEN, truncation is not licit. However, with disyllabic pronouns, stress and tone is retracted, (2a), and with monosyllabic pronouns, only stress is retracted (leaving behind an unstressed long vowel on the pronoun), (2b). With Ps taking the DAT, no stress shift to P is observed whatsoever, (3).

I assume a functional spine of (pro)nominal categories as in (4) (see for similar assumptions [1,2,7]). Heads are projected only, if the respective features are present. I further assume that personal pronouns in BCr do not possess a separate head for the NOM. 3P pronouns contain an exponent for the  $nP$  ( $n(j)(e)-$ ) which combines with exponents spelling out number, class and case. In contrast, 1P and 2P pronouns do not have a separate exponent for the  $nP$ . They rather spell it out together with person and number. Exponents for case are spelled out separately, cf. [7].

I propose that truncation is restricted by morphosyntactic locality. Truncation is possible only if the PP immediately dominates the minimal case projection (ACC) which in turn dominates a projection spelling out the deletable chain of segments. Thus, truncation is possible only in the ACC, (5a), not in the GEN, (5b), despite the fact that ACC and GEN 1.SG ( $m'ène$ ) and 2.SG pronouns ( $t'èbe$ ) have the same form in BCr. Only in (5a) the PP immediately contains the minimal case projection dominating the projection with the exponent which can be truncated ( $men$ ). With 3P pronouns in the ACC, truncation of  $n(j)(e)-$  is excluded, since the  $nP$  is not immediately dominated by the minimal case projection, (6), with CL intervening between ACC and  $nP$ . This leaves only the option to spell out the non-truncated form (but with stress and tone shift): *zá njega* ‘for him’.

Stress shift is also subject to locality conditions, exemplified here with the case of ‘associated stress shift’ (shifting stress *and* tone), i.e., with rising accents in disyllabic pronouns. Rising accents are spreading high tones which are always associated with the stressed position preceding the syllable with the underlying high tone. So, in *nj'èga* ‘him.ACC’, underlyingly the /a/ in the syllable /ga/ bears a high tone which spreads to /e:/ in the preceding syllable (the same with *nj'èmu* ‘him.DAT’ with an underlyingly high tone on the /u/). Pronouns seek to avoid being stressed, but associated stress shift to P has to obey locality within morphosyntactic derivations. Locality restrictions apply to the morphophonological shape of the minimal potential pronominal exponent which may bear a tone. This exponent is always the ACC exponent which combines with the exponent  $n(j)(e)-$  in its complement domain. /nje:/ receives the tone as soon as the exponent /ga/ bearing a high tone is inserted into the structure. If this minimal potential exponent bearing a tone is immediately dominated by the PP, stress (and tone) may shift to P. In the case of 3.M.SG, the minimal potential exponent bearing tone is the same for the ACC and the GEN (/ga/). Consequently, stress and tone are shifted to P in both cases, cf. *p''ó njega* ‘for him.ACC’ and *zb''óg njega* ‘because of him.GEN’. However, DATP which is spelled out as /mu/ contains a more minimal potential exponent bearing tone, viz. /ga/ spelling out GENP and ACCP which are projected if DATP is projected. This blocks associated stress shift with *po nj'èmu* ‘after/according to him.DAT’. In the paper, I will also account for cases of ‘dissociated stress shift’.

- (1) a. Marica je ná te gl'edala.  
Marica be.3SG on you.ACC look.LPT-F.SG  
'Marica cared for you.'
- b. Fele vá me v'eru je.  
Fele in(to) me.ACC believe.PRS.3SG  
'Fele believes in me.'
- c. Jive je ná \*ga / <sup>ok</sup> njega pazil.  
Jive be.3SG on him.ACChim.ACC watch.LPT-M.SG  
'Jive watched him.'
- d. Tome je to zà nās kupil.  
Tome be.3SG that.ACC for us buy.LPT-M.SG  
'Tome bought this for us.'
- (2) a. Joži je to <sup>u</sup>od \*me / <sup>ok</sup> mene d<sup>u</sup>ostal.  
Joži be.3SG this.ACC from me.GEN me.GEN get.LPT-M.SG  
'Joži got this from me.'
- b. Jurica je to òd nj'ě / njī d<sup>u</sup>ostal.  
Jurica be.3SG this.ACC from her.GEN them.GEN get.LPT-M.SG  
'Jurica got this from her/them.'
- (3) a. Tome je \*pò mi / <sup>ok</sup> po mēni d<sup>u</sup>ošal.  
Tome be.3SG after me.DAT after me.DAT come.LPT-M.SG  
'Tome arrived after me.'
- b. To je \*p<sup>u</sup>ó njemu / <sup>ok</sup> po nj'ěmu t<sup>u</sup>ojkavost.  
This be.3SG after him.DAT after him.DAT stupidity.NOM.SG  
'In his opinion, this is stupid.'
- (4) [PP P ... [K<sub>DAT</sub> K<sub>DAT</sub> [K<sub>GEN</sub> K<sub>GEN</sub> [K<sub>ACC</sub> K<sub>ACC</sub> [CLP CL [#P # [πP π [nP n ]]]]]]]]]
- (5) a. [PP P [ACC<sub>P</sub> ACC [πP π [nP n ]]]]  
zá e men (zá me 'for me')
- b. [PP P [GEN<sub>P</sub> GEN [ACC<sub>P</sub> ACC [πP π [nP n ]]]]]  
zb<sup>u</sup>óg e men (zb<sup>u</sup>óg mene 'because of me')
- (6) \* [PP P [ACC<sub>P</sub> ACC [CLP CL [nP n ]]]]  
zá ga nje (\*zá ga 'for him')

## Bibliography

- [1] Abels, Klaus. (2003) *Successive Cyclicity, Anti-locality, and Adposition Stranding*. Ph.D. dissertation, University of Connecticut.
- [2] Abels, Klaus. (2012) *Phases. An essay on cyclicity in syntax*. Berlin/Boston: De Gruyter.
- [3] Caha, Pavel. (2009) *The nanosyntax of case*. Ph.D. dissertation, University of Tromsø.
- [4] Caha, Pavel. (2021) "Modeling declensions without declension features. The case of Russian". *Acta Linguistica Academica*, 68(4): 385–425.
- [5] Franks, Steven and Tracy Holloway King. (2000) *A handbook of Slavic clitics*. New York: OUP.
- [6] Priestly, Tom. (1993) "Slovene". Bernard Comrie and Greville Corbett, eds. *The Slavonic languages*. London: Routledge, 388–454.
- [7] Puškar-Gallien, Zorica. (to appear) "Morphosemantic mismatches with pronouns as a consequence of their internal structure". Berit Gehrke et al., eds. *Advances in formal Slavic linguistics 2022*. Berlin: Language Science Press.

## The order of genitive and accusative clitics in Czech and Slovenian

Guy Tabachnick (University of Nova Gorica)

**Background** Pronominal clitics in Czech and Slovenian cluster at the “second position” of clauses. In both languages, clitics generally follow a set order according to their case and reflexive features: reflexive–dative–genitive/accusative. Sentences with both genitive and accusative clitics are rare and their relative order is disputed. Franks and King (2000) claim that accusative–genitive order is basic in Slovenian and that judgements in Czech are strained. Other work on Czech shows similar variation: Veselovská (1995) and Toman (1999) place genitive clitics before accusative clitics, Lenertová (2004) claims that accusative–genitive order is required, and Hana (2007), like Franks and King (2000), notes that speaker preferences vary between preferring one or the other or rejecting both.

**Study** I present a corpus study of 428 Czech and 192 Slovenian sentences with genitive and accusative clitics from the SYNv11 corpus of the Czech National Corpus (Křen et al., 2022) and the Slovenian Gigafida 2.0 corpus (Krek et al., 2020). Tokens belonged to one of two constructions. In what I call the *quantity* construction, many quantity terms (including numbers 5 or greater) take genitive complements, which can extract and cliticize. When a genitive extracts from a numeral subject, it can cooccur with an accusative direct object clitic, as in (1). In the *ditransitive* construction, genitive and accusative clitics can cooccur as objects of a small number of ditransitives, as in (2). I fit a regression model (see Table 1) testing the effect of language, construction, and the person of the two arguments—since a preference for non–third-person clitics to precede third-person clitics (the Person Case Constraint) can override the usual case-based order of clitics in both languages (Stegovec, 2020; Sturgeon et al., 2011).

**Results** Figure 1 shows a strong effect of construction: when both clitics are third-person, quantity sentences always have genitive–accusative order, while ditransitive sentences (which are much rarer) show both orders. We also see a strong effect of person: first- and second-person clitics usually precede third-person clitics. This pattern is nearly exceptionless in Slovenian and in ditransitives; a minority of Czech quantity sentences have genitive third-person clitics before accusative first- and second-person clitics. The strength of the person effect is the only significant difference between the two languages.

**Discussion** Combinations of genitive and accusative clusters are rare in both languages, matching previous reports of their marginality, but speakers do produce them and show distinct patterns in their usage. Contrary to some previous reports, the sentences with two third-person clitics point to an underlying preference for genitive–accusative order. However, this preference is much weaker than the analogous preference for dative–accusative order, and can be overridden by the Person Case Constraint and perhaps other factors in ditransitive sentences. This weakness of preference is likely due, at least in part, to widespread syncretism between the two types of clitics, which are only distinguished in Czech in the third-person feminine singular (genitive *jí* vs. accusative *ji*) and third-person plural (genitive *jich* vs. accusative *je*). The genitive–accusative preference is also weaker in Slovenian (relative to the person effect) than in Czech, likely reflecting even greater clitic syncretism: the third-person feminine singular forms are distinct (genitive *je* vs. accusative *jo*), but the third-person plural forms are not (both *jih*). One possibility is that syncretic clitics can be formally marked as either genitive or accusative to satisfy competing demands of the case-based template and factors like the Person Case Constraint.

Careful study of corpus examples clarifies previous conflicting reports in the literature to suggest an often violated but distinctly present pressure towards genitive–accusative order. I suggest that this pressure is grammatically grounded—specifically, a preference for clitics to obey the case hierarchy of dative–genitive–accusative proposed independently by Caha (2009) to explain morphological syncretisms. This study opens the path for future research into a difficult corner of the clitic system, including experimental testing (to see whether genitive–accusative clitic combinations are limited to a small number of permissive speakers) and extension to Serbo-Croatian, which shows a similar pattern of claims to Czech and Slovenian (e.g. Browne, 2004; Franks & King, 2000).

- (1) Při hlasování **jich** **ho** podpořilo 14 ze 16.  
 during voting them.GEN him.ACC supported 14 from 16  
 '14 of 16 of them supported him in the vote.' (Czech)
- (2) Osvobodila **nas** **jih** bo globoka, zavzeta, in dejavna vera ...  
 release us.ACC them.GEN FUT.3SG deep committed and active faith  
 'What will release us from them [everyday resentments and bonds to the past] is a deep,  
 committed, and active faith ... (Slovenian)

	estimate	standard error	z score	p value
<b>Intercept</b>	<b>-2.15</b>	<b>.32</b>	<b>-6.64</b>	<b>&lt;.0001</b>
Person (default: both 3)				
<b>GEN 3, ACC 1/2</b>	<b>3.62</b>	<b>.37</b>	<b>9.86</b>	<b>&lt;.0001</b>
GEN 1/2, ACC 3	-0.37	.80	-0.46	.6434
Construction (default: quantity)				
<b>ditransitive</b>	<b>3.77</b>	<b>.67</b>	<b>5.63</b>	<b>&lt;.0001</b>
Language (default: Czech)				
Slovenian	-0.63	.56	-1.13	.2586
Person * language (default: both 3 * Czech)				
<b>GEN 3, ACC 1/2 * Slovenian</b>	<b>2.47</b>	<b>.93</b>	<b>2.66</b>	<b>.0078</b>
GEN 1/2, ACC 3 * Slovenian	-15.59	676.00	-0.02	.9816

Table 2: Regression model predicting order of genitive and accusative clitics, with significant ( $p < .05$ ) effects bolded (higher = accusative–genitive more likely, lower = genitive–accusative more likely)

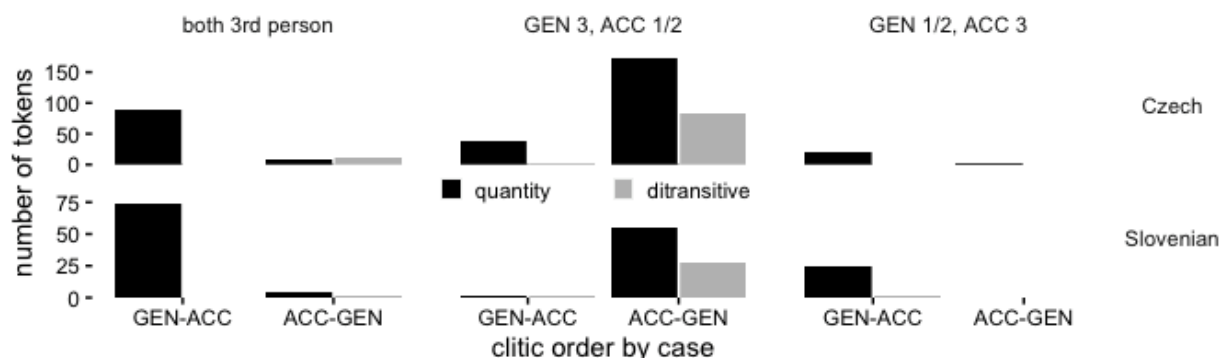


Figure 1: Case order of clitics in corpus tokens according to construction, person, and language

**Bibliography** Browne, W. (2004). Serbo-Croatian enclitics for English-speaking learners. *Journal of Slavic Linguistics*, 12(1/2), 249–283. • Caha, P. (2009). *The nanosyntax of case* [University of Tromsø]. • Franks, S., & King, T. H. (2000). *A handbook of Slavic clitics*. Oxford University Press. • Hana, J. (2007). *Czech clitics in Higher Order Grammar* [The Ohio State University]. • Krek, S., et al. (2020). Gigafida 2.0: The reference corpus of written standard Slovene. In N. Calzolari, et al. (Eds.), *Proceedings of the 12th Language Resources and Evaluation Conference* (pp. 3340–3345). • Křen, M., et al. (2022). Korpus SYN, verze 11 z 14.12.2022. *Ústav Českého národního korpusu FF UK*. • Lenertová, D. (2004). Czech pronominal clitics. *Journal of Slavic Linguistics*, 12(1–2), 135–171. • Stegovec, A. (2020). Taking case out of the person-case constraint. *Natural Language & Linguistic Theory*, 38(1), 261–311. • Sturgeon, A., et al. (2011). Revisiting the Person Case Constraint in Czech. In J. F. Bailyn, et al. (Eds.), *Formal Approaches to Slavic Linguistics #19: The 2nd College Park Meeting 2010* (pp. 116–130). Michigan Slavic Publications. • Toman, J. (1999). On clitic displacement. In M. Dimitrova-Vulchanova & L. Hellan (Eds.), *Topics in South Slavic syntax and semantics* (pp. 205–228). John Benjamins. • Veselovská, L. (1995). *Phrasal movement and X<sup>0</sup>-morphology: Word order parallels in Czech and English nominal and verbal projections* [Univerzita Palackého].

2

*Production  
and comprehension  
in adults  
from a psycholinguistic  
perspective*

organizers

Markéta Ceháková

James Brand

Jan Chromý

(ENG)

Bienále české lingvistiky 2024  
Biennial of Czech Linguistics 2024

# Workshop description

This workshop focuses on language processing in adults from a psycholinguistic perspective, offering empirical investigation within the domains of linguistics and cognitive science. Understanding how adults process language is a crucial pursuit, and this workshop aims to illuminate a wide spectrum of language processing issues, encompassing both production and comprehension.

Key themes and topics to be thoroughly examined in this workshop include:

(1) **Cognitive Mechanisms:** Through presentations and discussions, participants will focus on the cognitive mechanisms underpinning language processing. Researchers will explore how adults encode, store, retrieve, and manipulate linguistic information during both production and comprehension tasks, shedding light on the inner workings of language cognition (e.g. de Almeida & Gleitman, 2017; Rueschemeyer & Gaskell, 2018; Pickering & Garrod, 2013; Lewis et al., 2006; Van Dyke & Johns, 2012; Ferreira & Yang, 2019).

(2) **Processing Efficiency:** A central focus will be on how adults optimize language processing to enhance efficiency and comprehension speed. This discussion will encompass topics such as predictive processing (e.g. Levy, 2008; Futrell et al., 2020; Kuperberg & Florian Jaeger, 2016), context integration (e.g. Knoeferle, 2019), and the allocation of cognitive resources (e.g. Karimi & Ferreira, 2016; Traxler, 2014), providing insights into the strategies employed to navigate the complexities of language.

(3) **Psycholinguistic Variability:** The workshop will also explore individual and group differences in language processing, taking into consideration factors such as age, linguistic background, and cognitive abilities (e.g. Blasi et al., 2022; Gibson et al., 2013; Boland et al., 2016). Researchers will discuss how these variables influence language processing strategies and contribute to the rich diversity observed in adult language users. Moreover, the workshop also wants to stress out the importance of cross-linguistic differences in language processing and especially encourage submissions which aim to compare results obtained from different languages.

Furthermore, the workshop welcomes participants interested in investigating the neurological aspects of language processing in adult speakers. Additionally, it is open to those focused on exploring the practical implications of psycholinguistic research, with potential applications in fields such as clinical practice and human-computer interaction.

In conclusion, this workshop offers a comprehensive exploration of language processing in adults through the lens of psycholinguistics. Attendees would gain insights into the cognitive and neurological processes that underlie language use, with an emphasis on their practical applications across various domains.



## References

- de Almeida, R. G., & Gleitman L. R. (Eds.). (2017). *On Concepts, Modules, and Language: Cognitive Science at Its Core*. New York, Oxford Academic. doi: 10.1093/oso/9780190464783.001.0001
- Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, 26(12), 1153-1170. doi: 10.1016/j.tics.2022.09.015
- Boland, J. E., Kaan, E., Valdés Kroff, J., & Wulff, S. (2016). Psycholinguistics and variation in language processing. *Linguistics Vanguard*, 2(1), 20160064. doi: 10.1515/lingvan-2016-0064
- Ferreira, F., & Yang, Z. (2019). The Problem of Comprehension in Psycholinguistics. *Discourse Processes*, 56(7), 485–495. doi: 10.1080/0163853X.2019.1591885
- Futrell, R., Gibson, E., & Levy, R. P. (2020). Lossy-Context Surprisal: An Information-Theoretic Model of Memory Effects in Sentence Processing. *Cognitive Science*, 44(3). doi: 10.1111/cogs.12814
- Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A Noisy-Channel Account of Crosslinguistic Word-Order Variation. *Psychological Science*, 24(7), 1079-1088. doi: 10.1177/0956797612463705
- Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, 69(5), 1013-1040. doi: 10.1080/17470218.2015.1053951
- Knoeferle, P. (2019). Predicting (variability of) context effects in language comprehension. *Journal of Cultural Cognitive Science*, 3, 141-158. doi: 10.1007/s41809-019-00025-5
- Kuperberg, G. R., & Florian Jaeger, T. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32-59. doi: 10.1080/23273798.2015.1102299
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–77. doi: 10.1016/j.cognition.2007.05.006
- Lewis, R. L., Vasisht, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447-454. doi: 10.1016/j.tics.2006.08.007
- Pickering, M., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4), 329-347. doi: 10.1017/S0140525X12001495
- Rueschmeyer, S. A., & Gaskell, G. M. (Eds.). (2018). *The Oxford Handbook of Psycholinguistics*, 2nd edn. New York, Oxford Academic. doi: 10.1093/oxfordhb/9780198786825.001.0001
- Traxler, M. J. (2014). Trends in syntactic parsing: anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences*, 18(11), 605-611. doi: 10.1016/j.tics.2014.08.001
- Van Dyke, J. A., & Johns, C. L. (2012). Memory Interference as a Determinant of Language Comprehension. *Language and Linguistics Compass*, 6(4), 193–211. doi: 10.1002/lnc3.330

# Table of Contents

*The different roles of formal and conceptual properties in retaining morphosyntactic features among native and non-native speakers*

Denisa Bordag<sup>1,2</sup>, Andreas Opitz<sup>1</sup>, Bracha Nir<sup>2</sup> (<sup>1</sup>Leipzig University, <sup>2</sup>Haifa University)

*Diversity matters. Different types of garden-path sentences are processed differently.*

Markéta Ceháková, Jan Chromý (Charles University)

*HeCz: A large self-paced reading corpus of Czech newspaper headlines*

Jan Chromý<sup>1</sup>, Markéta Ceháková<sup>1</sup>, James Brand<sup>2</sup> (<sup>1</sup>Charles University, <sup>2</sup>Edge Hill University)

*Syntactic priming in Czech*

Maroš Filip (Institute of Psychology of CAS), Filip Smolík (Institute of Psychology of CAS)

*Immediate recall and information predictability in written and spoken comprehension*

Lucie Guštarová (Charles University)

*The negative concord illusion: An experiment with Czech neg-words*

Radim Lacina (Osnabrück University)

*The influence of frequency on disfluency distribution in connected speech of Czech speakers with aphasia*

Michal Láznička (Charles University)

*Examining language influence on categorization among Ukrainian-Russian simultaneous bilinguals: Insights from three-gendered languages*

Oleksandra Osypenko<sup>1</sup>, Silke Brandt<sup>1</sup>, Aina Casaponsa<sup>1</sup>, Panos Athanasopoulos<sup>2</sup> (<sup>1</sup>Lancaster University, <sup>2</sup>Lund University)

*Understanding time expression preferences and naming latencies from analog clocks: Insights from German and Czech speakers*

Elena Panfilova, Barbara Mertins (TU Dortmund)

*Genericity on a scale: Offline processing of masculine generic role-nouns in Czech*

Mikuláš Preininger<sup>1</sup>, James Brand<sup>2</sup>, Adam Kříž<sup>1</sup> (<sup>1</sup>Charles University, <sup>2</sup>Edge Hill University)

*How do people understand explicit and implicit contrastive and temporal discourse relations in Czech?*

Šárka Zikánová (Charles University, Faculty of Mathematics and Physics), Filip Smolík (Czech Academy of Sciences, Institute of Psychology & Charles University, Faculty of Arts)

## The different roles of formal and conceptual properties in retaining morphosyntactic features among native and non-native speakers

Denisa Bordag<sup>1,2</sup>, Andreas Opitz<sup>1</sup>, Bracha Nir<sup>2</sup> (<sup>1</sup>Leipzig University, <sup>2</sup>Haifa University)

It is uncontroversial that memory is a crucial component of the cognitive mechanisms underlying language processing. The present study aims to better understand how different types of information interact with the retention of linguistic material. Traditionally, studies investigating such interactions have focused on differences between memory for meaning/content and memory for verbatim (surface) linguistic forms. Results from this line of research suggest that memory for formal information (e.g., word order, grammatical voice) begins to decay rapidly shortly after processing, whereas propositional meaning/content information is more likely to be retained in long-term memory (Gernsbacher, 1985; Gurevich et al., 2010; Potter & Lombardi, 1998; Rummer et al., 2013). Furthermore, there is first evidence that non-native speakers have better memory for formal aspects of language than native speakers (Sampaio & Konopka 2013; Borag et al., 2021). In the present study, we set a slightly different focus and *investigated the role of differently salient formal and conceptual aspects of morphosyntactic information for retention*. Specifically, we investigated and compared the processing and retention of tense and number features in German, manipulating their formal and conceptual properties in both domains.

In two reading experiments, we tracked the eye movements of L1 (N=64) and L2 (N=63) German speakers while they read the same or grammatically modified sentences twice (12-16 sentences in between). When comparing changed and identical regions on the second presentation, longer fixations indicated that the change had been registered, implying that properties of the first presentation of the sentence were retained.

In the number conditions, we compared singular vs. plural marking and systematically manipulated whether the formal number differences included either a difference in number meaning (specific meaning, *'His son wants / His sons want to study medicine.'*) or not (generic / class meaning, *A bike is / Bikes are ecological means of transport*).

In the tense conditions, we systematically compared three tense forms (present, preterite, perfect) that are either similar or different in terms of their mode of formation (analytical vs. affixation) and in terms of their temporal meaning (referring to past vs. present events).

Our results (see also **Figures 1 & 2**) suggest that formal and conceptual information play different roles for native and non-native retention. For L1 participants, retention/registration effects were observed for grammatical changes related to imageable conceptual differences (e.g., specific readings in the number manipulation) or when both prominent formal and conceptual changes were involved at the same time (present↔perfect; affixation vs. analytic). Non-imageable tense changes that were not accompanied by salient formal changes were not registered (present↔preterite; both affixation). In contrast, L2 participants showed retention/registration effects (longer reading times in the changed condition) only when the manipulation involved salient formal changes (affixation vs. analytic forms). No effects were observed for less salient formal changes (affixation), independently of conceptual changes (e.g., no effects in number manipulation).

We conclude that formal aspects of grammatical features play a central role in L2 information processing and retention. In L1, the impact of formal aspects is less pronounced and only comes into play when accompanied by salient conceptual (functional) changes. Our study provides new insights into which

components (formal or conceptual) of grammatical features determine their retention in the memory of non-native and native readers.

Figure 1. Results of the number experiment.

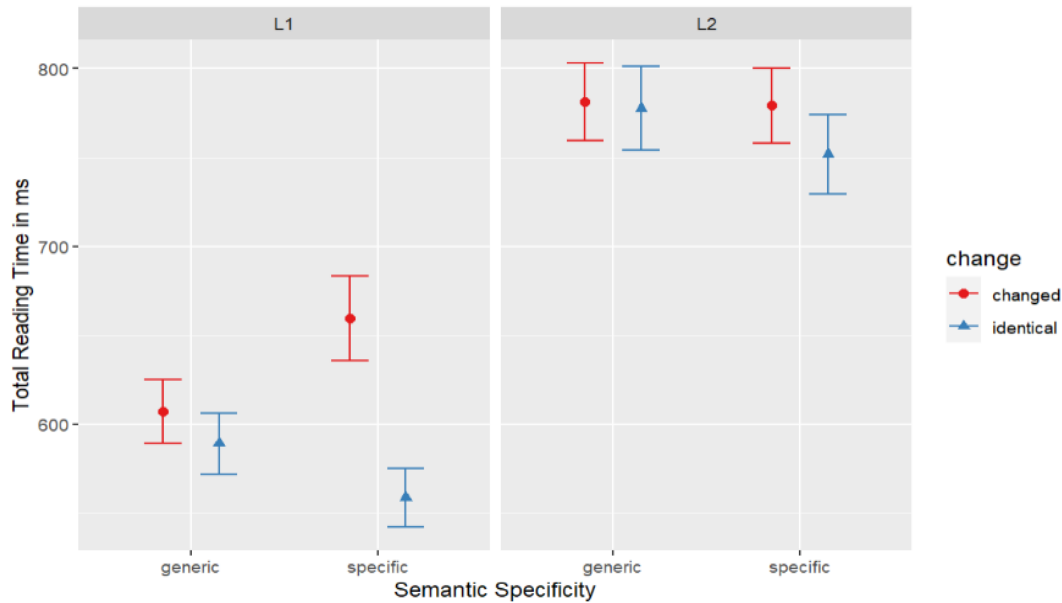
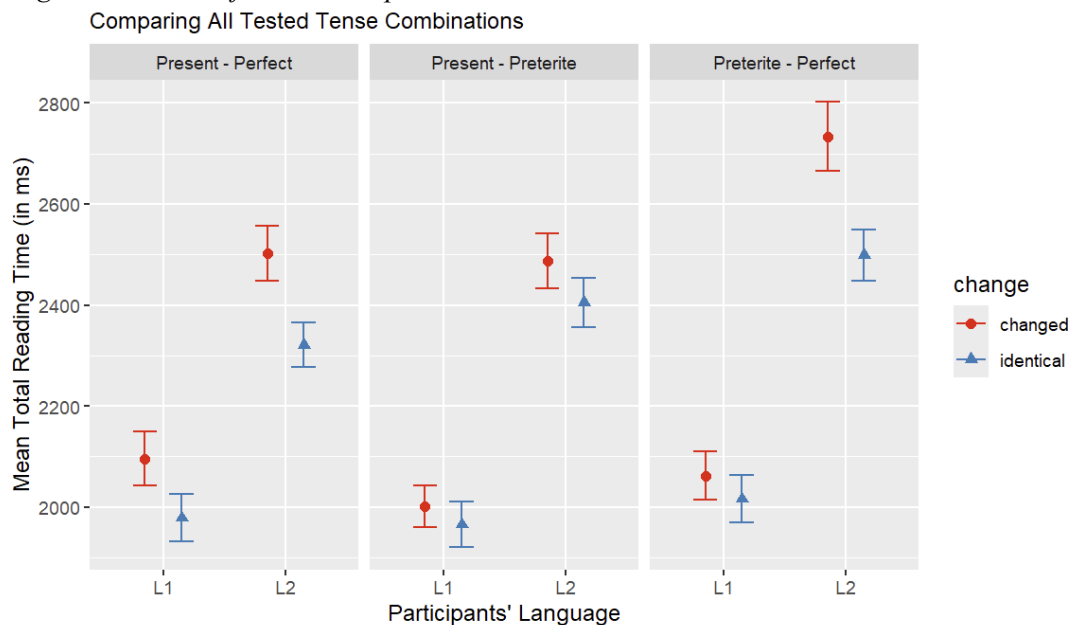


Figure 2. Results of the tense experiment.



## Bibliography

- Sampaio, C., & Konopka, A. E. (2013). Memory for non-native language: The role of lexical processing in the retention of surface form. *Memory, 21*(4), 537–544.
- Bordag, D., Opitz, A., Polter, M., & Meng, M. (2021). Non-native readers are more sensitive to changes in surface linguistic information than native readers. *Bilingualism: Language and Cognition, 24*(4), 599–611.
- Gurevich, O., Johnson, M. A., & Goldberg, A. E. (2010). Incidental verbatim memory for language. *Language and Cognition, 2*(1), 45–78.
- Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive psychology, 17*(3), 324–363.
- Potter, M. C., & Lombardi, L. (1998). Syntactic Priming in Immediate Recall of Sentences. *Journal of Memory and Language, 38*, 265–282.
- Rummer, R., Schweppe, J., & Martin, R. C. (2013). Two modality effects in verbal short-term memory: Evidence from sentence recall. *Journal of Cognitive Psychology, 25*, 231–247.

**Diversity matters. Different types of garden-path sentences are processed differently.**

Markéta Ceháková, Jan Chromý (Charles University)

**Background:** Multiple studies focusing on processing of garden-path sentences (GPs) showed that comprehenders often maintain the initial misanalysis even after they finish reading the whole sentence [1], [2], [3]. Comparing results of studies working with different GP structures also shows that the strength of this tendency depends on the type of ambiguity (e.g. the proportion of responses to comprehension questions which are consistent with the initial misanalysis can range from 10% in coordination GPs [4], to almost 80% in reduced relative clause GPs [5]). The nature of these differences and the driving force behind them (e.g. syntactic, semantic or pragmatic properties of the GPs in question) should be studied more deeply, because it can significantly contribute to our understanding of the processes involved in analyzing and interpreting GPs in general (e.g. whether comprehenders tend to compare the initial and the intended interpretation against each other in terms of their plausibility and then choose the more likely one, etc.).

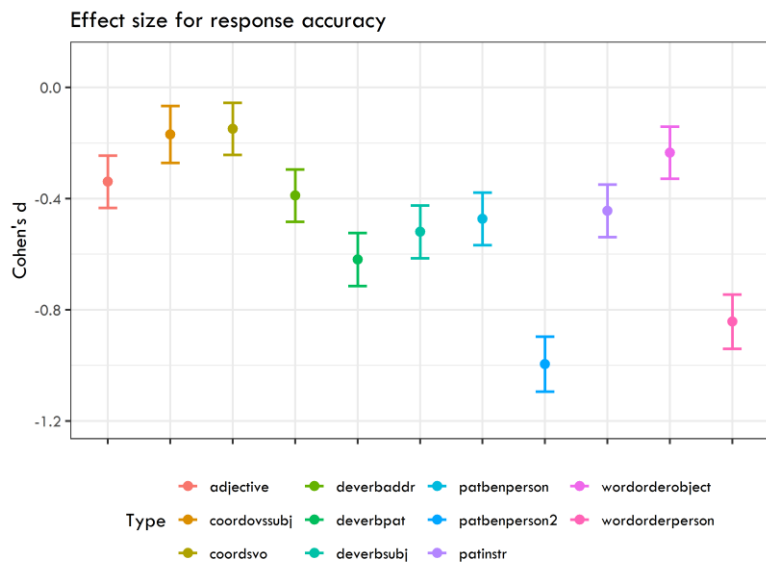
**Current study:** Since the presence and nature of such differences has not yet been consistently explored, our current study is mostly exploratory. Our main goal is to compare multiple types of GPs within the frame of one experiment and establish whether there are systematic differences in processing and interpretation between different types of structures. Our secondary goal is to pinpoint features which might be driving these differences for the purpose of future research. We expect to find differences in the size of the garden-path effect (reaction times [RTs] on the disambiguating region) and in the tendency of the initial misanalysis to linger (response accuracy to a question targeting the initial misanalysis).

**Materials & design:** We work with 11 types of Czech GPs, all matched for total length. Each type is represented by 6 items and each item also exists in a semantically similar non-ambiguous (nonGP) condition. The whole list of stimuli together with descriptions can be found on [https://osf.io/kstpe/?view\\_only=09bfbf5676a94f9da2be6acceddaae0f](https://osf.io/kstpe/?view_only=09bfbf5676a94f9da2be6acceddaae0f). We collected acceptability ratings (n=164) for a wider range of GP sentences and subsequently excluded types and items with poor ratings in the nonGP version, resulting in the final set of stimuli described above. To explore the differences, we conducted an online word-by-word self-paced reading experiment with yes/no comprehension questions (n=295). We used 66 experimental sentences and 90 fillers. We focused on RTs on the ambiguous region, disambiguating region (+ spillover region) and on response accuracy.

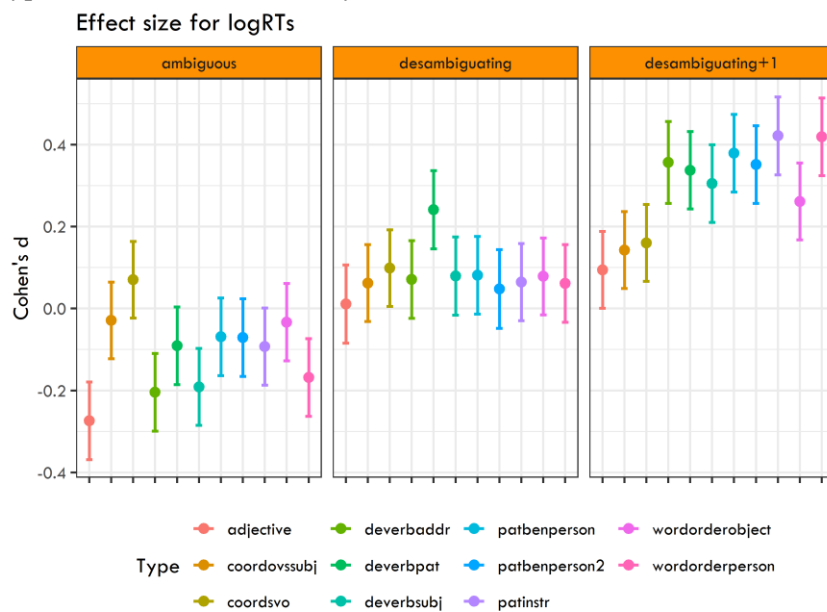
**Results & Analysis: Response accuracy.** We used logit mixed models with sentence type and ambiguity presence (GP vs. nonGP, nested under sentence type) as fixed effects and participants and items as random effects (both with ambiguity as random slope). We found a significantly ( $p < 0.05$ ) lower response accuracy for the GP condition in all sentence types. The effect size (Cohen's  $d$ ) varied between different types of GPs (Fig.1), with the highest values associated with sentences where the two potential interpretations were semantically similar. **RTs.** We used linear mixed models with  $\log(\text{RTs})$  as a dependent variable, sentence type and ambiguity presence (nested under sentence type) as fixed factors and participants and items as random effects. We found significantly longer RTs on the disambiguating region in the GP condition for three sentence types and the same effect on the spillover region for all types of sentences except one (with two of the remaining items only showing a weak effect). We also found longer RTs (with various effect sizes) on the ambiguous region in the nonGP condition in four types of sentences, which were most likely caused by low predictability of the noun form in the given context (e.g. a noun in instrumental case directly following a transitive verb).

**Discussion:** We found clear differences between the different types of GPs in terms of their effect sizes on RTs and response accuracy. The results also suggest that these differences might relate not only to

the syntactic properties of the ambiguity, but also to semantic and pragmatic properties of the initial and the intended analysis (such as their plausibility or mutual similarity).



**Fig 1.** Effect size (the difference in response accuracy between the GP and nonGP condition) for the 11 types of GPs used in the study.



**Fig 2.** Effect size (the difference in RTs on the ambiguous region, disambiguating region and spillover region) for the 11 types of GPs used in the study.

### Bibliography

- [1] Christianson K., Hollingworth A., Halliwell J.F., Ferreira F. (2001): Thematic roles assigned along the garden path linger. *Cognitive Psychology*. 42(4):368–407. [2] Slattery T.J., Sturt P., Christianson K., Yoshida M., Ferreira F. (2013): Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*. 2013;69(2):104–120 [3] Huang Y., Ferreira F. (2021): What causes lingering misinterpretations of garden-path sentences: Incorrect syntactic representations or fallible memory processes? *Journal of Memory and Language*. 121:104288 [4] Christianson K., Luke S.G. (2011): Context strengthens initial misinterpretations of text. *Scientific Studies of Reading*. 15(2):136–166. [5] Christianson K., Luke S.G., Hussey E.K., Wochna K.L. (2017):

Why reread? Evidence from garden-path and local coherence structures. *Quarterly Journal of Experimental Psychology*. 70(7):1380–1405.



**HeCz: A large self-paced reading corpus of Czech newspaper headlines**

Jan Chromý<sup>1</sup>, Markéta Ceháková<sup>1</sup>, James Brand<sup>2</sup> (<sup>1</sup>Charles University, <sup>2</sup>Edge Hill University)

**Introduction:** Linguistic corpora have been crucial to the study of language for nearly six decades (Kučera & Nelson, 1967), offering important insights into language use. More recently, researchers have begun to create corpora that not only include annotated texts, but also include behavioral measures of how these texts are processed by individuals (e.g., Kennedy et al., 2013; Kliegl et al., 2004 etc.). Currently, various eye-tracking and self-paced reading corpora are available, which differ in target language (cf. Siegelman et al., 2022), types of stimuli (natural or modified texts of various genres, isolated sentences), and additional information regarding both participants and stimuli (such as predictability norms, cf. Luke & Christianson, 2018), however such corpora tend to be relatively limited in size (cf. Futrell et al., 2020; Frank et al. 2013). This means that there are restrictions in terms of the scope of analyses that can be conducted using these corpora and the types of research questions being investigated.

**The HeCz Corpus Project:** This paper introduces the HeCz project – a large scale self-paced reading corpus of Czech newspaper headlines. The corpus comprises 1,919 headlines (23,634 tokens) sourced from Seznamzpravy.cz, the most widely read online news portal in Czech. The data were collected in two rounds (with an approximately one-month interval between the rounds) using word-by-word self-paced reading with comprehension questions following each headline. In the first round, we obtained data from 1,872 native Czech speakers. The headlines were randomly distributed into 16 lists, each comprising either 119 or 120 headlines. Each participant read the headlines from one list in a randomized order, with an average of 117 participants per list (range = 108–141 participants). In total, this core component of the corpus contains 2,986,615 observations. In the second round, roughly two thirds of the participants from the initial sample were re-tested, yielding another 1,853,460 observations (the average number of participants per list in the second round was 72.6; each was reading the same list of headlines as in the round 1). Additionally, the corpus provides comprehensive demographic data about the participants, including their age, L2, self-declared L2 proficiency, and foreign language exposure. Moreover, participants also completed the Czech adaptation of the Profile of Mood States questionnaire (Stuchlíková et al., 2005; cf. McNair et al, 1971, for original version), assessing the following moods ANGER, CONFUSION, DEPRESSION, FATIGUE, and TENSION. The corpus is automatically lemmatized and morphologically annotated using MorphoDiTa (Straková et al., 2014). As of March 2024, the automatic annotation is being manually checked, and the corpus is planned to be published in early Summer 2024.

**Goals:** The primary objective of the paper is to introduce the HeCz Corpus and outline its structure. We present four analyses from the corpus. First, we explore the relationship between headline length and reaction times (RTs) for individual words. A linear mixed-effects model reveals that the longer the sentence was, the faster the RTs were for individual words ( $\beta = -0.013$ ,  $SE = 0.0002$ ,  $t = -60.02$ ,  $p < 0.001$ ). This suggests sentence length plays an important role in the strategic allocation of cognitive resources during reading. Second, we address intra-participant reliability by analyzing data from participants who took part in both data collection rounds. We find a moderate correlation ( $r=0.42$ ) between RTs for individual words across the rounds. A linear mixed-effects model reveals a strong tendency for RTs to be faster in the second round ( $\beta = -0.097$ ,  $SE = 0.003$ ,  $t = -309.93$ ,  $p < 0.001$ ), emphasizing that task adaptation effects (Prasad & Linzen, 2021) may persist for a month or even longer. Third, we analyzed comprehension accuracy and its relation to headline length. The logit mixed-effects model yielded a significant effect of scaled length in words ( $\beta = -0.291$ ,  $SE = 0.027$ ,  $z = -10.624$ ,  $p < 0.001$ ) meaning that the longer the headline was, the lower was the comprehension accuracy. Fourth, we examined the effect of information position on comprehension accuracy and found that the closer was

the targeted information to the end of the headline, the higher was the accuracy ( $\beta = 0.431$ ,  $SE = 0.027$ ,  $z = 15.81$ ,  $p < 0.001$ ).

## References

- Frank, S. L., Monsalve, I. F., Thompson, R. L., & Vigliocco, G. G. (2013). Reading time data for evaluating broad-coverage models of English sentence processing. *Behavior Research Methods*, 45(4), 1182–1190.
- Futrell, R., Gibson, E., Tily, H. J., Blank, I., Vishnevetsky, A., Piantadosi, S. T., & Fedorenko, E. (2021). The Natural Stories corpus: a reading-time corpus of English texts containing rare syntactic constructions. *Language Resources and Evaluation*, 55, 63–77.
- Kennedy, A., Pynte, J., Murray, W. S., & Paul, S.-A. (2013). Frequency and predictability effects in the Dundee Corpus: An eye movement analysis. *Quarterly Journal of Experimental Psychology*, 66, 601–618.
- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16, 262–284.
- Kučera, H. & Francis, W. N. (1967). *Computational Analysis of Present-Day American English*. Providence: Brown University Press.
- Luke, S. G., & Christianson, K. (2018). The Provo Corpus: A large eye-tracking corpus with predictability norms. *Behavior Research Methods*, 50, 826–833.
- McNair, D., Lorr, M., & Doppleman, L. (1971). *POMS Manual for the Profile of Mood States*. San Diego, CA: Educational and Industrial Testing Service.
- Prasad, G., & Linzen, T. (2021). Rapid syntactic adaptation in self-paced reading: Detectable, but only with many participants. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(7), 1156–1172.
- Siegelman, N., Schroeder, S., Acartürk, C., Ahn, H. D., Alexeeva, S., Amenta, S., ... & Kuperman, V. (2022). Expanding horizons of cross-linguistic research on reading: The Multilingual Eye-movement Corpus (MECO). *Behavior Research Methods*, 54(6), 2843–2863.
- Straková, J., Straka, M., & Hajič, J. (2014). Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition. In *Proceedings of 52<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 13–18). Baltimore, Maryland: Association for Computational Linguistics.
- Stuchlíková, I., Man, F., & Hagtvet, K. (2005). Dotazník k měření afektivních stavů: konfirmační faktorová analýza krátké české verze. *Československá Psychologie*, 49(5), 459–469.

## Syntactic priming in Czech

Maroš Filip (Institute of Psychology of CAS), Filip Smolík (Institute of Psychology of CAS)

Repetition is a general ability of the cognitive system, often observed also in language production. For example, the repetition of previously perceived linguistic material can be explained by priming. In priming, exposure to a stimulus (*prime*) implicitly influences the response to a subsequent stimulus of the same type (*target*). In linguistic, there is an effect of syntactic priming (or structural priming), where target sentences are produced (or understood) more quickly because they share some structural features (e.g. syntax) with the prime sentences.

We conducted 6 experiments, similar to the pioneering study by Bock (1986), to investigate the effect of priming. Since syntactic priming has not yet been confirmed in Czech, the first step was to find out whether syntactic priming occurs in Czech, and then to look at other effects that influence it. Two different structures were used to test the priming effect. Transitive OVS and SVO structures (Sentences 1 and 2) and ditransitive structures with alternation of direct and indirect object (Sentences 3 and 4).

Sentence 1, SVO structure: Koza žrala trávu na louce. (ENG: A goat was eating grass in a meadow.)

Sentence 2, OVS structure: Trávu žrala koza na louce. (ENG: The grass was eaten by a goat in a meadow.)

Sentence 3, DAT/ACC structure: Veterinářka stříhá psovi drápy. (ENG: A vet is clipping dog's claws.)

Sentence 4, ACC/DAT structure: Veterinářka stříhá drápy psovi. (ENG: The vet is clipping claws of the dog.)

The six experiments varied slightly in design, but mainly in the accompanying effects that they studied (Table 1). The first investigated phenomenon was *lexical boost*, where repetition of the verb between prime and target sentences enhances the syntactic priming effect (Scheepers et al., 2017). Another focus was on role of morphology. We investigated the enhancing effect of repetition of case-marking ending of nouns. This effect should be analogous to the lexical boost effect, but it operates at the morphological level, hence we called it *morphological boost effect*. The previous studies focusing on the role of morphology yielded inconclusive results, finding some effect in Korean (Chung & Lee, 2017) but none in Basque (Santesteban et al., 2015). The third investigated element was the role of working memory in priming. As one study was conducted during the Covid pandemic in an online environment, we were also able to compare the impact of different social settings on priming. Generalized linear mixed models with random intercepts for subjects and items were used for the analyses.

Syntactic priming effect was observed only in ditransitive sentences (Fig. 1). Transitive structures with alternation of SVO/OVS order proved to be unsuitable for priming. The strong focus of the OVS structure on the object sounds probably strange if it is not accompanied by context, and then it is not primed. The effect of lexical boost was not confirmed in the experiments, but robust evidence from other studies suggest that low statistical power was probably the reason. Effect of working memory was also not observed. This may be because working memory can only affect structural priming in younger children, and in healthy adults it has developed to the point where it no longer plays a role. Interesting was also the observation that the online experiment did not produce as strong priming effects as the laboratory experiments. This is in line with the interactive alignment theory, which argues that priming is greater in social situations.

Probably the most important was the finding that the repetition of the case-marking morphemes of nouns can enhance the effect of structural priming. This suggests that syntactic processing is not completely independent of morphological processing, and they are processed at least partially in parallel. However,

it looks like not all case-marking morphemes had the same boost effect; the more prominent endings (-*ovi*) enhanced priming, but not the less prominent endings (-*e*).

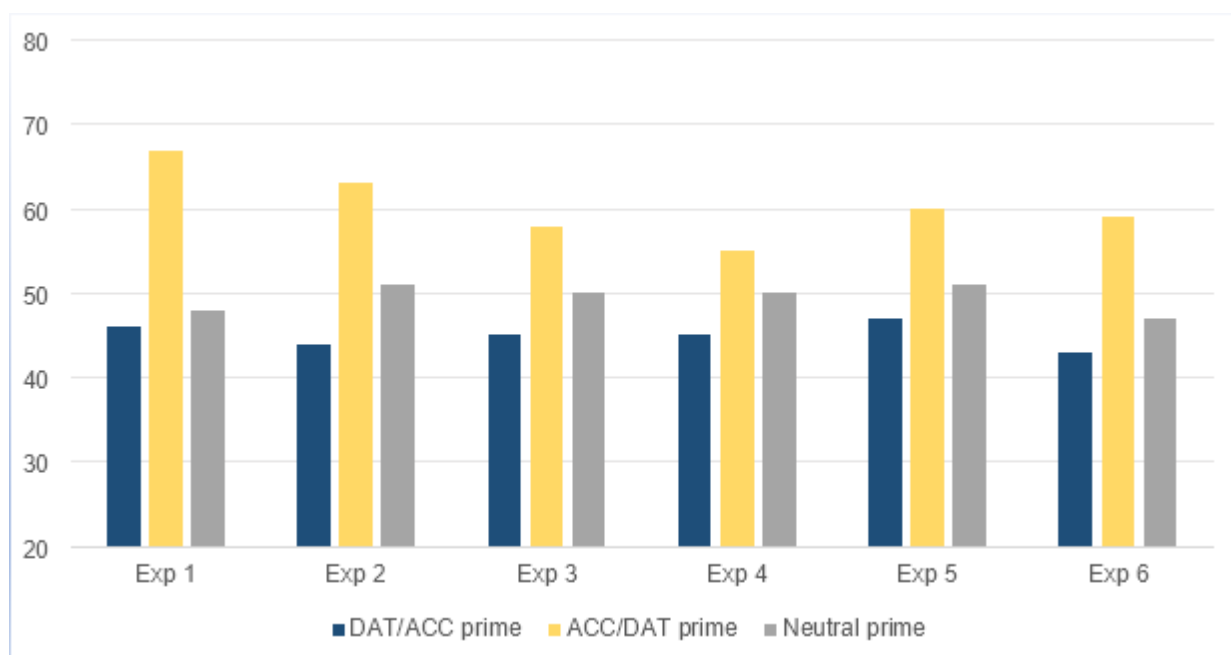
**Table 1**

Main differences between conducted experiments

	Investigated boost effect	Environment	Number of participants	Other investigated effects
<b>Experiment 1</b>	lexical	laboratory	62	SVO/OVS structure priming
<b>Experiment 2</b>	morphological	laboratory	63	working memory
<b>Experiment 3</b>	morphological	laboratory	64	-
<b>Experiment 4</b>	morphological	laboratory	99	-
<b>Experiment 5</b>	morphological	online	60	suffix type (-e vs -ovi)
<b>Experiment 6</b>	morphological	laboratory	61	suffix type (-e vs -ovi)

**Figure 1**

Proportion of ACC/DAT targets after different primes in six conducted experiments

**Literature**

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive psychology*, 18(3), 355-387.

Chung, E. S., & Lee, E. K. (2017). Morpho-syntactic processing of Korean case-marking and case drop. *Linguistic Research*, 34(2), 191-204.

Santesteban, M., Pickering, M. J., Laka, I., & Branigan, H. P. (2015). Effects of case-marking and head position on language production? Evidence from an ergative OV language. *Language, Cognition and Neuroscience*, 30(9), 1175.

Scheepers, C., Raffray, C. N., & Myachykov, A. (2017). The lexical boost effect is not diagnostic of lexically-specific syntactic representations. *Journal of Memory and Language*, 95, 102-115.

## Immediate recall and information predictability in written and spoken comprehension

Lucie Guštarová (Charles University)

**Background:** Recent studies [1, 2] have documented systematic differences in the extent to what readers recall certain types of information immediately after finishing reading a sentence. More precisely, it has been found that information conveyed by direct objects is recalled significantly better than information conveyed by temporal or locative adjuncts. This can be interpreted as a selective attention process: while reading, people drive their attention to information which they learned should be important/useful and they do not attend (or attend to a limited degree) to information that is peripheral [3]. The present study examines immediate recall of information conveyed by subjects and locative adjuncts. It focuses on the influence of predictability of locative information on information recall. Predictability has been intensively investigated in the domain of sentence processing [4, 5]. However, studies on the effects of predictability on information recall or later recognition have been rather scarce [6].

**Method:** First, predictability of locative adjuncts was normed. Fifty-seven transitive verbs not typically occurring with locative adjuncts were selected. Each of these verbs was combined with a predictable and non-predictable locative adjunct based on native speaker's intuition. Next, the degree to which a particular locative adjunct is expected in the combination with a given verb was judged by 115 native speakers and 24 verbs were selected for further testing (these verbs showed a robust difference in predictability between the two locative adjuncts). Next, stimuli for comprehension experiments were created. The written experiment (N=197) was conducted using a self-paced reading paradigm with whole sentences being presented at once. Each sentence was immediately followed by an open-ended comprehension question targeting either the information conveyed by the subject (Who did it?), or the information conveyed by the locative adjunct (Where did it happen?). The spoken experiment (almost ready to be run, as of March 2024) will use the same items, but audio-recorded by three native speakers of Czech with flat intonation (which is relatively common in Czech). The open-ended questions will be presented in writing and responses will be typed as in the written experiment. Both experiments use 24 experimental items together with 72 fillers and manipulate word order, information targeted by the comprehension question and locative adjunct predictability (see Table 1).

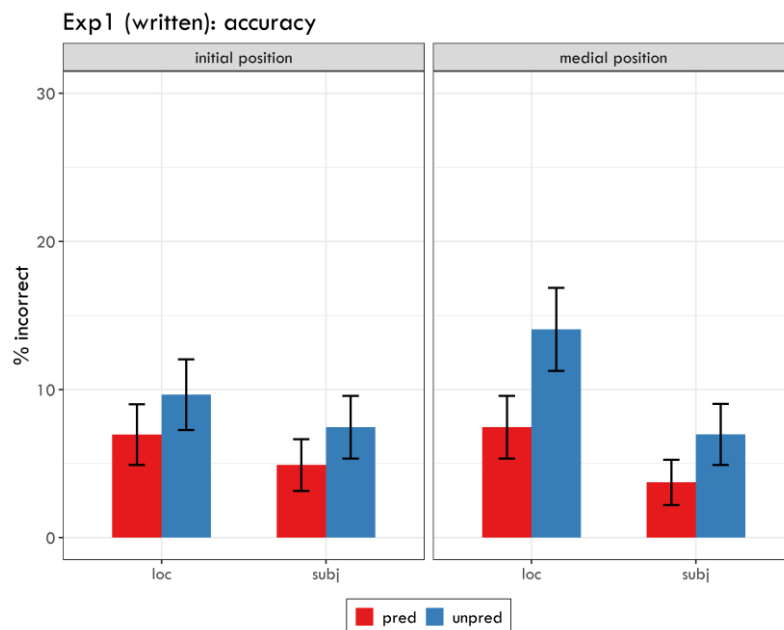
**Results:** Figure 1 shows the differences in recall accuracy between the four conditions in the written experiment. The nested logit mixed-effects model showed a general difference in recall: subjects were recalled better than locations ( $\beta = -0.704$ ,  $SE = 0.169$ ,  $z = -4.172$ ,  $p < 0.001$ ). Moreover, the model yielded a significant effect of predictability for both locative adjunct recall ( $\beta = -0.562$ ,  $SE = 0.145$ ,  $z = -3.869$ ,  $p < 0.001$ ) and subject recall ( $\beta = -0.567$ ,  $SE = 0.18$ ,  $z = -3.144$ ,  $p < 0.01$ ). No effect of word order has been observed, nor any significant interaction effect. The analysis of RTs yielded two significant effects. First, there was an effect of word order: sentences with LTVSO word order were processed slower than sentences with STVLO word order ( $\beta = 0.032$ ,  $SE = 0.013$ ,  $z = 2.565$ ,  $p < 0.05$ ). Second and crucially, there was an independent main effect of predictability: sentences with unpredictable locative adjuncts were processed slower than sentences with predictable locative adjuncts ( $\beta = 0.091$ ,  $SE = 0.012$ ,  $z = 7.765$ ,  $p < 0.001$ ).

**Discussion:** The present study shows a tendency to recall subjects better than locative adjuncts. This is in line with previous findings on Czech [1, 2]. Moreover, it was found that predictability may influence both online sentence processing (measured by RTs) and immediate recall. Sentences containing predictable locative adjunct were both read faster and yielded higher response accuracy. Interestingly, response accuracy was not only affected in case of the locative information (where predictability was manipulated), but also in case of the subject information (which stayed the same across the conditions). This suggests that low information predictability is related to a higher cognitive load which in turn influences attention patterns and memory encoding. The presentation will also include the results of the spoken experiment in which we expect similar tendencies, but generally higher rate of incorrect answers.

Table 1: Experimental item example. Word order values: *ltvso* = locative adjunct - temporal adjunct - verb - subject - object; *stvlo* = subject - temporal adjunct - verb - locative adjunct - object. Sentences have the same meaning and only differ in their word order and locative adjunct predictability: “Klára bought a really nice striped T-shirt in the store/in the park on Sunday.”

Word order	Predictability	Sentence
LTVSO	pred	V obchodě v neděli koupila Klára hrozně hezký pruho vaný tričko.
LTVSO	unpred	V parku v neděli koupila Klára hrozně hezký pruho vaný tričko.
STVLO	pred	Klára v neděli koupila v obchodě hrozně hezký pruho vaný tričko.
STVLO	unpred	Klára v neděli koupila v parku hrozně hezký pruho vaný tričko.

Figure 1: % of incorrect answers in the written experiment. *Loc* = locative adjuncts, *subj* = subjects, *pred* = predictable locative adjunct, *unpred* = unpredictable locative adjunct



## References

- [1] Chromý, J., & Vojvodić, S. (2024). When and where did it happen? Systematic differences in recall of core and optional sentence information. *Quarterly Journal of Experimental Psychology*, 77(1), 111–132.
- [2] Chromý, J. & Tomaschek, F. (submitted). Learning or boredom? Task adaptation effects in sentence processing experiments. *Open Mind*.
- [3] Ferreira, F., & Yang, Z. (2019). The problem of comprehension in psycholinguistics. *Discourse Processes*, 56(7), 485–495.
- [4] Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59.
- [5] Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: what is next? *Trends in Cognitive Sciences*, 27(11), 1032–1052.
- [6] Haeuser, K. I., & Kray, J. (2023). Effects of prediction error on episodic memory retrieval: evidence from sentence reading and word recognition. *Language, Cognition and Neuroscience*, 38(4), 558–574.



## The negative concord illusion: An experiment with Czech neg-words

Radim Lacina (Osnabrück University)

**Background:** In some languages, neg-words such as *nobody* (Laka, 1990) are grammatical only when the clause that they appear in also includes sentential negation. Take this Czech example:

- 1) Nikdo ne -udělal zkoušku z anatomie.  
 Nobody NEG-made exam from anatomy  
 ‘Nobody passed the anatomy exam.’

Even though (1) includes a case of double negation, the sentence is interpreted as only having one. This is the hallmark of *negative concord* (Giannakidou, 2006). Theoretical treatments of the phenomenon can be grouped into two families of accounts (Dočekal, 2017; Giannakidou & Zeijlstra, 2017). One essentially treats it as a subcase of negative polarity item licensing (NPI, e.g., Ladusaw, 1992). The other sees negative concord as a type of syntactic agreement (e.g., Zeijlstra, 2008).

Interestingly, both NPI and agreement phenomena give rise to the so-called *grammaticality illusions* in real-time processing (Phillips et al., 2011), whereby comprehenders can be induced into judging ungrammatical sentences as acceptable when these sentences include a grammatically irrelevant distractor element that nevertheless carries certain features important for the relation in question (morphosyntactic ones in agreement or those relevant to NPI licensing). In agreement, this is known as attraction and occurs in sentences such as *\*The key to the cabinets were rusty*, which comprehenders accept as grammatical in speeded acceptability tasks (Wagers et al., 2009) due to the number-matching word *cabinets*. Likewise, sentence such as *\*The student, who no teacher praises, has ever passed* with the unlicensed NPI *ever* are accepted (Orth et al., 2021) thanks to the phrase *no teacher* in the relative clause. Whether negative concord is subsumed under agreement or NPIs, similar illusions could arise therein. We attempt to test whether they do in the current study.

**Method:** We ran a speeded acceptability judgement experiment with native speakers of Czech (N = 112) testing whether the inclusion of an irrelevant negation into a relative clause could alleviate the ungrammaticality of the absence of the negative marker *ne-* in the matrix clause which included the neg-word quantifier *žádný* (*no*). Participants saw sentences presented word-by-word in the RSVP mode (350ms per word) and had 3s to binarily judge its acceptability (see an example item in Table 1). We manipulated the factors of matrix clause (MC, equivalent to grammaticality) and relative clause (RC) negation, which served as the illusion inducing element (see the crucial condition in (d) in Table 1).

**Results:** The responses by condition can be viewed in Figure 1 together with a plot of the associated response times. A mixed effects logit model showed that there was a main effect of MC negation on acceptability rates—negated sentences were judged as acceptable more often than affirmative ones ( $\beta = 1.83$ ,  $SE = 0.10$ ,  $z = 18.07$ ,  $p < 0.001$ ), in line with the grammar of Czech. Crucially, there was an interaction with RC negation ( $\beta = -0.39$ ,  $SE = 0.08$ ,  $z = -4.69$ ,  $p < 0.001$ ). A nested model further showed that there was a difference caused by RC negation within the affirmative MC condition (i.e. the ungrammatical one), where negation increased acceptability rates ( $\beta = 0.51$ ,  $SE = 0.14$ ,  $z = 3.58$ ,  $p < 0.001$ ). A model of log-RTs also showed a significant interaction of the two factors ( $\beta = -3.24$ ,  $SE = 0.009$ ,  $df = 3154$ ,  $t = -3.44$ ,  $p < 0.001$ ).

**Discussion:** The results show that there is indeed a “negative concord illusion” in Czech, since the presence of the irrelevant RC negation caused comprehenders to accept sentences with an unlicensed neg-word. This clear result contrasts with the lack of effects observed for number agreement attraction in the language (Chromý et al., 2023). The RT results support the illusion interpretation, as the presence of RC negation increased RTs in ungrammatical sentences. This first documentation of the effect opens new avenues for the study of linguistic processing illusions. Furthermore, future studies examining

under what conditions this effect is strengthened or goes away can help address whether negative concord should be treated as syntactic agreement or a case of NPI licencing.

Table 1: Example items from the experiment

- a) Minulý měsíc ne- vyhrál karbaník, který hledá novou práci, žádnou sázku.  
Last month NEG-won gambler who looks.for new job no wager
- b) \*Minulý měsíc vyhrál karbaník, který hledá novou práci, žádnou sázku.  
Last month won gambler who looks.for new job no wager
- c) Minulý měsíc ne- vyhrál karbaník, který ne- hledá žádnou práci, žádnou sázku.  
Last month NEG-won gambler who NEG-looks.for no job no wager
- d) \*Minulý měsíc vyhrál karbaník, který ne- hledá žádnou práci, žádnou sázku.  
Last month won gambler who NEG-looks.for no job no wager  
'Last month, the gambler, who is/isn't looking for a new/no job, didn't win/won no wager.'

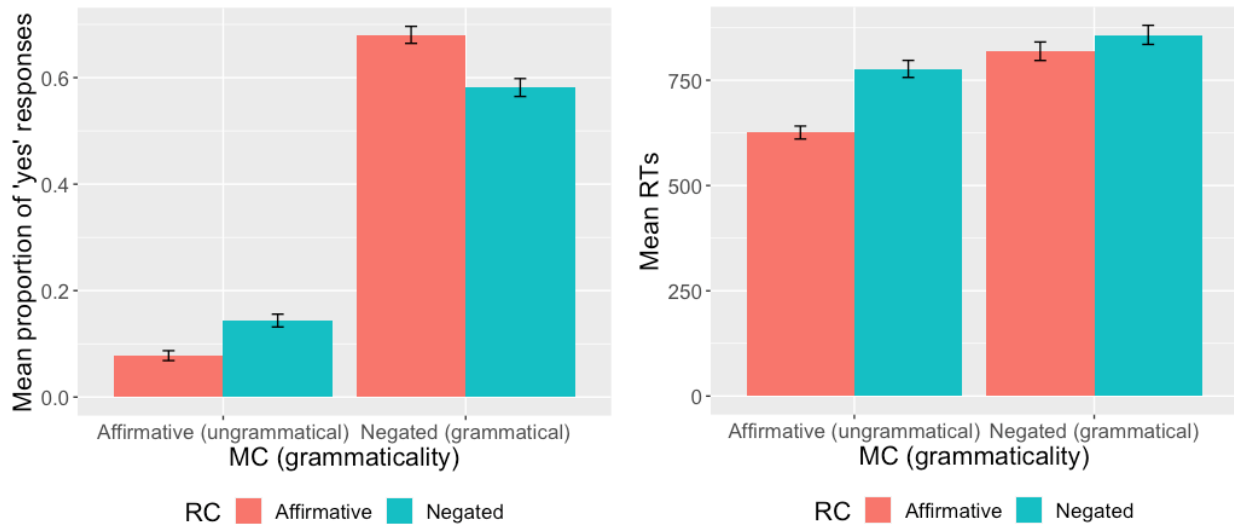


Figure 1: Mean responses (proportion of 'yes' answers) per condition on the left and mean RTs (both 'yes' and 'no' responses) per condition on the right

## Bibliography

Chromý, J., Lacina, R., & Dotlačil, J. (2023). Number Agreement Attraction in Czech Comprehension: Negligible Facilitation Effects. *Open Mind*, 7, 802-836.; Dočekal, M. (2017). N-words and NPIs: Between syntax, semantics, and experiments. *Advances in formal Slavic linguistics 2017*, 15.; Giannakidou, A. (2006). N-words and Negative Concord. *The Blackwell companion to syntax*, 327-391.; Giannakidou, A., & Zeijlstra, H. (2017). The landscape of negative dependencies: negative concord and n-words. *The Wiley Blackwell companion to syntax*, 1-38.; Ladusaw, W. A. (1992). Expressing negation. In *Semantics and linguistic theory* (pp. 237-260).; Laka, I. (1990). *Negation in syntax*. PhD Dissertation. MIT.; Orth, W., Yoshida, M., & Sloggett, S. (2021). Negative polarity item (NPI) illusion is a quantification phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47(6), 906.; Phillips, C., Wagers, M. W., & Lau, E. F. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In *Experiments at the Interfaces* (pp. 147-180). Brill.; Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, 61(2), 206-237.; Zeijlstra, H. (2008). Negative concord is syntactic agreement. *Ms., University of Amsterdam*, 5, 113.

## The influence of frequency on disfluency distribution in connected speech of Czech speakers with aphasia

Michal Láznička (Charles University)

Language in aphasia presents a unique testing ground for the models of language processing and representation, as such models should in principle be able to account for the patterns of language use observed in this syndrome. For instance, agrammatic aphasia has been used to argue for generativist models of language representation (e.g. Grodzinsky 2000). More recently, usage-based and cognitive linguistics have been gaining attention in the study of aphasic language with results suggesting that these approaches may have the benefit of greater descriptive adequacy and potentially also explanatory power (Martínez-Ferreiro et al. 2018; Hatchard & Lieven 2019; Bruns et al. 2019).

Frequent disfluencies are one of the hallmark features of language in aphasia. Disfluencies are associated with increased processing load and various factors have been investigated that may influence the occurrence of disfluencies, including strength of association between words (Goldman-Eisler 1958; Schneider 2016), which can be connected to chunkhood and entrenchment (Bybee 2010; Divjak 2019). Usage-based approaches view the deficits observed in aphasia as a result of limited processing resources and interpret the differences between neurotypical language users and persons with aphasia as quantitative in nature. This should hold not only for word-level phenomena but for syntactic patterns as well, as argued by Gahl & Menn (2016). In this paper, I present a usage-based analysis of the distribution of disfluencies in prepositional phrases (PP) produced in connected speech by ten Czech speakers with different types of aphasia and varying levels of fluency and severity, focusing on the role of frequency. I analyzed a total of 647 PPs with nominal complements. The data was annotated for complement lemma, number and case, presence and type of adnominal modifier, and for the presence, place of occurrence, and type of disfluency. Furthermore, cumulative frequency of the complement lemma, bigram frequency of the PP, relative frequency of the word form, and forward and backward transitional probability between the preposition and the complement were obtained from three corpora of Czech (spoken and written language, and movie subtitles). These distributional characteristics were combined using hierarchical clustering in order to obtain a rough measure of retrieval difficulty.

PPs containing an adnominal modifier were generally produced less fluently than bare P N constructs, suggesting increased processing load associated with increased phrase length and complexity. Crucially, there are visible tendencies in the data that suggest the influence of frequency. High frequency PP tokens exhibit higher levels of fluency. Fully fluent production is associated with complements with high lemma frequency (e.g. *za babičkou* ‘to grandma’) and with lexicalized PPs with reduced compositionality (*v pohodě* ‘alright’). Furthermore, PPs that are parts of larger patterns such as *prosil o pomoc* ‘was begging for help’ also tend to be fluent. The data also shows an interesting difference between pre-phrasally and phrase-internally disfluent production. High frequency PP tokens have a higher proportion of pre-phrasal disfluencies, indicating the phrase’s chunkhood. For instance, *do klece* ‘into (the) cage’, which has a relatively weaker association, occurs several times with phrase-internal disfluencies in the corpus, while *v kleci* ‘in (the) cage’, a token with stronger association, was produced fluently or with a pre-phrasal disfluency. Interestingly, this pattern holds for speakers with mild fluent aphasia, but also for more severely impaired fluent speakers and even for speakers with a reduced fluency level.

While this is a descriptive analysis based on observational data, the results are in line with previous usage-based studies of aphasic language. The data can thus be interpreted as supporting evidence for the usage-based construction grammatical view of language representation. Simultaneously, the observed patterns could be used for the development of novel strategies in aphasia assessment and therapy, based on item-specific distributional characteristics.



## Bibliography

- Bruns, Claudia, Rosemary Varley, Vitor C. Zimmerer, Marcella Carragher, Gwen Brekelmans & Suzanne Beeke. 2019. "I don't know": a usage-based approach to familiar collocations in non fluent aphasia. *Aphasiology* 33(2). 140–162. <https://doi.org/10.1080/02687038.2018.1535692>.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Leiden: Cambridge University Press.
- Divjak, Dagmar. 2019. *Frequency in Language: Memory, Attention and Learning*. 1st edn. Cambridge University Press. <https://doi.org/10.1017/9781316084410>.
- Gahl, Susanne & Lise Menn. 2016. Usage-based approaches to aphasia. *Aphasiology* 30(11). 1361–1377. <https://doi.org/10.1080/02687038.2016.1140120>.
- Goldman-Eisler, Frieda. 1958. The Predictability of Words in Context and the Length of Pauses in Speech. *Language and Speech* 1(3). 226–231. <https://doi.org/10.1177/002383095800100308>.
- Grodzinsky, Yosef. 2000. The Trace Deletion Hypothesis and the Tree-Pruning Hypothesis: Still valid characterizations of Broca's aphasia. *Behavioral and Brain Sciences* 23(1). 55–64. <https://doi.org/10.1017/S0140525X00582399>.
- Hatchard, Rachel & Elena Lieven. 2019. Inflection of nouns for grammatical number in spoken narratives by people with aphasia: how glass slippers challenge the rule-based approach. *Language and Cognition* 11(3). 341–372. <https://doi.org/10.1017/langcog.2019.21>.
- Martínez-Ferreiro, Silvia, Byurakn Ishkhanyan, Vicente Rosell-Clarí & Kasper Boye. 2018. Prepositions and pronouns in connected discourse of individuals with aphasia. *Clinical Linguistics & Phonetics* 33(6). 497–517. <https://doi.org/10.1080/02699206.2018.1551935>.
- Schneider, Ulrike. 2016. Hesitation placement as evidence for chunking: A corpus-based study of spoken English. In *Experience Counts: Frequency Effects in Language*, 61–90. De Gruyter. <https://doi.org/10.1515/9783110346916-004>.

**Examining language influence on categorization  
among Ukrainian-Russian simultaneous bilinguals:  
Insights from three-gendered languages**

Oleksandra Osypenko<sup>1</sup>, Silke Brandt<sup>1</sup>, Aina Casaponsa<sup>1</sup>, Panos Athanasopoulos<sup>2</sup>

(<sup>1</sup>Lancaster University, <sup>2</sup>Lund University)

**Background.** This study explores the linguistic relativity principle (Whorf, 1956): how do the languages people speak affect their cognitive processes? We address this question in the domain of grammatical gender, in simultaneous bilinguals of Ukrainian and Russian with contrasting grammatical three-gendered systems. Incorporating three-gendered languages, particularly the neuter gender, addresses a significant critique in the field of linguistic relativity, which argues that the cognition of speakers of three-gendered languages is not as influenced by gender compared to speakers of two-gendered languages, due to the binary nature of the latter that makes grammatical gender more salient (Sera et al., 2002). Adopting an object similarity-judgement paradigm (Phillips & Boroditsky 2003), we designed two experiments. Experiment 1 incorporated items with neuter grammatical gender, while Experiment 2 omitted them.

**Methods and Study Design.** In both experiments, we recruited English monolinguals and Ukrainian-Russian bilinguals, who rated the similarity of pairs of conceptually neutral object images, each presented with a male or female character image (e.g., notebook and a ballerina), on a 9-point Likert scale. The nouns varied in their grammatical gender alignment in Ukrainian and Russian, including those with opposite grammatical genders in both languages (e.g., masculine in Ukrainian and feminine in Russian) and nouns with matching genders (e.g., masculine in both). Two key predictions were tested using linear mixed models. First, we predicted that bilinguals would demonstrate stronger language effects than English monolinguals, for nouns with matching grammatical gender in Ukrainian and Russian. Specifically, we expected higher ratings for pairs where the biological sex of the character and the grammatical gender of the object were congruent in both Ukrainian and Russian. Second, for nouns with mismatching genders in Ukrainian and Russian, we predicted that the ratings would be influenced by the bilinguals' most proficient language, with higher ratings for pairs congruent in that language.

**Results and Discussion.** Experiment 1 showed no significant grammatical gender effects on bilingual categorisation, supporting the notion that neuter grammatical gender yields weaker effects compared to two-gendered languages. Conversely, Experiment 2 revealed a significant language effect, with bilinguals rating pairs more similar when grammatical gender and biological sex were congruent in their more proficient language, compared to monolinguals. The findings highlight the role of neuter gender in modulating the influence of grammatical gender on cognitive processes. The absence of significant effects in Experiment 1 aligns with previous research suggesting that neuter gender diminishes grammatical gender effects. In contrast, our results in Experiment 2 demonstrate that excluding neuter gender can reveal these effects, indicating that grammatical gender does impact categorization processes in three-gendered languages. The current study also emphasizes the importance of language proficiency in cognition of simultaneous/early bilinguals. The tendency of Ukrainian-Russian bilingual participants to align their similarity ratings with their more proficient language suggests that cognitive processes might be influenced to a greater extent by the more proficient grammatical system.

Table 3. Examples of the stimuli

Grammatical gender congruency in Ukrainian and Russian	Gender of the character	Example of a fixed pairing across the sample
<b>1. Neutral in both (Experiment 1 only)</b>	(a) Feminine	Apple and a girl
	(b) Masculine	Apple and an architect
<b>2. Masculine in both</b>	(a) Feminine	Key and a bride
	(b) Masculine	Key and a groom
<b>3. Feminine in both</b>	(a) Feminine	Candle and witch
	(b) Masculine	Candle and king
<b>4. Feminine in Ukrainian, Masculine in Russian</b>	(a) Feminine	Umbrella and a female ogre
	(b) Masculine	Umbrella and a male ogre
<b>5. Masculine in Ukrainian, Feminine in Russian</b>	(a) Feminine	Copybook and a ballerina
	(b) Masculine	Copybook and a man

## Bibliography

Phillips, W., & Boroditsky, L. (2003). Can quirks of grammar affect the way you think? Grammatical gender and object concepts. In *Proceedings of the 25th annual cognitive science society* (pp. 928–933). Psychology Press.

Sera, Elieff, C., Forbes, J., Clark Burch, M., Rodriguez, W., & Poulin Dubois, D. (2002). When Language Affects Cognition and When It Does Not. *Journal of Experimental Psychology: General*, *131*(3), 377–397. <https://doi.org/10.1037/0096-3445.131.3.377>

Whorf, B. L. (1956). In J. B. Carroll (Ed.). *Language, thought, and reality: selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press



## Understanding time expression preferences and naming latencies from analog clocks: Insights from German and Czech speakers

Elena Panfilova, Barbara Mertins (TU Dortmund)

**Background** Several studies have found that naming times from analog clocks requires conceptual preparation (Meeuwissen et al., 2003; Korvorst, Roelofs, & Levelt, 2006). This process involves understanding abstract representations (e.g., the position of the hands) and linking them to specific time expressions (e.g., ‘half past three’, or ‘two thirty’). Korvorst et al. (2007) developed a model for Dutch time telling that identifies three stages of speech planning: conceptual preparation, lemma retrieval and syntax, and form encoding. The study revealed that relative expressions from both analog and digital clocks require conceptual preparation and form encoding. In contrast, absolute expressions from digital clocks require only form encoding, without the need for conceptual preparation. In the study by Korvorst et al. (2007), participants did not produce absolute time expressions when responding to analog clocks. Given the preference of Dutch speakers for relative time expressions, the researchers concluded that producing absolute time expressions might be challenging for Dutch speakers.

**Contribution** This research investigates how preferences for absolute (e.g. ‘two-twenty’) and relative (e.g. ‘twenty past two’) time expressions in German and Czech affect naming latencies from the analog clock. Specifically, we aim to determine whether it takes more or less time to respond to analog clock displays using absolute time expressions compared to relative ones.

To identify the preferences present in each language we collected time expressions through a paper-and-pencil elicitation task completed by German and Czech students (n=264). The outcomes indicate that Germans and Czechs predominantly employ relative time expressions. In the next step, we examined how these preferences affect naming latencies in time telling from analog clock displays. To address this research question, we conducted an eye-tracking experiment (n=80). During the experiment, participants named times from analog clocks in their native language. They did this without receiving specific instructions on which format of time expression to use. The Whisper V3 speech recognition system automatically detected speech onset latencies and transcribed the time expressions, which we manually verified and adjusted the acquired timestamps and transcripts using the EXMARaLDA transcription and annotating tool Partitur-Editor (Schmidt & Wörner, 2014).

Naming latencies provided evidence that speakers of Czech and German needed more time to produce absolute time expressions from analog clocks than relative expressions (650ms for German and 450ms for Czech). The analysis of first fixations in the eye-tracking data of both participant groups revealed the tendency to fixate the region of the minute hand first even in situations when speakers produced absolute expressions.

**Implications** The results highlight several important considerations. Firstly, it seems reasonable to suggest that an analog clock, serving as a stimulus, may provide a foundation for relative expressions. This is due to the perceptual context offered by the spatial geometry and motion direction of the clock hands, particularly for German and Czech speakers. Secondly, the preference for relative time expressions influences how speakers conceptualize time information from analog clocks contributing to longer naming latencies when producing absolute time expressions.

These findings contribute to our understanding of the cognitive processes involved in time perception and language production across different linguistic contexts. The evidence that absolute time expressions from analog clocks require more conceptual preparation than relative ones in both German and Czech supports and extends existing models of speech planning, such as the one proposed by Korvorst et al. (2007). This model can be refined to better account for the cognitive demands of different types of time expressions in various languages. Furthermore, insights from this research can contribute to educational practices, especially in teaching time-telling skills. Recognizing that analog clocks align

more with relative expressions can help educators develop strategies that leverage these cognitive preferences.

## **Bibliography**

- Bock, Irwin, Davidson & Levelt (2003). Minding the clock. *Journal of Memory and Language*, 48
- Korvorst, M., Roelofs, A., & Levelt, W. J. M. (2006). Incrementality in naming and reading complex numerals: Evidence from eyetracking. *Quarterly Journal of Experimental Psychology*, 59(2), 296–311. <https://doi.org/10.1080/17470210500151691>
- Korvorst, M., Roelofs, A., & Levelt, W. J. M. (2007). Telling Time from Analog and Digital Clocks. *Experimental Psychology*, 54(3), 187–191. <https://doi.org/10.1027/1618-3169.54.3.187>
- Meeuwissen, M., Roelofs, A., & Levelt, W. J. M. (2003). Planning levels in naming and reading complex numerals. *Memory & Cognition*, 31(8), 1238–1248. <https://doi.org/10.3758/BF03195807>

## Genericity on a scale: Offline processing of masculine generic role-nouns in Czech

Mikuláš Preininger<sup>1</sup>, James Brand<sup>2</sup>, Adam Kříž<sup>1</sup> (<sup>1</sup>Charles University, <sup>2</sup>Edge Hill University)

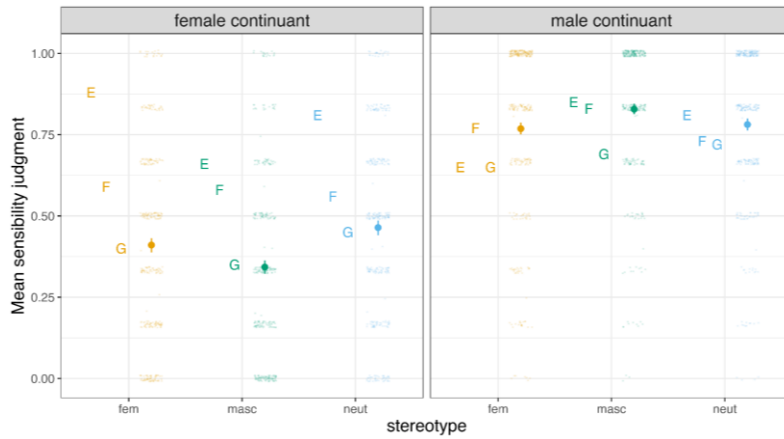
**Introduction** Masculine forms are often used generically, to refer not exclusively to males, but also to females. This has led some to critique the use of masculine generic forms as they might not represent males and females equally (for overview, see Sczesny & Formanowicz, 2016). In an anaphora resolution task, Gygax et al. (2008) have shown that with no previous context, masculine forms are interpreted not by default as generic, but instead as more likely to refer to males. Moreover, they showed that acceptability judgements of sentences where these forms referred to women differed between languages, being higher for French than for German, possibly due to a higher degree of gender marking in German. In this paper, we first replicate the Gygax et al. study in Czech to test whether the effects hold, or perhaps increase given the high degree of gender marking in Czech (Exp. 1). In Exp. 2, we address issues related to the coreference in the original stimuli and their overall naturalness, which might have confounded the results.

**Method** In Exp. 1, participants (n = 306) read pairs of sentences and their task was to assess whether the second sentence was a sensible continuation of the first one, in a binary choice task. The first sentence always included a role-noun in a masculine plural form (e.g. *novináři*, ‘journalists’). and the second sentence included a coreferring gendered noun (e.g. *někteří z mužů*, ‘some of the men’). Gender stereotypicality of the role-noun was manipulated (feminine/masculine/neutral), as well as grammatical gender of the anaphoric noun (feminine/masculine). We assumed that lower proportion of positive responses and higher response times in positive responses indicate anaphora resolution issues and suggest specific rather than generic interpretation of the role-noun. In Exp. 2, a new set of Czech stimuli was designed, pretested for coreference strength and translated into English, which served as a baseline. Czech (n = 98) and English participants (n = 43) judged the sentences in their L1 on a 5-point Likert scale (1 = not at all sensible, 5 = highly sensible).

**Results** In Exp. 1, we observed a significant effect of continuation and an interaction between continuation and stereotype in glmer models predicting sensibility, with the female and stereotype incongruent continuations being less sensible (see Fig. 1). For response times, we observed a significant effect of continuation with the positive ratings being made quicker in male continuations (see Fig. 2). In the Czech version of Exp. 2, we observed significant main effects of continuation and stereotypicality and the interaction of continuation and stereotypicality, again with the female and stereotype incongruent continuations being rated as less sensible, especially in stereotypically masculine conditions. In the English version, no significant effects were found (see Fig. 3).

**Discussion** The results show that when interpreting anaphora, Czech speakers rely on both gender stereotypicality of the antecedent as well as on its grammatical gender. Generically intended masculine forms favor a masculine interpretation, further supporting the claim that generic forms are not representing their referents equally. Exp. 1 not only corroborates the findings from Gygax et al., where for Czech we observed more pronounced effects, but also extends them by finding significant effects of stereotypicality and effects in response times, which were predicted but not confirmed in the original study. However, Exp. 2 shows that if the coherence of the two sentences increases and the way sensibility is measured changes, the magnitude of the effect reduces to minimum. In the future, we aim to test whether these effects hold even when using online processing methods (such as self-paced

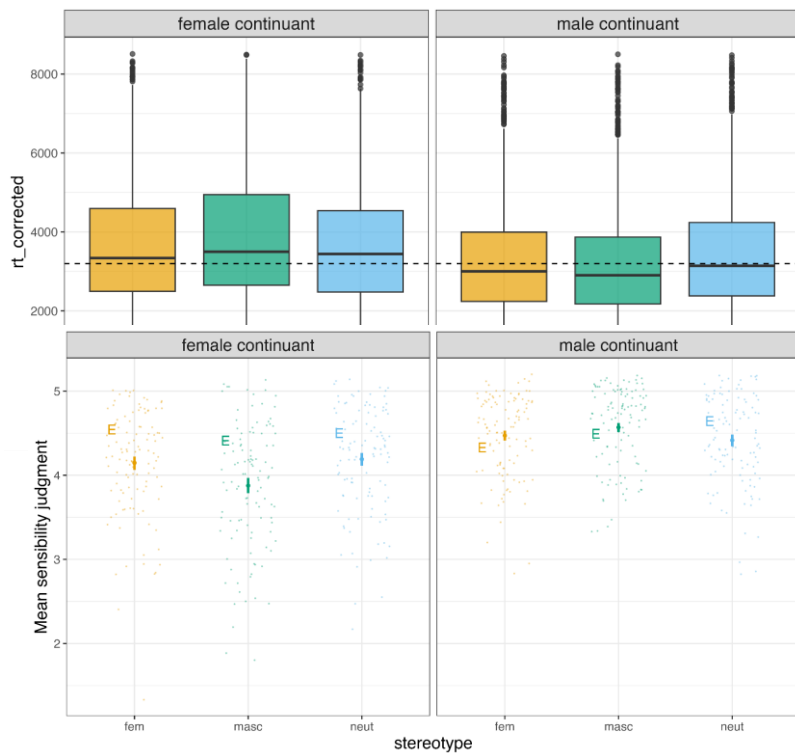
reading) and whether the results differ for participants from outside of a university sample who might be less sensitive towards gender inclusive language.



### Figures

**Figure 1.** Mean sensibility judgments in Exp. 1. Dots represent Czech data, the smaller ones for item means and the bigger ones for condition means. Letters represent condition means from Gygax et al. (2008), with E for English, F for French and G for German. The plot is faceted by the gender in the continuation. The colors represent gender stereotypicality of the role-noun (feminine in orange, masculine in green, neutral in blue).

neutral in blue).



**Figure 2.** Response times in Exp. 1. Each box represents response times in milliseconds for each condition. The plot is faceted by the gender in the continuation. The colors represent gender stereotypicality of the role-noun (feminine in orange, masculine in green, neutral in blue). The dashed line represents median response time across all conditions.

neutral in blue).

**Figure 3.** Mean sensibility judgments in Exp. 2. Dots represent Czech data, the smaller ones for item means and the bigger ones for condition means. Letters represent condition means from the English version of the experiment. The plot is faceted by the gender in the continuation. The colors represent gender stereotypicality of the role-noun (feminine in orange, masculine in green, neutral in blue).

### Bibliography

Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and cognitive processes*, 23(3), 464-485.

Sczesny, S., Formanowicz, M., & Moser, F. (2016). Can gender-fair language reduce gender stereotyping and discrimination?. *Frontiers in psychology*, 7, 154379.

## How do people understand explicit and implicit contrastive and temporal discourse relations in Czech?

Šárka Zikánová (Charles University, Faculty of Mathematics and Physics), Filip Smolík (Czech Academy of Sciences, Institute of Psychology & Charles University, Faculty of Arts)

The Facilitator Connective hypothesis (Grisot et al., 2022; Köhne-Fuetterer et al., 2021) assumes that discourse connectives enable readers to interpret easily discourse relations between upcoming discourse arguments. It follows from the hypothesis that implicit discourse relations (pairs of discourse arguments without discourses connectives) should demand a bigger effort from readers for the proper understanding. We tested the predictions of the hypothesis in two self-paced reading experiments that examined the processing of four distinct discourse relations in Czech, marked by explicit connectives or left implicit.

Experiment 1 examined the contrastive relations of concession and confrontation marked by connectives *přesto* “nevertheless/still” and *zato* “but, on the contrary”. Experiment 2 examined temporal relations marked by connectives *mezitím* “meanwhile” (synchrony) and *potom* “then” (asynchrony). The current concept of discourse semantic types was adopted from the Prague Discourse Treebank approach (Zikánová et al., 2015).

Participants (N=94 and 105, respectively) read four-sentence passages in which the second and third sentences were related by the discourse relations under investigation. Each participant read 20 concessive, 20 confrontation and 40 filler items, resp. 20 synchronous, 20 asynchronous and 40 filler items. The experiments were presented online via IbexFarm. The regions shown as one unit in the self-paced reading are separated by / below:

- (1) Odjakživa nenáviděl / rozbouřené moře.

Always he\_had\_hated<sub>3SG. PAST MASC.</sub> stormy<sub>ACC. SG.</sub> sea<sub>ACC. SG.</sub>.

*He had always hated stormy seas.*

**[Přesto]** / vloni / na jachtě obeplul / celý svět.

Nevertheless last\_year on\_a\_yacht he\_sailed\_around<sub>3SG. PAST MASC. SG.</sub> whole<sub>ACC. SG.</sub> world<sub>ACC. SG.</sub>.

*Nevertheless, he sailed around the world last year.*

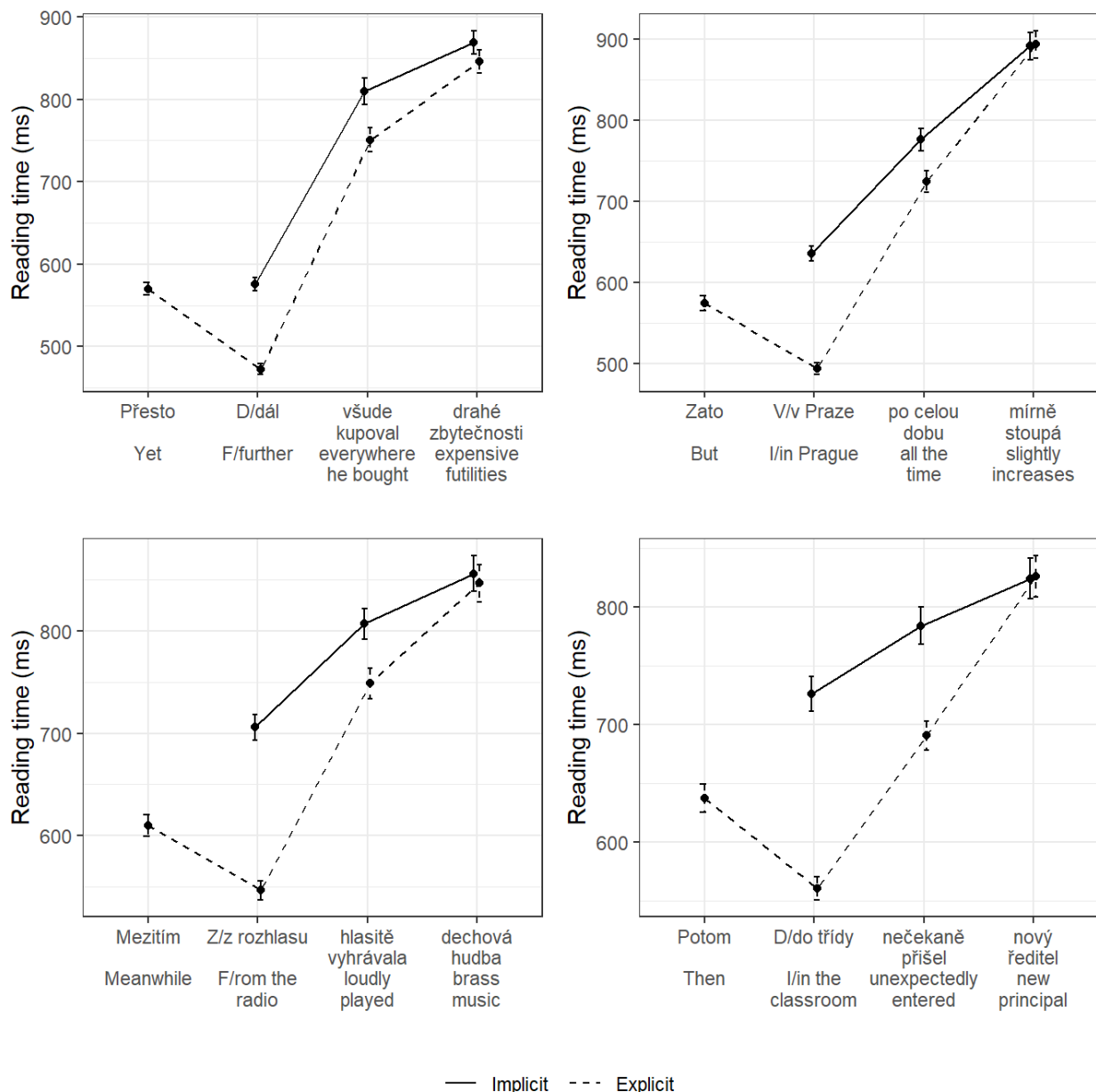
According to the Facilitator Connective hypothesis, we expected that the absence of explicit discourse relation marking would result in slower reading times in the third sentence of the passage. The results were analyzed using mixed models with random effects for persons and items, separately for each region of the target sentence.

The reading times were slower on the first and the second region of the target sentences in the implicit condition, compared to the first and second region after the connective. No effects were observed in the sentence-final regions. The pattern was consistent across all four types of structures. These findings provide strong evidence for the claim that the absence of a discourse connective makes sentences harder to process. On the other hand, we see little evidence for any differences between the structures, although some previous work suggested that this could be the case (Evers-Vermeul – Sanders, 2009; Köhne-Fuetterer et al., 2021): according to the Cognitive Approach to Coherence Relations (CCR), discourse

relations with negative polarity (e.g., contrastive relations) are more complex for processing than relations with positive polarity (e.g., temporal relations).



Figure 1: Reading times for the target sentence segments in all types of structures.



## Bibliography

Evers-Vermeul, J., Sanders, T. (2009). The emergence of Dutch connectives: How cumulative cognitive complexity explains the order of acquisition. *Journal of Child Language*, 36(4), 829–854. <https://doi.org/10.1017/S0305000908009227>

Grisot, C., Blochowiak, J., Degand, L. (2022). From implicit to explicit: The processing of forward causal and temporal relations. *Pragmatics and Cognition*, 29(1), 30–60.

Köhne-Fuetterer, J., Drenhaus, H., Delogu, F., Demberg, V. (2021). The online processing of causal and concessive discourse connectives. *Linguistics*, 59(2), 417–448. <https://doi.org/10.1515/ling-2021-0011>

Zikánová, Š., Hajičová, E., Hladká B., Jínová, P., Mírovský, J., Nedoluzhko, A., Poláková, L., Rysová, K., Rysová, M., Václ, J. (2015). *Discourse and Coherence. From the Sentence Structure to Relations in Text*. Prague: Charles University, Faculty of Mathematics and Physics.

3

*Metody korpusové analýzy  
diskurzu v češtině*

*Methods of corpus-based  
discourse studies  
in the Czech context*

organizují / organizers

Václav Cvrček

Ondřej Dufek

Michal Hořejší

(CZE, ENG)

Bienále české lingvistiky 2024  
Biennial of Czech Linguistics 2024

## Popis panelu / Workshop description

**(CZE)** Korpusové přístupy se v analýze diskurzu uplatňují už více než 20 let. Tyto metody využívají rozsáhlých textových korpusů k odhalování opakujících se vzorců v textech, diskurzního rámování pojmů a k poznání dynamiky diskurzu (Partington & Duguid, 2013). Tento konferenční panel se zabývá aplikací těchto technik (a představením nových) na analýzu českých textů. Zaměřuje se především na parolové jevy, tj. na analýzu konkrétních textů či jejich souborů, nikoli na popis systémových funkcí nadvětných jazykových jevů, pro něž se zejména v NLP rovněž vžilo označení diskurzni analýza. Panel bude klást důraz spíše na metodologické přístupy a jejich aplikace na texty, neklade si tedy konkrétní tematická omezení; cílem je poskytnout prostor pro prezentaci metod, nástrojů, aplikací i případových studií v oblasti diskurzni analýzy českých textů.

Přístupy užívané v korpusových analýzách diskurzu zahrnují různé metody, např. frekvenční analýzu, kolokační analýzu či identifikaci klíčových slov (Baker, 2006; McEnery & Hardie, 2012). Kromě toho se v analýze diskurzu uplatňují i další metody, které se běžně používají při práci s korpusy nebo v NLP experimentech (např. Cvrček & Fidler, 2022; Schweinberger et al., 2021). Zkoumání těchto jazykových jevů v textech nebo souborech textů pomáhá identifikovat klíčová témata, sdílené předsudky či rétorické strategie. Zásadní výhoda korpusových metod přitom spočívá v jejich schopnosti účinně vyvažovat výsledky kvalitativních analýz (srov. Baker et al., 2008). Zatímco kvalitativní analýza umožňuje hloubkové zkoumání rysů diskurzu, kvantitativní přístup může poskytnout empirické důkazy, statistickou přesnost a zmírnit výzkumný bias či jiné nežádoucí jevy (např. cherry-picking). Korpusové metody mají nadto schopnost odhalit dříve nepozorované jevy či netušené souvislosti.

Hlavní náplní tohoto konferenčního panelu je diskuse o celém spektru korpusových metod, které lze efektivně aplikovat na analýzu českých textů (včetně těch, které jsou přizpůsobeny specifickým rysům češtiny nebo českého diskurzu). Panel nebude tematicky omezen, což účastníkům umožní získat povědomí o širokém spektru metod použitelných v různých oblastech výzkumu.

Výzkumné otázky, na něž se panel zaměřuje:

- měření prominence jednotek (keyness)
- korpusová analýza, diskurzy a reprezentativnost
- identifikace (vzdálených) asociací nebo diskurzivních rámců
- aplikace/nástroje pro korpusovou analýzu diskurzu
- metody zohledňující specifické rysy češtiny (např. bohatá morfologie)
- triangulace a kombinování metod

**(ENG)** Corpus-based approaches have more than 20 years long tradition in discourse analysis. These methods harness the power of large, systematically collected text corpora to unveil hidden patterns in texts, uncover discourse framing, and gain insights into discourse dynamics (Partington & Duguid, 2013). This conference panel delves into the application of these techniques (and introduction of new ones) to the analysis of Czech texts. The focus is primarily on parole phenomena, i.e. analysis of particular texts or their collections, not on description of systemic functions of language elements surpassing sentence level which is sometimes also referred to as discourse level namely within NLP. The panel will emphasize methodological approaches and their applications on texts rather than specific thematic constraints, thus providing a platform for presentation of methodologies, tools, applications as well as case-studies in discourse analysis of Czech texts.

Approaches used in corpus-based discourse studies encompass various techniques, such as frequency analysis, collocation analysis, and the identification of keywords (Baker, 2006; McEnery & Hardie, 2012). Furthermore, several other tools commonly used in corpus analyses in general or in NLP experiments find their application in studies of public discourse (e.g. Cvrček & Fidler, 2022; Schweinberger et al., 2021). Examination of linguistic phenomena in texts or text collections by means of these methods can shed light on underlying themes, biases, and rhetorical strategies. The key advantages of corpus-based methods lies in their ability to complement qualitative findings (cf. Baker et al., 2008). While qualitative analysis allows for in-depth exploration of discourse features, quantitative approach can provide empirical evidence, statistical rigor and mitigate research biases or cherry-picking. Moreover, corpus methods have the remarkable ability to uncover previously unseen phenomena or relations within a large body of text.

The primary focus of this conference panel is to discuss a wide range of corpus-based methods that can be effectively applied to the analysis of Czech texts (including those which are custom-tailored to the specific features of Czech or Czech discourse). The panel will not be constrained by specific thematic limitations, allowing attendees to gain an awareness of a wide range of tools and methods applicable to diverse research areas.

Research questions addressed by the panel:

- measuring prominent units (keyness)
- corpus analysis, discourses and representativeness
- identifying (distant) associations or discursive frames
- applications/tools for corpus-assisted discourse analysis
- methods anchored in specific features of Czech (e.g. rich morphology)
- triangulation and methodology blending

### **Literatura / References**

Alan Partington, Alison Duguid, C. T. (2013). *Patterns and Meanings in Discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. John Benjamins Publishing Company.

Baker, P. (2006). *Using Corpora in Discourse Analysis*. Continuum.

Baker, P., Gabrielatos, C., Khosravini, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society*, 19(3), 273–306.

Cvrček, V., & Fidler, M. (2022). No Keyword is an Island: In search of covert associations. *Corpora*, 17(2), 259–290.

McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.

Schweinberger, M., Haugh, M., & Hames, S. (2021). Analysing discourse around COVID-19 in the Australian Twittersphere: A real-time corpus-based analysis. *Big Data & Society*, 8(1), 1–17.

## Obsah / Table of Contents

*Sibling-texts keyword analysis: exploring topic and register keywords*

Martina Berrocal (Friedrich Schiller University Jena)

*Methodological challenges of corpus-assisted discourse analysis of anti-system news portals*

Václav Cvrček (Charles University, Faculty of Arts), Masako Fidler (Brown University)

*Thematic concentration vs keyness: A comparison of methods and their relevance for discourse analysis*

Ondřej Dufek (Charles University)

*Discursive representations of nationalities in Czech news media*

Irene Elmerot (Stockholm University)

*Divočina v českém veřejném diskurzu. Korpusová analýza kulturního konceptu*

Michal Hořejší (Ústav českého jazyka a teorie komunikace FF UK)

*Serbian (de)verbal nouns ending in –nje as a peculiarity of some formal registers in public discourse of the internet*

Borko Kovačević (University of Belgrade – Faculty of Philology)

**Sibling-texts keyword analysis: exploring topic and register keywords**

Martina Berrocal (Friedrich Schiller University Jena)

Over the last few decades, keywords (KWs) have emerged as one of the most influential concepts in corpus linguistics. It has been noted multiple times [e.g., 1, pp. 347; 2, pp. 97–98; 5, pp. 43; 6, pp. 354] that KWs can be associated with the topic/theme of the text or its register, genre, or style. This paper aims to present a method that facilitates the distinction between these two types of keywords: those related to the topic and those related to the genre/register. This method is based on comparing the results of multiple parallel keyword analyses (KWAs) performed on texts that share similar genre and register characteristics. These texts are referred to as 'sibling texts', thus termed sibling-texts keyword analysis. The identification of these texts is aided by multi-dimensional analysis (MDA) [4].

To test the proposed method on real data, we apply it to texts from the Parlcop corpus [3], which compiles transcripts of speeches delivered in the Czech Lower Parliamentary Chamber (1993–2021). The results demonstrate the method's ability to distinguish between register-related and topic-related keywords, even in highly homogeneous corpora. This differentiation holds significant potential for advancement in various research domains. For instance, register-related keywords have proven valuable in register studies, for example, by enhancing the interpretation of parliamentary subregisters (administrative & procedural, deliberating & debating, and government policies), as demonstrated in the present study. Furthermore, it facilitates a more nuanced approach to keyword analysis by refining the focus of topic-oriented discourse studies. This refinement is achieved by eliminating the 'noise' caused by register-related keywords, which, in such studies, may distort the results and their interpretation.

**Bibliography**

- [1] Baker, Paul. 2004. Querying keywords: Questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics* 32(4), 346–359.  
<https://doi.org/10.1177/0075424204269894>
- [2] Baker, Paul, Andrew Hardie, and Tony McEnery. 2006. *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.
- [3] Berrocal, Martina, and Manuel Berrocal. 2021. *ParlCorp: Corpus of Czech Parliamentary speeches (Version 1)* [Computer software]. Institute of Czech National Corpus, Faculty of Arts, Charles University. [www.korpus.cz](http://www.korpus.cz)
- [4] Cvrček, Václav, Zuzana Komrsková, David Lukeš, Petra Poukarová, Anna Řehořková, and Adrian Jan Zasina. 2021. From extra- to intratextual characteristics: Charting the space of variation in Czech through MDA. *Corpus Linguistics and Linguistic Theory* 17(2), 351–382.
- [5] Scott, Mike. 2010. Problems in investigating keyness, or clearing the undergrowth and marking out trails.... In Marina Bondi and Mike Scott (eds.), *Keyness in texts*, 43–58. Amsterdam: John Benjamins.
- [6] Stefanowitsch, Anatol. 2020. *Corpus linguistics: A guide to the methodology*. 1st ed. Berlin: Language Science Press.

### Methodological challenges of corpus-assisted discourse analysis of anti-system news portals

Václav Cvrček (Charles University, Faculty of Arts), Masako Fidler (Brown University)

The main benefit of bringing corpora into discourse analysis (CADS, Baker & McEnery, 2005; Partington, 2004) is the ability to quantify linguistic features and thus generalize qualitative hypotheses based on close reading of small number of texts to the whole discourse. The inventory of CADS methods is relatively stable and the methodological innovations mostly focus at refinement of procedures (cf. Egbert & Biber, 2019; Gabrielatos & Marchi, 2012). This paper proposes potentially innovative corpus-based study of associations, a hither-to-overlooked aspect of discourse that could play a key role in comparing mainstream (MS) anti-system (ANTS) media.

Keywords (KWs) (Culpeper & Demmen, 2015; Scott & Tribble, 2006), a commonly used starting point in discourse analysis, are defined as units with a strikingly higher relative frequency in the target text compared to that in a reference corpus. They indicate prominent topics of a given text/discourse (e.g. Baker et al., 2013; Tabbert, 2015; Walker, 2010).

The KWs in the ANTS reveals two basic findings: (1) some point to obvious differences in the thematic preferences (e.g. Pope Francis as a heretic, global conspiracy, alleged Russophobia in the EU); (2) most topics represented by the ANTS KWs appear to be shared by MS. The KWs, however, do not inform of the fundamental differences in the description of such recurrent topics. This is the lacuna that we address via probing associated units.

We present two methods for association research in corpora. By associations we mean the regular co-occurrence of lexical units that do not have to be in a syntagmatic relation (i.e., no necessarily in proximity to each other). One of them is the Market Basket Analysis (MBA), a data-mining technique applied to texts, and the other Companions.

Both methods are applied to data from the ONLINE corpus (Cvrček et al., 2022; Cvrček & Procházka, 2020) that covers a wide range of Czech news portals and classified on the basis of audience overlap between the portals (Šlerka, 2018).

MBA (Cvrček & Fidler, 2022) identifies the co-occurrence of keywords in texts. Analogous to its original use in marketing, where their goal was to identify items in a shopping cart that are frequently purchased together, it assesses the strength of relationships between all KWs in the corpus texts and finds those pairs, triplets of quads of keywords that regularly occur together in the texts. A group of KWs co-occurring with a word can be considered its associative frame, which characterizes the type of text where the word occurs and contributes significantly to the overall reception of the topic in the texts (e.g. *the EU* is embedded in texts about the types of negotiations and/or their actors–politicians in MS, while the same word appears in texts about the EU sanctions against Russia, loss of national identity, dangerous migrants, impoverishment of member states in ANTS).

Companions looks at the frequency patterns of two words over a stretch of time using correlation (Cvrček & Fidler, in print). If two words show the same (or very similar) trends, we can assume they are associated. For example, ANTS, unlike MS, shows similar frequency trends for *Covid* or *coronavirus* and (simple) *flu* in the first phase of the pandemic; ANTS thereby consistently underestimated covid, justifying its criticism of the government action.

Both MBA and Companions empower CADS by informing more on the relationships between lexemes that occur over a greater distance (e.g., compared to collocations). They thus allow us to grasp the nature of the entire text where KWs occur.

## Bibliografie

- Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press*. Cambridge University Press.
- Baker, P., & McEnery, T. (2005). A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Journal of Language and Politics*, 4(2), 197–226.
- Culpeper, J., & Demmen, J. (2015). Keywords. In D. Biber & R. Reppen (Eds.), *The Cambridge Handbook of English Corpus Linguistics* (pp. 90–105). Cambridge University Press.
- Cvrček, V., & Fidler, M. (in print). From News to Disinformation: Unpacking a Parasitic Discursive Practice of Czech Pro-Kremlin Media. *Scando-Slavica*.
- Cvrček, V., & Fidler, M. (2022). No Keyword is an Island: In search of covert associations. *Corpora*, 17(2), 259–290.
- Cvrček, V., Jeziorský, T., & Henyš, J. (2022). *ONLINE2: Monitorovací korpus internetové češtiny*. Institute of the Czech National Corpus. FF UK. [www.korpus.cz](http://www.korpus.cz)
- Cvrček, V., & Procházka, P. (2020). *ONLINE: monitorovací korpus internetové češtiny*. Institute of the Czech National Corpus. FF UK. [www.korpus.cz](http://www.korpus.cz)
- Egbert, J., & Biber, D. E. (2019). Incorporating text dispersion into keyword analyses. *Corpora*, 14(1), 77–104.
- Gabrielatos, C., & Marchi, A. (2012, September 14). *Keyness: Appropriate metrics and practical issues*. CADS International 2012. Corpus-assisted Discourse Studies: More than the sum of Discourse Analysis and computing?, University of Bologna, Italy.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3 edition). Morgan Kaufmann.
- Miner, G. (2012). *Practical text mining and statistical analysis for non-structured text data applications.: Vol. 1st ed.* Academic Press.
- Partington, A. (2004). Corpora and discourse, a most congruous beast. In A. Partington, J. Morley, & L. Haarman (Eds.), *Corpora and discourse* (pp. 11–20). Peter Lang.
- Scott, M., & Tribble, C. (2006). *Textual Patterns: Key words and corpus analysis in language education*. John Benjamins.
- Šlerka, J. (2018). *Typologie domácích zpravodajských webů*. Nadační fond nezávislé žurnalistiky. <https://www.nfnz.cz/studie-a-analyzy/typologie-domacich-zpravodajskych-webu/>
- Tabbert, U. (2015). Crime and Corpus. In *Lal.20*. John Benjamins.
- Walker, B. (2010). Wmatrix, key-concepts and the narrators in Julian Barnes' Talking It Over. In D. McIntyre & B. Busse (Eds.), *Language and Style* (pp. 364–387). Palgrave MacMillan.



**Thematic concentration vs keyness:  
A comparison of methods and their relevance for discourse analysis**

Ondřej Dufek (Charles University)

Discourse analysis using big data and automatized, replicable procedures benefits productively from methods that point to prominent characteristics of a text, corpus or discourse. Such methods may represent one of the gateways into the target discourse, especially in corpus-driven approaches, because they enable to free oneself from preassumptions and may draw attention to what one could not think about. They constitute a useful path towards notification of significant concepts in the discourse.

This study focuses on the comparison of two of these methods: commonly used keyness or key words analysis (KW) on one hand and rarely used thematic concentration of text or thematic words analysis (TC) on the other hand.

“[K]eyness is a quality words may have in a given text or set of texts, suggesting that they are important, they reflect what the text is really about, avoiding trivia and insignificant detail” (Scott – Tribble, 2006, s. 55–56). For discourse analysis, key words analysis was introduced by Scott (1997); it can “direct the researcher to important concepts in a text (in relation to other texts) that may help to highlight the existence of types of (embedded) discourse or ideology” (Baker, 2004, s. 347).

Thematic concentration also – in a way – highlights linguistic items that are in the target data important (frequent) beyond expectation. Its primary aim is to “capture to what extent the author concentrates on a given theme in the text” (Čech, 2016, s. 9, translation mine). Basically, it is grounded in frequency characteristics of a text and its h-point, inspired by Hirsch index. It was introduced by Popescu (2007), and comprehensively addressed by Čech (2016, also particularly in relation to KW analysis: pp. 101-110).

This study’s aim is to compare these two methods regarding their usability in discourse analysis on two levels: 1) theoretical and 2) empirical. The analysis makes use of two apps by Czech National Corpus: KWords (<https://kwords.korpus.cz/>; Horký Vondříčka Cvrček 2023) and QuitaUP (<https://korpus.cz/quitaup/>); Cvrček Čech Kubát 2020).

Using one single text, various parameters of both methods are examined regarding their influence onto the results – in KW analysis, different outputs are gained regarding the statistical test, effect size measure, minimal frequency, significance threshold and, of course, reference corpus. In TC analysis, the effect of primary and secondary TC is observed.

As a “reference” perspective or base line for assessing how to consider what method gives what outputs, a human-conducted qualitative analysis of the target text is used. To mitigate the influence of subjectivity, the qualitative analysis is carried out by three independent analysts – one of them is the author of this study, another one is an independent experienced researcher in the field of corpus-based discourse studies, and yet another one is linguist with no substantial practice in the corpus-based discourse studies. All of them were instructed just to identify what autosemantic words (lemmas) represent the text’s nature, most important concepts, in other words, what the text is about. All of them were also asked to comment on why they chose particular words and what based their decisions.

Results of the study show what method comes closer to the human reading of the text. It is considered what each of them (and under what parameters) mean for the purposes of discourse analysis. The study ends with suggestions towards use and interpretation of KW and TC analysis.

## Bibliography

Baker, P. (2004). Querying keywords: questions of difference, frequency, and sense in keyword analysis. *Journal of English Linguistics*, 32(4), 346-359.

Cvrček, V., Čech, R., & Kubát, M. (2020). *QuitaUp – nástroj pro kvantitativní stylometrickou analýzu* [online]. Praha: Ústav Českého národního korpusu FF UK – Ostrava: Ostravská univerzita. Cit. 19. 5. 2023. <<https://korpus.cz/quitaup/>>.

Čech, R. (2016). *Tematická koncentrace textu v češtině*. Praha: Ústav formální a aplikované lingvistiky MFF UK v Praze.

Horký, V., Vondříčka, P., & Cvrček, V. (2023). *KWords* [online]. Praha: Ústav Českého národního korpusu FF UK. Cit. 19. 5. 2023. <<http://kwords.korpus.cz>>.

Popescu, I.-I. (2007). Text ranking by the weight of highly frequent words. In: P. Grzybek & R. Köhler (Eds.), *Exact Methods in the Study of Language and Text*. Berlin: Mouton de Gruyter, pp. 557-567.

Scott, M. (1997). PC analysis of key words – and key key words. *System*, 25(2), pp. 233-245.

Scott, M., & Tribble, Ch. (2006). *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Philadelphia: John Benjamins.

**Discursive representations of nationalities in Czech news media**

Irene Elmerot (Stockholm University)

When humans often, or even consistently, read that certain groups of people act or are in a certain way, we tend to think that there at least may be a grain of truth in that representation. As argued many times during the past century, the things we repeatedly read or hear turn first into a discourse, and sometimes into matters-of-fact (Šklovsky 1971[1925]; Van Dijk 1988: 147; Fairclough 2006:3). Discourses, however fluid they may be, may help us understand and make meaning of our world (Filardo-Llamas and Morales-López 2021, Reisigl and Wodak 2016: 27), but if it is a biased discourse, our picture of the world may also become biased. This presentation will show analytical results created with different tools, from different kinds of Czech public discourses.

Researchers of the Czech language have access to longitudinal corpora, such as the synchronous corpus series with journalistic text, domestic and European parliamentary debates, and fiction, all within the Czech National Corpus ([www.korpus.cz](http://www.korpus.cz)). This presentation aims to examine how nations and nationalities are reflected in different discourses from some of these sources; Czech news papers, tv news shows and some other news-related media, mainly in the 21st century. Recent research has delved deeper into this than previously, and showed that different sources may include very different discourses on both the own and other nationalities (e.g. Cvrček and Fidler 2022, Elmerot 2019; Berrocal 2017).

Previous research (Elmerot 2021, 2022; Pekáček and Elmerot 2023) has shown that the Czech media represent its own country as one among its Western European neighbours, and they clearly create others out of low-income nationalities and especially nationalities of Arab origin or where the Muslim faith is dominant. In the latter study, we also see how the representations of Ukraine and ukrainians have changed in the last decade, due to geopolitical actions. This workshop presentation will first discuss results from these previous studies, as well as works in progress (see graph below), showing how Czechs seem to view themselves and others in media, and then focus on discussing how we best may reveal the discourses that may turn into matters-of-course. Is there an attractive alternative to the black boxes of chatbots?

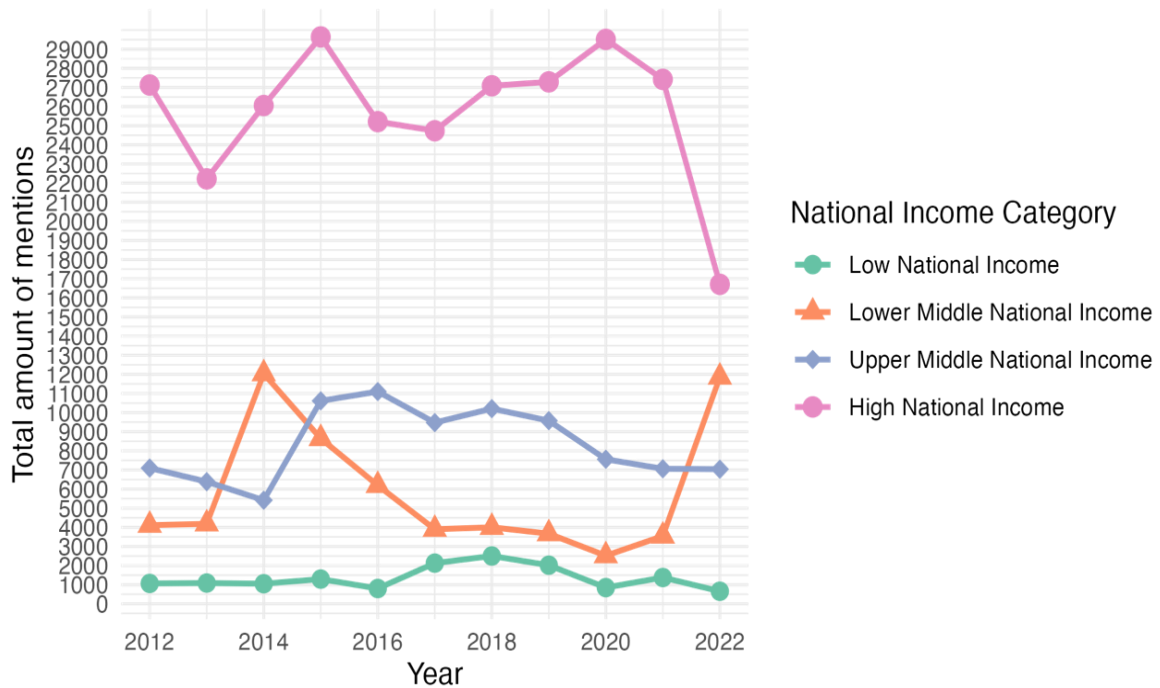


Figure 3: Raw frequencies of country and territory nouns per national income group in Czech news television 2012–2022. Source: Culková & Elmerot forthcoming.

## Bibliography

- Berrocal, Martina. 2017. 'Proximizing the Ukraine Conflict: The Case of the United States and the Czech Republic'. *Lodz Papers in Pragmatics* 13 (2). [DOI](#).
- Cvrček, Václav, and Masako Fidler. 2022. 'No Keyword Is an Island: In Search of Covert Associations'. *Corpora* 17 (2): 259–90. [DOI](#).
- Dijk, Teun Adrianus van. 1988. *News as Discourse*. Communication. Hillsdale (NJ): L. Erlbaum associates.
- Elmerot, Irene. 2019. 'Dirtbags, Drunkards and Miniature Mutes: Czech Subjectivity Revealed Through Corpus-Based Discourse Analysis'. *Scando-Slavica* 65 (2): 127–45. [DOI](#).
- . 2021. 'Income, Nationality and Subjectivity in Media Text'. *Jazykovedný Časopis/Journal of Linguistics* 72 (2): 667–78. [DOI](#).
- . 2022. 'Constructing "us and Them" through Conflicts – Muslims and Arabs in the News 1990–2018'. In *Discursive Approaches to Socio-Political Polarization and Conflict*, edited by Laura Filardo Llamas, Esperanza Morales López, and Alan Floyd, 1st ed., 122–36. Routledge Research in Language and Communication 11. New York; Oxon: Routledge. [DOI](#).
- Fairclough, Norman. 2006. *Language and Globalization*. Abingdon: Routledge.
- Filardo-Llamas, Laura, and Esperanza Morales-López. 2021. 'Introduction'. In *Discursive Approaches to Sociopolitical Polarization and Conflict*, by Laura Filardo-Llamas, Esperanza Morales-López, and Alan Floyd, 1st ed., 1–22. New York: Routledge. [DOI](#).
- Pekáček, Ondřej, and Irene Elmerot. 2023. 'When Is a Crisis Really a Crisis? Using NLP and Corpus Linguistic Methods to Reveal Differences in Migration Discourse across Czech Media'. *Jazykovedný Časopis/Journal of Linguistics* 74 (1): 369–80. [DOI](#).
- Reisigl, Martin, and Ruth Wodak. 2016. 'The Discourse-Historical Approach (DHA)'. In *Methods of Critical Discourse Studies*, edited by Ruth Wodak and Michael Meyer, 23–61. Sage: Sage.
- Šklovskij, Viktor. 1925. *O teorii prozy*. Moskow: Krug.

**Divočina v českém veřejném diskurzu. Korpusová analýza kulturního konceptu**

Michal Hořejší (Ústav českého jazyka a teorie komunikace FF UK)

Cílem mého výzkumu je rekonstruovat, jakým způsobem se ve středoevropském a konkrétně českém prostředí historicky etabloval a dlouhodobě provozuje koncept divočiny. Smyslem (budoucího) výzkumu je na základě analýzy dlouhodobého provozu konceptu vytvořit základ pro interpretaci konfliktů v nastavování recentní environmentální politiky v ČR potažmo Evropské unii, které jsou zapříčiněny napětím mezi ideologií *re-wilding* a ideologií *nature restoration* (srov. Bastmeijer 2016, Büscher – Fletcher 2019) a které se opakovaně stávají bází pro konflikty komplexnější povahy (Forchtner 2020). Je přitom pozoruhodné, že termín *divočina/wilderness* do evropského práva přineslo české předsednictví EU v roce 2009 (Bastmeijer 2016).

Chci otestovat tradovanou představu vycházející především z angloamerické literatury, podle níž je slovo *divočina* ve významu ‚divoká příroda‘ součástí teprve lexikonu moderního světa a je předně vymezeno esteticky (Nash 2014 [1967], Stankey 1989, Oelschlaeger 1991, Buell 1995, Cronon 1996). Ač této představě odpovídá i rešerše českého kontextu, v němž se slovo *divočina* coby pojmenování přírodního prostoru sporadicky nachází až v literatuře počátku 19. století, frekventovaněji pak v beletrii druhé poloviny 19. století, je zřejmé, že divokou přírodu, resp. opuštěné přírodní prostory vně obývaného prostoru jazyk reflektuje mnohem déle a zhusta prostřednictvím jiných slov či slovních spojení než právě prostřednictvím slova *divočina*. Podobně zřejmé je, že variabilitu v čase a prostoru vykazuje také sama sémantika slova. Argumentem pro to je například ve staré češtině (tj. v českojazyčných pramenech před rokem 1500) hojně využívání lexému *poušť* v biblických textech na místech, kde angličtina téže doby užívá *wilderness*, nebo analýza nedávného diskurzu o Šumavě, v němž lexém *divočina* převážně nenesl význam ‚místa krásné či nedotčené přírody‘, nýbrž ‚místa ponechaného přírodním procesům‘ (Hořejší 2022, srov. též Krajhanzl et al. 2015). Samo slovo tudíž nejen že nepokrývá historii provozu konceptu ‚divoké přírody‘, ale je i významově ambivalentní. Ptám-li se proto, jaká je historie toho, co nazývám „konceptem divočiny“, musím nejen vykročit z hranic slova, ale i z hranic představy, která koncept divočiny umisťuje na přelom 18. a 19. století a přisuzuje mu primárně estetické rámování.

Těžištěm příspěvku proto bude především metoda vytvoření samotného korpusu, resp. otázka jeho reprezentativnosti (Biber 1993, Baker 2006, Chromý 2014). Textová populace, jejímž je můj korpus vzorkem, totiž nejen že musí obsahovat materiál z více než třech století a různých – v čase proměnlivých – žánrů, zároveň však nesmí být vymezena prostými klíčovými slovy. Příspěvek představí kvalitativní sondy do materiálu a navrhne design finálního korpusu.

## Bibliografie

- Baker, P. (2006): *Using Corpora in Discourse Analysis*. London: Continuum
- Bastmeijer, K. (2016). *Wilderness protection in Europe: The role of international, European and national law*. Cambridge: Cambridge University Press
- Biber, Douglas (1993): Representativeness in Corpus Design. *Literary and Linguistic Computing* 8, s. 243–257
- Buell, L. (1995). *The environmental imagination: Thoreau, nature writing, and the formation of American culture*. Cambridge: The Belknap Press
- Büscher, B. – Fletcher, R. (2019). Towards Convivial Conservation. *Conservation & Society* 17 (3), s. 283–296
- Cronon, W. (1996). The Trouble with Wilderness: Or, Getting Back to the Wrong Nature. *Environmental History* 1 (1), s. 7–28
- Forchtner, B., (ed.) (2020). *The far right and the environment: politics, discourse and communication*. London: Routledge, Taylor & Francis
- Hořejší, M. (2022): *Politika lesa*. Praha: FF UK
- Chromý, J. (2014). Korpus a reprezentativnost. *Naše řeč* 97 (4–5), s. 185–193
- Krajhanzl, J. – Skalík, J. – Špaček, O. – Chabada, T. – Čada, K. – Lechnerová, Z. – Svozdová, R. (2015): *Ochrana divoké přírody očima české veřejnosti*. Brno: KES FSS MU.
- Nash, R. (2014). *Wilderness and the american mind*. Fifth edition. New Haven: Yale University Press
- Oelschlaeger, M. (1991). *The Idea of Wilderness. From Prehistory to the Age of Ecology*. New Haven & London: Yale University Press
- Stankey, G. H. (1989). Beyond the Campfire's Light: Historical Roots of the Wilderness Concept *Natural Resources Journal* 29 (1), s. 9–24

**Serbian (de)verbal nouns ending in *-nje* as a peculiarity of some formal registers  
in public discourse of the internet**

Borko Kovačević (University of Belgrade – Faculty of Philology)

Deverbal nouns ending in *-nje* are the most frequent tool for deriving nominalizations in Serbian. So, the sentence *Oni su se prijavili* (They applied) would be usually nominalized in the form of the next nominal phrase: *njihovo prijavljivanje* (their applying/application), with the noun *prijavljivanje* as the center of the nominalization. Lot of research has been done in the Serbian linguistics on showing the fact that the use of nominalizations is a characteristic of an extremely formal way of language use, like the language of bureaucracy, journalistic, political, or scientific style (Radovanović 1990, Klikovac 2008). The most part of the previous research was focused on nouns ending in *-nje* as the main type of nominalizations in Serbian.

The aim of the paper is to test the statement that the use of deverbal nouns ending in *-nje* is a characteristic of some formal registers, as well as to extract some additional markers of the use of these nouns in formal registers, more specifically in formal registers in public discourse of the internet. Particularly, the modification of nouns ending in *-nje* by adverbs and by the phrase *od strane* (equivalent to the English *by* phrase) is analyzed. This has been done by working on the greatest and most modern contemporary corpus of the Serbian language (*Serbian Web Corpus PDRS 1.0*), which is the only one with tools for automatic search in the text types/areas (*media, inform, shop, forum, state, science, company, ngo, portal, database, community*) that more or less correspond to different registers (in the first place regarding the opposition between formal and informal registers). In doing this search two parameters of the use of the two modifiers were measured – relative frequency (frequency per million tokens) and relative density (relative text type frequency).

The research has been done in the PDRS (*Public discourse in the Republic of Serbia*) project, supported by the Science Fund of the Republic of Serbia. A part of the project is extraction and analysis of the grammatical characteristics of the public discourse within different registers, where by register is meant “a variety associated with a particular situation of use (including particular communicative purposes)” (Biber & Conrad 2019: 6).

The expected results should show that the modification of deverbal nouns ending in *-nje* by adverbs and by the *od strane* phrase is typical for some formal text types, as *media, inform, state, science, company, database* and *community* are. However, the results of the research emphasize that this is not the case for all of the formal text types/areas. Because of the fact that the both here examined modifiers (adverbs and the *od strane* phrase) are typical verbal characteristics of deverbal nouns (Sleeman & Brito 2010), besides their importance in the extraction and analysis of the grammatical characteristics of the Serbian public discourse, the results of the research could also give some insight into the broader architecture of the nominalizations in Serbian (and generally), in the course of the *structural model* (Alexiadou 2001).

## **Bibliography**

- Alexiadou, A. (2001). *Functional Structure in Nominals: Nominalization and Ergativity*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Biber, D. & S. Conrad (2019). *Register, genre, and style* [2nd edition]. Cambridge: Cambridge University Press.
- Klikovac, D. (2008). *Jezik i moć*. Beograd: XX vek.
- Radovanović, M. (1990). *Spisi iz sintakse i semantike*. Sremski Karlovci: Izdavačka knjižarnica Zorana Stojanovića / Novi Sad: Dobra vest.
- Sleeman, P. & A.M. Brito (2010). Aspect and argument structure of deverbal nominalizations: a split vP analysis. From: A. Alexiadou & M. Rathert (ed.) *Nominalizations across Languages and Frameworks*, 191-209. Berlin: Mouton de Gruyter.



4

*Onomastika  
v době digital humanities*

organizují

Jaroslav David

Michal Místecký

Tereza Klemensová

(CZE)

# Popis panelu

Panel je zaměřen na aktuální trendy a témata výzkumu nejen v současné onomastice, ale i v dalších výzkumných oblastech, kde onomastika nachází své uplatnění. Ačkoliv se onomastika jako samostatná disciplína již dávno etablovala – u nás zejména zásluhou teoretického a metodologického rámce vypracovaného R. Šrámkem (1999), stále je velká část výzkumů vedena v duchu tradičního etymologizujícího nebo klasifikačně zaměřeného přístupu (srov. Hough, 2016). Většina monografií a studií má charakter konkrétně zacílených mikroanalýz a větší syntetizující díla, jako je například gramatika proprií či celočeský onomastický atlas, chybí. To je dáno zejména dominantní metodou výzkumu, jenž byl doposud z důvodu vazby disciplíny především k historii, potřeby vytváření materiálové báze (slovníky) a vzhledem ke specifickým propriím a absenci onomasticky zpracovaných korpusů realizován kvalitativní cestou. Úkolem tohoto panelu je nalézat a diskutovat možnosti a směry, kterými se může výzkum proprií nadále ubírat, přičemž se nechceme omezovat pouze na moderní lingvistické přístupy, ale v přístupech k onymickému materiálu naopak klademe důraz na interdisciplinaritu. V tomto ohledu se nabízí nejen tradiční spojení s historií, archeologií, geografii a kartografií, ale styčné plochy se otevírají také pro spolupráci se sociolingvistikou, antropologií, versologií, literární vědou či politologií (viz např. David – Mácha, 2014; Plecháč – Kolár, 2017; David – Klemensová – Místecký a kol., 2022). Tematické vymezení panelu tak klade důraz především na metodologické otázky a problémy využití nových přístupů (aplikace digital humanities) nejen v tradiční onomastice, ale i při otevírání nových témat, přičemž se akcentují korpusové a kvantitativní přístupy k výzkumu jazyka (viz např. Straková – Straka a kol., 2017; Cvrček a kol., 2020; Motschenbacher, 2020; David – Místecký, 2023).

## **Výzkumná témata:**

- vlastní jména v perspektivě kvantitativní analýzy
- metodologie tvorby onomasticky zaměřeného korpusu
- software NER a jeho aplikace v češtině (možnosti a úskalí; viz např. Straková – Straka a kol., 2017)
- vlastní jména jako stylový prostředek (od deskripce k registrům; viz např. Cvrček a kol., 2020)
- literární onomastika a problematika určování autorství
- vlastní jména v kartografické perspektivě – výzvy prezentace a vizualizace onomastických dat
- aplikace kvantitativněonomastických metod v historii a archeologii
- elektronické databáze pro sběr a prezentaci vlastních jmen (tvorba, využití)
- konkrétně zacílené kvantitativněonomastické analýzy proprií – podmínkou je obecnější metodologický přesah (např. diverzifikace českých antroponym; vymezení vlastních jmen; apelativizace, proprializace, transonymizace; sémantika)

## Literatura

- CVRČEK, V. a kol. (2020): Registry v češtině. Praha: Nakladatelství Lidové noviny.
- DAVID, J. – KLEMENSOVÁ, T. – MÍSTECKÝ, M. a kol. (2022): Od etymologie ke krajině. Onomastika pro 21. století. Brno: Host.
- DAVID, J. – MÁCHA, P. (2014): Paměť míst. Paměť, identita, kulturní dědictví. Brno: Host.
- DAVID, J. – MÍSTECKÝ, M. (2023): Prolegomena ke kvantitativní onomastice. Acta onomastica, 64(2), s. 301–320.
- HOUGH, C. (ed., 2016): The Oxford Handbook of Names and Naming. Oxford: Oxford University Press.
- MOTSCHENBACHER, H. (2020): Corpus Linguistic Onomastics. A Plea for a Corpus Based Investigation of Names. Names, 68(2), s. 88–103.
- PLECHÁČ, P. – KOLÁR, R. (2017): Kapitoly z korpusové versologie. Praha: Akropolis.
- STRAKOVÁ, J. – STRAKA, M. a kol. (2017): Czech Named Entity Corpus. In: N. Ide – J. Pustejovsky (eds.), Handbook of Linguistic Annotation. Dordrecht: Springer Netherlands, s. 855–873.
- ŠRÁMEK, R. (1999): Úvod do obecné onomastiky. Brno: Masarykova univerzita.

# Obsah

*A pojmenování se stalo obrazem: geoinformační možnosti vizualizace geografických jmen*  
Jan D. Bláha, Martin Bartůněk (Univerzita J. E. Purkyně v Ústí nad Labem)

*Doba zkoušek a naděje – kvantitativní analýza propriální složky textů jako užitečný nástroj pro historiky*

Jaroslav David, Jana Davidová Glogarová (Ostravská univerzita)

*Korpus vlastních jmen OnomOs*

Jaroslav David (Ostravská univerzita), Tomáš Jeziorský, Michal Křen (Univerzita Karlova)

*Onomastika v době digital humanities*

Jaroslav David, Tereza Klemensová, Michal Místecký (University of Ostrava)

*Contact onomastics and digitalization: a toponomastic dataset for geospatial comparison in northern Bavaria*

Nicolas M. Jansens (Heidelberg University / Charles University)

*Place names in time and space. The application of geostatistical analysis to research on medieval settlement in Bohemia and Bavaria*

Tomáš Klír, Martin Janovský, Viktorie Janovská (Charles University), Nicolas M. Jansens (Heidelberg University, Charles University)

*Rekonstrukce jazykové krajiny v minulosti? Možnosti a limity výzkumu urbanonym prostřednictvím historické fotografie*

Jan Lhoták (Katedra historických věd, FF ZČU)

*Toponyma jako zbraň: role mapového názvosloví v česko-německých národnostních sporech v 19. a 1. polovině 20. století*

Jitka Močičková (Historický ústav AV ČR)

*Možnosti a meze korpusového výzkumu proprií*

Klára Osolsobě, Hana Žižková (Masarykova univerzita)

*Rozpoznávání pojmenovaných entit v básnických textech*

Petr Plecháč<sup>1</sup>, Silvie Cinková<sup>2,1</sup>, Robert Kolár<sup>1</sup> (<sup>1</sup>Ústav pro českou literaturu AV ČR, <sup>2</sup>Univerzita Karlova)

*Vlastní jména v perspektivě kvantitativní analýzy divadelních her*

Petr Pořízka (Univerzita Palackého)

*Kapitulace na konci světa: kvantitativní analýza jmen v současných francouzských dystopiích*

Karel Střelec, Michal Místecký (Ostravská univerzita)

*NimiSampo: Databáze finských toponym online*

Soňa Wojnarová (ÚJČ AV ČR)

**A pojmenování se stalo obrazem: geoinformační možnosti vizualizace geografických jmen**

Jan D. Bláha, Martin Bartůněk (Univerzita J. E. Purkyně v Ústí nad Labem)

Toponyma (geografická jména) jsou odpradáвна nedílnou součástí studia geografie a obsahu map. Existuje však jen málo výzkumných záměrů, v nichž se toponyma stávají základním předmětem zájmu vizuálního vyjádření tematické kartografie, a stávají se tak jedním z obrazů reality. V rámci tohoto příspěvku bude představeno propojení onomastiky a geografie ve výzkumu toponym prostřednictvím tematického kartografického vyjádření. Cílem příspěvku je nejen poskytnout základní pojednání o znázornění toponym a jejich vlastností v mapě, ale také představit konkrétní možnosti kartografické vizualizace toponym na základě několika případových studií z lokální a regionální úrovně.

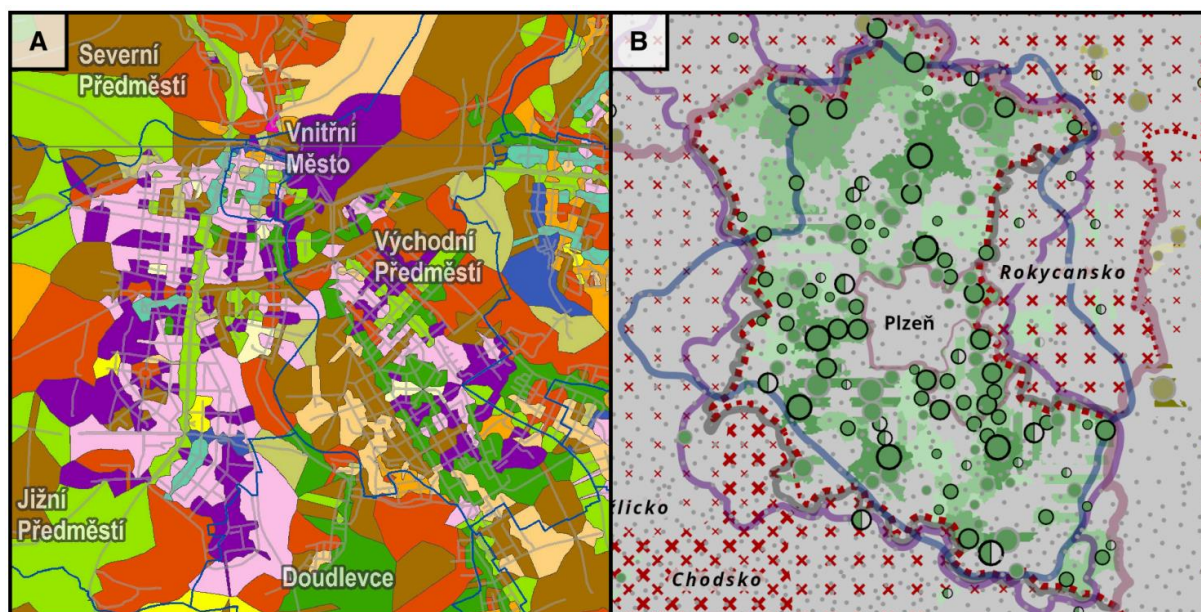
Velmi důležitou vlastností toponym je jejich prostorový aspekt, který lze územně ukotvit (lokalizovat) a následně vyjádřit v mapě. Tato jména však představují mnohem více než jen prostředek pro orientaci v prostoru. Jsou klíčovými prvky územní identity a bohaté databáze kulturní krajiny/území, která odkazuje na propojení mezi jménem a přírodní sférou, historií a společností, přičemž tyto vztahy jsou dynamické a proměnlivé v čase (Helleland, 2012; Alderman, 2008; Berg & Kearns, 1996; Tuan, 1991). S ohledem na tuto komplexnost je možné vytvářet nejen topografické, ale i tematické mapy (Dent et al., 2008; Slocum et al., 2009) za využití různorodých grafických vyjadřovacích prostředků (Bertin, 1983).

Při studiu toponym představených v případových studiích autoři využili různých metod sběru dat a zdrojů, což jim mj. umožňuje představit různorodé formy kartografické vizualizace. Jedním z cenných zdrojů dat je *Registr územní identifikace, adres a nemovitostí* pod správou ČÚZK (2024b), který poskytuje velmi detailní a aktuální údaje o urbanonymech v Česku včetně přesahu do analýzy starých mapových dokumentů (např. staré plány měst) – (např. Bláha, 2017). Dalším zdrojem je *Databáze geografických jmen ČR* (ČÚZK, 2024a), která má navíc od roku 2023 volnou licenci. Z tradičních sociologických metod lze realizovat dotazníkové šetření, které umožňuje získat informace od obyvatel o vnímání toponym ve vztahu k symbolické rovině regionu (např. Šifta & Chromý, 2014). Další možností je použití korpusu (ČNK, 2024) a analyzovat frekvenci a kontext použití geografických jmen v textech. Analýzou toponym v učebnicích lze získat představu o prezentaci regionálně-geografického kurikula v podobě *mapy četnosti geografických jmen* (Kučerová et al., 2014). Tato variabilita v metodách sběru dat a jejich zdrojů předurčuje možnosti jejich analýzy a následné vizualizace za pomoci nástrojů geoinformačních systémů (GIS) a metod tematické kartografie.

Myšlenka propojovat onomastiku a geografii skrze kartografické/obrazové vyjádření vzešla z akademického prostředí, především z prací realizovaných v rámci studijních projektů nebo v závěrečných pracích. V rámci několika příkladů realizace (viz Obrázek 1), především od studentů geografie, budou představeny případové studie věnující se především jménům regionů (choronyma), jménům ulic a náměstí (urbanonyma), ale i dalším druhům toponym a ergonym (obecně geografická jména) a jejich vyjádření v mapě na základě konceptů regionální a kulturní geografie.

Toponyma jsou objektem zájmu všech tří oborů (onomastika, geografie a kartografie) a interdisciplinarita ve výzkumu může obohatit metodologii výzkumů. Pouze kombinace přístupů těchto disciplín umožní hlubší porozumění toponymům v jejich širším kontextu, což může být aplikovatelné například v regionálním rozvoji.

Obrázek 1. (a) Příklad vizualizace klasifikované urbanonymie města; (b) Příklad vizualizace rozsahu choronyma dle korpusu



## Bibliografie

- ALDERMAN, D. H. (2008). Place, naming, and the Interpretation of Cultural Landscape. In: B. J. Graham, P. Howard (eds.). *The Ashgate Research Companion to Heritage and Identity*. Farnham, UK: Ashgate Press, s. 195–213.
- BERG, L. D. A KEARNS, R. A. (1996). Naming as norming: race, gender, and the identity politics of naming places in Aotearoa/New Zealand. *Environment and Planning D: Society and Space*, 14(1), s. 99–122.
- BERTIN, J. (1983). *Semiology of Graphics*. Madison: University of Wisconsin Press.
- BLÁHA, J. D. (2017). Geografická analýza vývoje místních jmen ulic ve městě Chotěboři aneb Jak si Buttula dává Ovesnou kaši. In: L. Křížová, J. Martínek (eds.). *Od Karlova mostu ke Gottwaldovu: Osobnosti v názvech měst a míst*. Praha: Historický ústav, s. 147–169.
- ČNK (2024). *Český národní korpus*. [cit. 2024-03-03]. Dostupné z: <<https://korpus.cz>>
- ČÚZK (2024a). *Databáze geografických jmen České republiky (Geonames)*. [cit. 2024-03-03]. Dostupné z: <<https://ags.cuzk.cz/geoprohlizec/?k=490>>
- ČÚZK (2024b). *Registr územní identifikace, adres a nemovitostí (RÚIAN)*. [cit. 2024-03-03]. Dostupné z: <<https://www.cuzk.cz/ruian/RUIAN.aspx>>
- DENT, B. D., TORGUSON, J. S., HODLER, T. W. (2008). *Cartography: Thematic Map Design*. New York: McGraw-Hill Education.
- HELLELAND, B. (2012). Place names and identities. *Names and Identities. Oslo Studies in Language*, 4(2), s. 95–116.
- KUČEROVÁ, S. R., et al. (2014). Geographical Names Frequency Map as a Tool for the Assessment of Territorial Representations in Geography Textbooks. In: P. Knecht, E. Matthes, S. Schütze, & B. Aamotsbakken (eds.). *Methodologie und Methoden der Schulbuch- und Bildungsmedienforschung*. Bad Heilbrunn: Klinkhardt, s. 264–275.
- SLOCUM, T., et al. (2009). *Thematic Cartography and Geovisualization*. Upper Saddle River, NJ: Pearson Prentice Hall.
- ŠIFTA, M., CHROMÝ, P. (2017). The importance of symbols in the region formation process. *Norsk Geografisk Tidsskrift*, 71(2), s. 98–113.
- TUAN, Y-F. (1991). Language and the Making of Place: A Narrative-Descriptive Approach. *Annals of the Association of American Geographers*, 81(4), s. 684–696.

## Doba zkoušek a naděje – kvantitativní analýza propriální složky textů jako užitečný nástroj pro historiky

Jaroslav David, Jana Davidová Glogarová (Ostravská univerzita)

Jazykový obrat (linguistic turn, Sprachwende), jenž je spojován především se stejnojmenným souborem textů editovaných Richardem Rortym (1967), lze zjednodušeně popsat jako změnu ve filozofickém myšlení. V jejím pozadí je poznání, že pro řešení filosofických otázek je zcela zásadní analyzování jazyka. Dopad proměny paradigmatu výzkumu byl však mnohem širší a výrazně přispěl k posílení mezioborového aspektu v řadě disciplín. V historii pak obnovena důvěra ve svědectví jazykových faktů našla své místo ve zkoumání dějin všedního dne i idejí, v mikroperspektivách či zkoumání antropologických aspektů kulturních a politických dějin: „Lingvistický obrat se stal v rámci historického bádání součástí úsilí o překonání socioekonomického determinismu a o posílení role kulturních faktorů, mezi nimiž jazyk zaujímá klíčové místo“ (Iggers, 2022, s. 124). Tento obrat nalezl ohlas v metodologii desítek monografií a studií, v nichž se jako jedna z nejsilnějších oblastí jeho aplikace vyprofilovala na lexikum zaměřená historická sémantika (přehledově viz Macek, 1991; David et al., 2014). Specifické oblasti vlastních jmen však doposud jak ve výzkumu historické sémantiky, tak při zkoumání historických textů nebyla věnována významnější pozornost. Přitom se jedná o frekvenčně významnou a sémanticky důležitou složku textů zejména publicistické povahy (ze starších analýz srov. např. David et al., 2013; Davidová Glogarová – Kubát, 2020; David – Davidová Glogarová – Místecký, 2021).

V souvislosti s nástupem moderních metod do onomastického výzkumu, zejména kvantitativní a korpusové analýzy (David – Místecký, 2023; David et al., 2022; Motschenbacher, 2022), se však nabízí při úzkém propojení obou disciplín nový potenciál, a to využití kvantitativních metod při analýze historických pramenů, ale i jejich propriální složky.

Příspěvek proto představuje možnosti, které kvantitativní analýza vlastních jmen (antroponym, toponym, popř. chrématonym) nabízí historické vědě při studiu moderních dějin. Na základě analýzy různorodých textů z období druhé světové války si klade za cíl ověřit využitelnost a meze dostupných kvantitativně lingvistických metod, jako jsou např. frekvenční a kolokační analýza, analýza tematické koncentrace či keywords, ve výzkumu vlastních jmen v historických pramenech. Metody budou aplikovány na české texty z let 1939–1945, a to jak publicistické (Moravec, 1942; Beneš, 1945; Gottwald, 1946; Masaryk, 1945), tak memoárové (deník manětínského děkana Wonky, 2010). Výsledky analýzy, k níž bude využít jednak nástroj #Lancsbox, jednak aplikace dostupné v rámci Českého národního korpusu, se zaměří na způsoby využití a tematizace vlastních jmen v textech sledovaného období. Závěry pak budou komparovány s interpretacemi a závěry tradičně pojatého historického výzkumu. Příspěvek si klade za cíl rozšířit okruh využití onomastiky a moderních lingvistických metod v historických vědách, zejména pro účely textových analýz. Chce rovněž otevřít diskuzi, zda může být moderní onomastika vědou nápomocnou historii – nikoliv pouze pomocnou vědou historickou.

## Bibliografie

- [1] BENEŠ, E. 1945. *Šest let exilu a druhé světové války*. Praha: Týdeník Českoslovák.
- [2] *Český národní korpus*. Dostupný online: <https://www.korpus.cz/>.
- [3] DAVID, J. et al. 2013. *Slovo a text v historickém kontextu. Perspektivy historickosémantické analýzy jazyka*. Brno: Host.
- [4] DAVID, J. et al. 2022. *Od etymologie ke krajině. Onomastika pro 21. století*. Brno: Host.
- [5] DAVID, J. – DAVIDOVÁ GLOGAROVÁ, J. – MÍSTECKÝ, M. 2021. A Two-Man Show: Stylometric Analysis of Personal Names in Rudolf Slánský's Staged Trial Newspaper Reports. *KALBU STUDIJOS – STUDIES ABOUT LANGUAGES* 38(1), s. 115–130.
- [6] DAVID, J. – MÍSTECKÝ, M. 2023. Prolegomena ke kvantitativní onomastice. *Acta onomastica* 64(2), s. 301–320.
- [7] DAVIDOVÁ GLOGAROVÁ, J. – KUBÁT, M. 2020. Srovnávací frekvenční analýza exilových projevů Klementa Gottwalda a Edvarda Beneše z let 1939–1945. *Slovo a slovesnost* 81(1), s. 65–77.
- [8] GOTTWALD, K. 1946. *Deset let: Sborník statí a projevů 1936–1946*. Praha: Svoboda.
- [9] IGGERS, G. G. 2002. *Dějepisectví ve 20. století. Od vědecké objektivy k postmoderní výzvě*. Praha: NLN.
- [10] #LancsBox: Lancaster University corpus toolbox. Dostupný online: <http://corpora.lancs.ac.uk/lancsbox>.
- [11] MACEK, J. 1991. Historická sémantika. *Český časopis historický* 89(1), s. 1–30.
- [12] MASARYK, J. 1945. *Volá Londýn*. Londýn: Lincolns-Prager.
- [13] MORAVEC, E. 1942. *Tři roky před mikrofonem*. Praha: Orbis.
- [14] MOTSCHENBACHER, H. 2020. Corpus Linguistic Onomastics: A Plea for a Corpus-Based Investigation of Names. *Names* 68(2), s. 88–103.
- [15] RORTY, R. 1967. *The Linguistic Turn*. Chicago and London: The University of Chicago Press.
- [16] WONKA, F. 2010. *Doba zkoušek a naděje. Deník děkana Františka Wonky z let 1938–1945*. Manětín: Machart, s.r.o.



### Korpus vlastních jmen OnomOs

Jaroslav David (Ostravská univerzita), Tomáš Jeziorský, Michal Křen (Univerzita Karlova)

Česká onomastika od svého vzniku v 19. století prošla dlouhým vývojem. Prvotní pozitivisticko-etymologická fáze charakterizovaná silnou vazbou k historii byla ve 2. polovině 20. století překonána posílením její lingvistické složky, plným využitím mezioborového potenciálu a postupným důrazem na empirizaci a metodologické aspekty. Onomastika tak po překonání jisté “splendid isolation” vstoupila do nového tisíciletí nejen posílena o impulzy z oblasti antropologie, kulturní a sociální geografie, ale také postupně vstřebává kvantitativní a korpusový přístup k vlastním jménům (Šrámek, 2007; nejnověji David – Místecký, 2023; David et al., 2022; Motschenbacher, 2022), stejně jako moderní trendy ve vizualizaci a způsobu sběru dat (projekty “Mapa českého verše” nebo “Živá jména”, Vrbík – Lábus, 2022).

Tuto proměnu paradigmatu onomastického výzkumu ilustruje vznik specializovaného Korpusu OnomOs (2023). Jde o první korpus ČNK se značkováním vlastních jmen v souladu s jejich onomastickým pojetím, který by v budoucnu mohl sloužit jako metodologické východisko pro projekty podobného typu. Korpus OnomOs představuje lingvisticky zpracovanou databázi textů z periodik Rudé právo (vycházelo 1920–1995) a Právo (1995–dosud). Obsahuje vždy jedno číslo z každé dekády, v níž (Rudé) Právo vycházelo. Do korpusu byly zahrnuty texty, v nichž dominuje jazyková složka; vypouštěny byly proto např. reklamy a inzeráty, programy kin, divadel a rádia, některé typy textů ze sportovní rubriky. Celkem korpus obsahuje 255 149 tokenů.

Základem pro vytvoření korpusu byly pečlivě manuálně korigované články (Rudého) Práva ve formátu XML, které byly dále anotovány již plně automaticky (tokenizace, větná segmentace, lemmatizace, morfologické značkování, označení jmenných entit). Přitom se projevil obecný problém většiny anotačních nástrojů, totiž že na vstupu typicky vyžadují prostý text. To komplikuje zpracování textů v XML (které již anotovány de facto jsou) a zejména zapojování anotačních nástrojů sekvenčně za sebou: typicky je totiž potřeba původní anotaci v každém kroku zachovat a “jenom” přidat další. Tento problém byl vyřešen sadou pythonovských skriptů, které XML na vstupu rozdělí na prostý text a tzv. stand-off anotaci, v níž je uložena vstupní anotace spolu s odkazy na její přesné umístění v původním textu. Poté, co je prostý text anotován, je nová výstupní anotace “slita” se stand-off anotací původní. V případě korpusu OnomOs byla navíc anotace jmenných entit nástrojem NameTag2 (2019) z technických důvodů předřazena anotaci morfologické (Jelínek et al., 2021).

V druhé části pak příspěvek na konkrétním výzkumu ukáže možnosti využití korpusu. Výzkumná sonda se pokusí ukázat, zda a jak se proměňuje frekvence příjmení, jejich kolokací (rodné jméno, označení funkce, pán/paní, přechylování) a pádových forem. Korpus bude pro tyto účely rozdělen na dva menší celky zahrnující Rudá práva z období před sametovou revolucí (ročníky 1956, 1969, 1979 a 1988) a po ní (1996, 2003 a 2014).

## Bibliografie

- [1] DAVID, J. et al. 2022. *Od etymologie ke krajině. Onomastika pro 21. století*. Brno: Host.
- [2] DAVID, J. – MÍSTECKÝ, M. 2023. Prolegomena ke kvantitativní onomastice. *Acta onomastica* 64(2), s. 301–320.
- [3] JELÍNEK, T. – KŘIVAN, J. – PETKEVIČ, V. – SKOUMALOVÁ, H. – ŠINDLEROVÁ, J. 2021. SYN2020: A new corpus of Czech with an innovated annotation. In: K. Ekštejn – F. Pártl – M. Konopík (eds.) *Text, Speech, and Dialogue*. TSD 2021. Lecture Notes in Computer Science, vol. 12848. Cham: Springer, s. 48–59.
- [4] *Korpus OnomOs*. 2023. Praha: Ústav Českého národního korpusu FF UK. Dostupný online: <http://www.korpus.cz>.
- [5] LÁBUS, V. – VRBÍK, D. 2023. *Atlas libereckých živých jmen*. Liberec: Technická univerzita v Liberci.
- [6] *Mapa českého verše*. Praha: Ústav pro českou literaturu AV ČR. Dostupný online: [https://versologie.cz/v2/tool\\_map/](https://versologie.cz/v2/tool_map/).
- [7] MOTSCHENBACHER, H. 2020. Corpus Linguistic Onomastics: A Plea for a Corpus-Based Investigation of Names. *Names* 68(2), s. 88–103.
- [8] *NameTag2*. 2019. Praha: Ústav formální a aplikované lingvistiky MFF UK. Dostupný online: <https://ufal.mff.cuni.cz/nametag/2>.
- [9] ŠRÁMEK, R. 2007. Onomastika. In: Pleskalová, J. et al. *Kapitoly z dějin české jazykové a bohemistiky*. Praha: Academia, s. 377–425.
- [10] *Živá jména*. Liberec: Technická univerzita v Liberci. Dostupný online: <https://mapy.fp.tul.cz/zivajmena>.

**Onomastika v době digital humanities**

Jaroslav David, Tereza Klemensová, Michal Místecký (University of Ostrava)

Úvodní příspěvek nejprve stručně charakterizuje současnost české onomastiky (David –Mácha 2014; David – Klemensová – Místecký 2022), její východiska (Šrámek 1999) a postavení ve světovém kontextu. Důraz je položen na postupující kvantifikaci disciplíny související s obecným trendem akcentujícím exaktní a empirické přístupy v humanitních vědách (Motschenbacher 2020; David – Místecký 2023). Přednosti, ale i meze takto pojatého přístupu v onomastice jsou představeny na konkrétních výzkumných tématech, jež byla v uplynulých letech řešena (apelativizace, honorifikační motiv, přechylování příjmení, morfologie vlastních jmen atd.). Součástí prezentace je stručné představení základní bibliografie v poslední době se dynamicky rozvíjejícího oboru.

V druhé části příspěvku bude pozornost upřena na budoucnost disciplíny a možnosti výzkumu. Budou formulovány badatelské otázky, témata a úkoly, které se před onomastikou otevírají a o nichž bude možno diskutovat v kuloárech panelu. Mezi tyto lze zařadit např. 1) NER, disambiguace proprií a potřeba rozsáhlých korpusů se značkovánými vlastními jmény (Straková – Straka a kol. 2017); 2) přítomnost proprií v jazykových slovnících; 3) využití onomastiky při diagnostice psychických nemocí (Collins a kol. 2023); 4) honorifikační motiv a politická kolonizace prostoru v celoslovenské perspektivě (Klemensová – Místecký 2022); 5) role onomastiků při tvorbě nové urbanonymie; 6) využití onomastiky ve výzkumu politicky angažovaných projevů a propagandy; 7) role umělé inteligence při onomastickém výzkumu (např. při modelování); 8) využití statistických metod při ověřování hypotéz.

V závěru pak budou představena témata, jimž se věnují příspěvky v panelu – kvantitativní onomastika, onymické korpusy a databáze, využití kvantitativních metod v literární onomastice a analýze textů, moderní interdisciplinární propojení onomastiky a geografie, archeologie a historie.

## Bibliografie

COLLINS, L. a kol. (2023): Corpus linguistics and clinical psychology: Investigating personification in first-person accounts of voice-hearing. *International Journal of Corpus Linguistics*, 28(1), s. 28–59.

CVRČEK, V. a kol. (2020): *Registry v češtině*. Praha: Nakladatelství Lidové noviny.

DAVID, J. – KLEMENSOVÁ, T. – MÍSTECKÝ, M. a kol. (2022): *Od etymologie ke krajině. Onomastika pro 21. století*. Brno: Host.

DAVID, J. – MÁCHA, P. (2014): *Paměť míst. Paměť, identita, kulturní dědictví*. Brno: Host.

DAVID, J. – MÍSTECKÝ, M. (2023): Prolegomena ke kvantitativní onomastice. *Acta onomastica*, 64(2), s. 301–320.

KLEMENSOVÁ, T. – MÍSTECKÝ, M. (2022): Schiller kontra Komenský – uliční názvy v pohraničí a honorifikace. In: J. David – T. Klemensová – M. Místecká a kol. (eds.), *Od etymologie ke krajině. Onomastika pro 21. století*. Brno: Host, s. 335–351.

MOTSCHENBACHER, H. (2020): Corpus Linguistic Onomastics. A Plea for a Corpus Based Investigation of Names. *Names*, 68(2), s. 88–103.

STRAKOVÁ, J. – STRAKA, M. a kol. (2017): Czech Named Entity Corpus. In: N. Ide – J. Pustejovsky (eds.), *Handbook of Linguistic Annotation*. Dordrecht: Springer Netherlands, s. 855–873.

ŠRÁMEK, R. (1999): *Úvod do obecné onomastiky*. Brno: Masarykova univerzita.

**Contact onomastics and digitalization:  
a toponomastic dataset for geospatial comparison in northern Bavaria**

Nicolas M. Jansens (Heidelberg University / Charles University)

For well over a century, scholars have endeavored to identify and explain the oldest toponyms of northern Bavaria, particularly those with a Slavic origin. The fruits of their labor now lie scattered among numerous contributions in books, journals and online resources.<sup>1</sup> Once extracted and linked to geographic coordinates, the data contained in these sources may be used to reconstruct distributions of toponyms according to their historical linguistic features and compare them to archeological finds using geographic information systems (GIS). To this end, the results of these highly qualitative, philological name studies must first be converted for quantitative use, entailing manual adaptation to standardized schemas and cataloguing of the morphological components of each toponym.

A suitable database structure and notational schemas able to capture the necessary levels of analysis were largely developed over the course of a pilot project looking at toponymy in three early medieval language contact zones. Among these was northern Bavaria (esp. *Bavaria Slavica*), an area of continued interest to our team of researchers centered around archeologist Tomáš Klír in Prague. It was he who collected the initial data on Slavic toponyms in the region from the available qualitative sources. My job has been to linguistically verify and extend his work and to enable systematic geospatial comparison. This latter task involved integration of the Slavic data (700+ entries, at various stages of elaboration) into a much more complete dataset of Bavarian toponyms (some 52,000 entries) taken from *GeoNames*, an open index of geographical objects. I then performed a data-filter-based mass annotation of German topoformants in the additional data entries.

The dataset presently facilitates geospatial comparison of distributions for objects with names bearing particular German formants (e.g. *-dorf*, *-heim*, *-hof|en*, *-ing|en*) vs. those of Slavic origin vs. those of so-called mixed Slavic/German origin. Nearing completion is the further task of morphologically elaborating all etymological proposals for names of Slavic origin, which, once accomplished, will allow us to also reveal and compare distributions within that subset of objects.

In this connection, I have devised solutions for several data-structural challenges in onomastics, including instances of multiple names given to a single geographical object, the embedding of one name form in another and the linguistic handling of names combining elements from multiple languages.

---

<sup>1</sup> A long path of progress can be traced across such major onomastic works as Gradl 1890/1892; Ziegelhöfer – Hey 1911; 1920; Bachmann 1926; Schwarz 1960; Eichler 1962; Eichler *et al.* 2001/2006, which are accompanied by numerous more recent contributions from further scholars, such as Wolf Armin Frhr. v. Reitzenstein, Albrecht Greule, Wolfgang Janka and Harald Bichlmeier.

## Bibliography

- Bachmann, Margarete. 1926. *Die Verbreitung der slawischen Siedlungen in Nordostbayern*. Erlangen: Verlag von Max Mencke.
- Eichler, Ernst. 1962. Zur Etymologie und Struktur der slawischen Orts- und Flußnamen in Nordostbayern. *Wissenschaftliche Zeitschrift der Karl-Marx-Universität Leipzig. Gesellschafts- und sprachwissenschaftliche Reihe*, 11, S. 365–95. = Eichler, Ernst. 1985. Zur Etymologie und Struktur der slawischen Orts- und Flußnamen in Nordostbayern. In: Ernst Eichler, *Beiträge zur deutsch-slawischen Namenforschung (1955–1981)*. Leipzig: Zentralantiquariat der DDR, S. 269–395.
- Eichler, Ernst – Greule, Albrecht – Janka, Wolfgang – Schuh, Robert (Hg.). 2001/2006. *Beiträge zur slawisch-deutschen Sprachkontaktforschung*, 1: *Siedlungsnamen im oberfränkischen Stadt- und Landkreis Bamberg* – 2: *Siedlungsnamen im oberfränkischen Stadt- und Landkreis Bayreuth (Slavica. Monographien, Hand-, Lehr- und Wörterbücher, Bd. 2, 4)*. Heidelberg: Universitätsverlag Winter.
- Gradl, Heinrich. 1890/1892. Die Ortsnamen am Fichtelgebirge und in dessen Vorlanden. [1. Abtheilung: Deutsche Ortsnamen]. In: *Archiv für Geschichte und Alterthumskunde von Oberfranken*, 18.1, S. 1–177. – [2. Abtheilung: Slawische Ortsnamen]. *Archiv für Geschichte und Alterthumskunde von Oberfranken*, 18.3, S. 81–179. = Gradl, Heinrich. 1891/1892. *Die Ortsnamen am Fichtelgebirge und in dessen Vorlanden*, [1. Abtheilung: Deutsche Ortsnamen] – 2. Abtheilung: *Slawische Namen*. Eger: in Kommission bei J. Kobrtsch und Gschihay.
- Schwarz, Ernst. 1960. *Sprache und Siedlung in Nordostbayern (Erlanger Beiträge zur Sprach- und Kunstwissenschaft, 4)*. Nürnberg: Verlag Hans Carl.
- Ziegelhöfer, Adam – Hey, Gustav. 1911. *Die Ortsnamen des ehemaligen Hochstifts Bamberg*. Bamberg: Verlag des Historischen Vereins.
- Ziegelhöfer, Adam – Hey, Gustav. 1920. *Die Ortsnamen des ehemaligen Fürstentums Bayreuth (Archiv für Geschichte und Alterthumskunde von Oberfranken, 27)*. Bamberg: Selbstverlag von Ziegelhöfer.

**Place names in time and space. The application of geostatistical analysis to research on medieval settlement in Bohemia and Bavaria**

Tomáš Klír, Martin Janovský, Viktorie Janovská (Charles University),  
Nicolas M. Jansens (Heidelberg University, Charles University)

In the Early Middle Ages, the territory of present-day north-eastern Bavaria was part of a wide and long border zone and an important communication corridor at the east of the Franconian Empire and later the German kingdom. Over the years, this territory has attracted the long-term and intensive attention of linguists, historians and archaeologists. This is due to the fact that an extraordinary confluence of rich written, linguistic, and archaeological evidence is available for this area, which makes it possible to address complex issues such as the social, economic, and linguistic processes occurring on the borders of the early empires [3,4]. The aim of the present contribution is to present methods and preliminary results of an ongoing interdisciplinary research project, which is primarily concerned with the relationship between the speakers of Slavic and German languages in this area, against the background of the expanding settlement oikumene and the growing power structures of the Franconian Empire and German kingdom. It will present the method by which qualitative data and knowledge from history, linguistics (onomastics) and archaeology are combined, as well as methods of quantitative digital processing and evaluation [1, 2]. Through the use of geostatistical and statistical methods, distribution patterns of different types of evidence are recognized, which help us to reconstruct the development of the medieval settlement oikumene and the spread of the Slavic and German languages in different social contexts (first part of the paper). On the basis of the successful digitization of toponomastic and archaeological data, the study area can be placed in a broader geographical context that includes the entire territory of Bohemia, West-Bohemia particularly, and a large part of present-day Bavaria (second part of the paper).

## Bibliography

- [1] Bichlmeier, Harald and Nicolas M. Jansens. 2024. Vorstellung des Projekts „Sprach- und Kulturdynamik in einer Frontiergesellschaft: Neue Perspektiven auf Nordostbayern und Westböhmen im frühen Mittelalter“ samt einiger topolinguistischer Problemstellungen. In Peter Ernst and Stephan Gaisbauer (eds.), *Akten der Jubiläumstagung 50 Jahre ‚Österreichische Namenforschung‘: ‚Onomastik: Bestandsaufnahmen und Zukunftsperspektiven‘*, 55–101. (= Jahrbuch Österreichische Namenforschung 1) Wien: Praesens-Verlag 2024.
- [2] Janovský, Martin, Nicolas M. Jansens, Viktorie Janovská and Tomáš Klír. Forthcoming. A Landscape of Two Linguistic Worlds and its Investigation Using a Clustering Method: A Contribution to the Possibilities of DBSCAN in Archaeological-Linguistic Research. *Journal of European Landscapes*.
- [3] Klír, Tomáš 2020. Language, ethnicity, and material culture: Northeastern Bavaria in the Early Middle Ages. In Tomáš Klír, Vít Boček and Nicolas Jansens (eds.), *New Perspectives on the Early Slavs and the Rise of Slavic. Contact and Migrations*, 113–130. (= Empirie und Theorie der Sprachwissenschaft, Volume 6) Heidelberg: Winter Verlag.
- [4] Schwarz, Ernst 1960. *Sprache und Siedlung in Nordostbayern*. (Erlanger Beiträge zur Sprach- und Kunstwissenschaft 4) Nürnberg: Verlag Hans Carl.



**Rekonstrukce jazykové krajiny v minulosti?  
Možnosti a limity výzkumu urbanonym prostřednictvím historické fotografie**

Jan Lhoták (Katedra historických věd, FF ZČU)

Psaný jazyk ve veřejném prostoru se nejčastěji zachycoval terénním výzkumem synchronizovaným se samotnou (ponejvíce kvantitativní) analýzou sebraných dat, pozvolna stále intenzivněji doplňovanou diachronní perspektivou zkoumaných subjektů zohledněním historicity (Blommaert – Dong, 2020). Diachronní přístup se mohl dále uplatnit časosběrným shromažďováním fotografií v řádu měsíců či let ad hoc pro výzkum proměn jazykové krajiny (Backhaus, 2005; Pavlenko, 2010; Tufi, 2022), přičemž zvolený časový rámec nepřekročil jednu dekádu. Transpozice výzkumu více do minulosti při zachování metodologických postupů zůstává reálnou, užijí-li se ikonografické soubory, které vznikly z evidenčních důvodů v krátkém časovém intervalu a zachytily veřejný prostor v relativně ucelené podobě (kupř. snímkování pro potřeby asanačních plánů měst, pořizované zejména v letech 1948–1989). Cílem předkládaného příspěvku je pokus představit na příkladu jazykové krajiny výzkumný potenciál interdisciplinárního přístupu propojujícího odborné metody onomastiky (potažmo sociolingvistiky) a historie. Centrem pozornosti předkládaného příspěvku budou urbanonyma a způsob jejich prezentace ve veřejném prostoru v kvalitativním i kvantitativním kontextu jazykové krajiny. Primárním pramenem budou historické fotografie pořízené za různým účelem a různými autory v dlouhodobém časovém horizontu od rozšíření fotografie v 50. letech 19. století. Jejich význam v tomto případě přerůstá funkci prostého pomocného zaznamenání terénní situace (která zůstává pramenem primárním) a stávají se nositelkou již neexistující reality. Perspektivnost tohoto přístupu vzrůstá díky postupující digitalizaci sbírek v paměťových institucích a tedy stále snazší dostupnosti zdrojového pramenného materiálu.

Zvolený přístup kontextualizace urbanonym s „historickou jazykovou krajinou v minulosti“ skýtá několik metodické úskalí: rekonstrukce prostřednictvím historického fotografického materiálu je limitována nahodilým a neúplným dochováním snímků, které nebyly pořízeny k jedinému okamžiku a na přesně vymezeném prostoru. Naopak se vyznačují prostorovou i časovou heterogenitou: zachycená místa jsou značně variabilní, fotografie se pořídily v širokém časovém intervalu a je nutné mít zasvěcenou představu o jejich dataci, tvůrcích atd. Zmíněný přístup a priori determinuje vyvážené užití kvalitativního (Blommaert, 2022) a kvantitativního (Backhaus, 2007) přístupu, díky historickým metodám a přístupům však lze indexikalitu a „odkrytí příběhu“ (Hult, 2009) přinejmenším do značné míry rekonstruovat zapojením tradičních historických pramenů (archivní materiál, zmínky v dobovém tisku), reliktní hmotné kultury (staré uliční tabulky, vývěsní štíty) i ojedinělé terénní pozůstatky (uliční nápisy v omítce na fasádách atd.).

Nastíněné možnosti a limity budou aplikovány na vybrané lokalitě – okresním městě Sušice v jihozápadních Čechách. Pro kvalitativní výzkum jde o volbu výhodnou díky autorově znalosti lokálních dějinných souvislostí a podrobně prozkoumaným a zpracovaným uplatněním fotografie v místě (Lhoták, 2010), které také určilo časové ohraničení výzkumu lety 1873–1918. Výzkumný potenciál výběru je posílen polohou města v těsné blízkosti historické česko-německé jazykové hranice v podhůří Šumavy, takže tu uplatnění prezentace urbanonym a různojazyčných nápisů ve veřejném prostoru podléhalo politicko-nacionálnímu tenzím (na příkladu Jeseníku Klemensová, 2019), interpretovaným dosud výlučně historickou metodologií (skladba obecní reprezentace, obcovací řeč, spolkový život atd.) bez zohlednění každodenně utvářeného veřejného prostoru. Multinacionální projevy v jazykové krajině jsou přitom tradičními objekty výzkumu (z domácích prací Klemensová, 2017; Mácha, 2020; Váňa 2022). Vzhledem k tomu, že ve zvolené době byly ještě poměrně zřetelné morfologické struktury a vizuální hierarchie města (David, Klemensová, 2020), vymezily se pro potřeby analýzy dvě oblasti – město v hradbách (po stránce jazykové krajiny nejbohatší) a předměstí. Vlastní zpracování bude následkem povahy zdrojového pramenného materiálu kombinovat tradiční kvantitativní pojetí s kvalitativními hledisky (jejichž plně efektivní vyhodnocení je částečně omezeno časovým odstupem) sledováním indexikalit diachronní optikou (Pavlenko – Mullen, 2015).

## **Bibliografie:**

- BACKHAUS, P. 2005. Signs of Multilingualism in Tokyo. A Diachronic Look at the Linguistic Landscape. In: *International Journal of the Sociology of Language* 2005, roč. 175/176, s. 103–121.
- BACKHAUS, P. 2007. *Linguistic Landscapes. A Comparative Study of Urban Multilingualism in Tokyo*. Clevedon: Multilingual Matters.
- BLOMMAERT, J. 2022. *Jazyková krajina jako kronika komplexity*. Praha: Karolinum.
- BLOMMAERT, J. – DONG, J. 2020. *Ethnographic Fieldwork. A Beginner's Guide*. Bristol: Multilingual Matters.
- DAVID, J. - KLEMENSOVÁ, T. 2020. Mnohost v jazykové krajině – od morfologie přes gramotnost k diskuzím o vizuálním smogu. In: *Od textu k překladu XIV. 2. část. Linguistic Landscape*. Ed. Alena Ďuricová. Praha: Jednota tlumočnicků a překladatelů, s. 6–15.
- FRÝDL, H. 2020. *Rozbor jazykové krajiny v centru Hrádku nad Nisou*. Bakalářská práce. Praha: Filozofická fakulta UK.
- HULT, F. 2009. Language Ecology and Linguistic Landscape Analysis. In: *Linguistic Landscape. Expanding the Scenery*. Ed. E. Shohamy – D. Gorter. New York: Routledge, s. 88–103.
- KLEMENSOVÁ, T. 2017. V Gruntě a Špicberk – toponyma německého původu v pohraničí v současnosti (na příkladu vybraných obcí Jesenicka). *Naše řeč* 100, s. 300–311.
- KLEMENSOVÁ, T. 2019. Böhmisch Gasse, Schillerstrasse, Titova, Školní - one street, many names. Urbanonyms in the Czech borderlands during the 20th century. In: *Acta onomastica* 60, s. 45–62.
- KOMOROWSKA, H. 2014. Analyzing Linguistic Landscapes. A Diachronic Study of Multilingualism in Poland. In: *Teaching and Learning in Multilingual Contexts. Sociolinguistic and Educational Perspectives*. Ed. Agnieszka Otwinowska – Gessica De Angelis. Bristol-Buffalo-Toronto: Multilingual Matters, s. 19–31.
- LHOTÁK, J. 2010. *Dějiny fotografie v Sušici*. Sušice : Muzeum Šumavy.
- MÁCHA, P. 2020. Dvojjazyčné nápisy na Těšínsku aneb těšínský venkov jako experiment s pluralitou. *Národopisná revue* 30, s. 191-201.
- PAVLENKO, A. 2010. Linguistic Landscape of Kyiv, Ukraine: A Diachronic Study. In: *Linguistic Landscape in the City*. Edd. Elana Shohamy, Eliezer Ben-Rafael, Monica Barni. Bristol: Blue RidgeSummit, s. 133–150.
- PAVLENKO, A. - MULLEN, A. 2015. Why diachronicity matters in the study of linguistic landscapes. In: *Linguistic Landscape* 2015, roč. 1, s. 114-132.
- TUFI, S., 2022. Situated spatialities and the linguistic landscape: a diachronic account of an emblematic square in Naples. In: *Social Semiotics* 2022, roč. 32, s. 240–261.
- VÁŇA, O. 2022. *Vícejazyčnost v jazykové krajině města Písku*. Bakalářská práce. Praha: Filozofická fakulta UK.
- WITTLICH, F. 2012. *Fotografie – přímý svědek?! Fotografický obraz a jeho význam pro historické a jeho význam pro historické poznání*. Praha: Nakladatelství Lidové noviny.

**Toponyma jako zbraň: role mapového názvosloví  
v česko-německých národnostních sporech v 19. a 1. polovině 20. století**

Jitka Močičková (Historický ústav AV ČR)

Zejména od druhé poloviny 19. století došlo v prostoru střední Evropy k výraznému rozvoji žánru národnostních map, a to v souvislosti s rozvojem geografie jako vědní disciplíny (pronikavý nástup tematické kartografie), technologickým pokrokem a zavedením pravidelných sčítání lidu, která poskytla cenná (byť v mnoha ohledech diskutabilní) statistická data pro zpracování map. Do nárůstu významu a produkce etnických map se však výraznou měrou promítl rovněž silící nacionalismus a odstředivé tendence formujících se středoevropských národů. Vedle oficiálních státem vydávaných děl pozoruhodnou měrou narostl podíl soukromé kartografie a publikování rozsáhlého množství kartografických dokumentů různé kvality, které měly široký dosah a oslovovaly nejširší vrstvy populace. V průběhu turbulentního 20. století tento trend ještě zesílil: národnostní mapy, které se dostaly např. i do školních atlasů, často sloužily jako ideologický nástroj (a „důkazní materiál“) k prosazování mocenských ambicí států i jednotlivých etnických skupin a jsou tak pozoruhodným a dosud nedostatečně prozkoumaným pramenem k moderním politickým, sociálním i kulturním dějinám.

Příspěvek se na základě digitalizovaného korpusu cca 350 map znázorňujících česko-německou národnostní hranici (1800–1945) zaměří na jeden z aspektů kartografické tvorby, a to roli toponym a andronym v etnické kartografii a jejich manipulativnímu využití pro ideologické účely. Podrobná analýza kartografických dokumentů umožní představit některé strategie a vzorce myšlení, které autoři do map více či méně vědomě promítali ve snaze ovlivnit veřejné mínění. Komparovány budou mapy české, německé a rakouské provenience vzniklé v tomto období, přičemž bude v synchronní i diachronní perspektivě sledováno mj. užívání exonym, dvojjazyčných názvů a další specifika etnografických map. Jedním z nich je pojmenování etnik na mapách, které nabízelo prostor k poměrně zásadním manipulacím a prezentaci různých idejí o etnické diverzitě obyvatelstva ve středoevropském prostoru (pro české prostředí např. spojení nebo naopak oddělení Čechů a Moravanů v polovině 19. století a o něco později otázka Čechů a Slováků jako jednoho národa/kmene).

## Bibliografie

Black, Jeremy, *Maps and Politics*. Chicago, University of Chicago Press, 1997.

Boria, Edoardo, *Carte come armi: geopolitica, cartografia, comunicazione*. Roma: Edizioni Nuova Cultura, 2012.

David, Jaroslav - Mácha, Přemysl, Možnosti studia toponym ve vztahu ke krajině, prostoru a identitě. *Acta onomastica* 53 (2012), s. 28–45.

Hagen, Joshua, 'Mapping the Polish Corridor: ethnicity, economics and geopolitics', *Imago Mundi* 62:1 (2010): 64.

Holubec, Stanislav - Močičková, Jitka: Zobrazování česko-německé národnostní hranice na mapách v období cca 1820-1940, *Historická geografie* 46:2 (2020), s. 283–340.

Holubec, Stanislav – Močičková, Jitka. Ethnic Mapping in Central Europe, 1810–1945: The Case of the Czech–German Language Border, *Imago Mundi*, 75:1 (2023), 45-71, DOI: [10.1080/03085694.2023.2223068](https://doi.org/10.1080/03085694.2023.2223068)

Jordan, Peter, Some Considerations on the Function of Place Names on Maps. *ICC Proceedings*, 2009, online: [https://icaci.org/files/documents/ICC\\_proceedings/ICC2009/html/nonref/12\\_2.pdf](https://icaci.org/files/documents/ICC_proceedings/ICC2009/html/nonref/12_2.pdf)

Močičková, Jitka – Holubec, Stanislav – Justová, Petra, Národnostní diverzita na mapách českých zemí – kartografický obraz a realita, 2020–2023; webový portál, online: <https://cha.fsv.cvut.cz/web/NarodnostniMapy/>

Perono Cacciafoco, F., Cavallaro F., Toponymy and Cartography. In: *Place Names: Approaches and Perspectives in Toponymy and Toponomastics*. Cambridge University Press; 2023, s. 242–265.

**Možnosti a meze korpusového výzkumu proprií**

Klára Osolsobě, Hana Žižková (Masarykova univerzita)

V příspěvku bychom chtěli na základě zkušeností s českými jazykovými korpusy ukázat meze a možnosti výzkumu proprií, a to s ohledem na stav morfologického značkování užívaného v českém prostředí. Objasníme, jak jednotlivé kroky automatické morfologické analýzy ovlivňují stav lemmatizace a značkování v případě proprií. Dotkneme se problému tokenizace a víceslovných proprií, problému doplňování morfologického slovníku ve vztahu k propriím, zvláštností flexe proprií a jejich homonymie s apelativy ve vztahu k značkování a disambiguaci. Upozorníme na případy, kdy není vhodné při výzkumu spoléhat na morfologické značkování, na konkrétních příkladech ukážeme zkreslení výzkumných dat způsobené chybným morfologickým značkováním. Nastíníme možnosti, jak se zkreslení analyzovaných dat vyhnout. V příspěvku dále ukážeme možnosti využití různých počítačových nástrojů na konkrétních příkladech onomastického výzkumu. Představíme rozdíly v použití získávání dat z korpusů ČNK, SketchEngine, Aranea a malých specializovaných korpusů. Ukážeme možnosti složitějších CQL dotazů při třídění dat.

Zamyslíme se nad budoucností zpracování lingvistických dat v době převratných změn v oblasti strojového učení.

## Rozpoznávání pojmenovaných entit v básnických textech

Petr Plecháč<sup>1</sup>, Silvie Cinková<sup>2,1</sup>, Robert Kolár<sup>1</sup>  
 (<sup>1</sup>Ústav pro českou literaturu AV ČR, <sup>2</sup>Univerzita Karlova)

Příspěvek představí aktuálně řešený projekt „Evropská poezie: distant reading“ (GA23-07727S), jehož cílem je (mimo jiné) detekce pojmenovaných entit (NER) a jejich propojení se znalostní bází v sedmi rozsáhlých korpusech básnických textů (*cs, de, en, fr, hu, it, ru*). Dále ukážeme několik příkladů, jak lze získaná data vytěžovat pro potřeby literárněhistorického výzkumu.

Při výběru metod NER jsme vycházeli z testování oproti zlatým standardům sestávajících ze 100 náhodně vybraných básní z každého korpusu. V nich byly manuálně anotovány entity typu LOC (toponyma) a PER (antroponyma); zbylé dva standardní typy ORG a MISC byly považovány za v dané doméně irelevantní.

Úspěšnost jsme testovali u třech NER nástrojů (komplexnější tagsety byly zjednodušeny na výše uvedené dva relevantní typy):

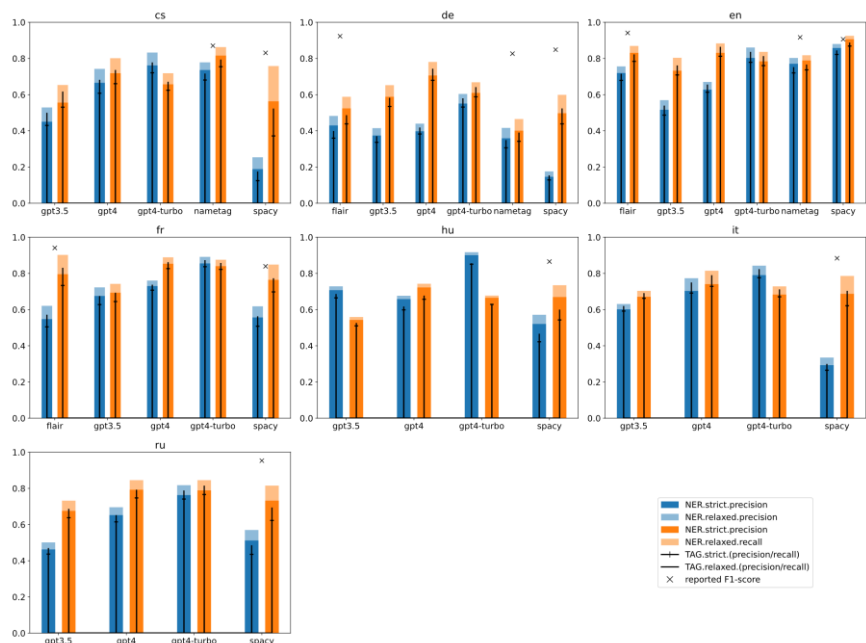
- spaCy (Honnibal et al., 2020); model *en\_core\_web\_trf* pro *en*, mnohojazyčný model *xx\_ent\_wiki\_sm* pro *cs*, *{lang}\_core\_news\_lg* pro ostatní jazyky
- NameTag 2 (Straková, Straka, & Hajič, 2019); dostupné modely pro *cs, de a en*
- Flair (Akbik et al., 2019); dostupné modely pro *de, en a fr*

Vedle toho jsme vzorky proznačkovali i pomocí few-shot promptování třech modelů GPT: GPT 3.5 Turbo, GPT 4 a GPT 4 Turbo.

Výsledky (OBR. 1) ukazují očekávatelný a už dříve pozorovaný efekt (srov. Foley, 2019, p. 41–47): při aplikaci na doménu poezie dosahují NER nástroje významně nižší úspěšnost než u standardních testovacích sad sestávajících z neveršovaných soudobých textů. To lze přičíst nejenom specifickému lexiku a slovosledu, ale i časté kapitalizaci počátečních slov veršů, případně nestandardním kapitalizacím uvnitř verše („*tep srdcí, zrychlen Blížením*“). GPT 4 a GPT 4 Turbo dosahují ve srovnání se standardními nástroji lepší úspěšnost – výjimku tvoří *cs* a *en* – přičemž GPT 4 obvykle překonává GPT 4 Turbo ve výtěžnosti (recall), GPT 4 Turbo naopak významně vyniká v přesnosti (precision).

S ohledem na výše uvedené jsme proznačování celého korpusu *cs* provedli pomocí NameTag 2, proznačování *en* pomocí spaCy a ostatní korpusy proznačkovali pomocí GPT 4 Turbo (ten jsme před GPT 4 upřednostnili zejm. z finančních důvodů).

Aby bylo možné s detekovanými entitami pracovat napříč korpusy, je třeba je propojit se společnou znalostní bází. Pro tyto potřeby jsme zvolili Wikidata. Tak lze nejen určit, že např. *Řím, Rome, Rom, Róma, Roma, Puum* odkazují k jediné entitě, ale i získat další strojově čitelná metadata (např. že se jedná o město v Itálii rozkládající se někde okolo 41°53'35" s.š. a 12°28'58" v.d.). Úspěšnost dostupných nástrojů na propojování entit (NEL) se ovšem u básnických textů ukázala být mizivá. Nezbylo tak, než se vydat náročnější cestou: každé unikátní pojmenované entitě jsme provizorně přiřadili první položku, kterou Wikidata API na příslušný dotaz vrátila. Propojení každé entity je pak manuálně kontrolováno a korigováno. (S ohledem na množství dat (TAB. 1) jsme se prozatím omezili na entity typu LOC. V této chvíli je zpracováno cca. 50 % *cs* a *en*, 10 % *de, ru* a *hu*.)



OBR. 1: Úspěšnost rozpoznávání pojmenovaných entit v náhodných vzorcích 100 básnických textů. NER.strict: přesná shoda anotace bez ohledu na typ. NER.relaxed: překrývající se anotace bez ohledu na typ. TAG.strict: přesná shoda anotace a shodný typ. TAG.relaxed: překrývající se anotace a shodný typ.

korpus	# básní	LOC		PER	
		# výskytů	# unikátních entit	# výskytů	# unikátních entit
cs	80229	43661	10322	186682	31840
de	53133	40404	10548	76629	18508
en	40735	55358	12439	138199	29687
fr	18226	24646	5046	41715	9504
hu	12955	9568	3813	15900	5945
it	4689	9478	2652	14602	3866
ru	45563	35438	9561	56329	16034

TAB. 1: Pojmenované entity rozpoznané v jednotlivých korpusech.

## Bibliografie

- Akbik, A., Bergmann, T., Blythe, D., Rasul, K., Schweter, S., & Vollgraf, R. (2019). FLAIR: An easy-to-use framework for state-of-the-art NLP. In NAACL 2019, 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations) (pp. 54–59). <https://doi.org/10.18653/v1/N19-4010>
- Foley, J. (2019). Poetry: Identification, Entity Recognition, and Retrieval [dizertační práce]. University of Massachusetts Amherst. <https://maroo.cs.umass.edu/getpdf.php?id=1355>
- Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python. Zenodo. <https://doi.org/10.5281/zenodo.1212303>
- Straková, J., Straka, M., & Hajič, J. (2019). Neural Architectures for Nested NER through Linearization. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 5326–5331). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1527>

## Vlastní jména v perspektivě kvantitativní analýzy divadelních her

Petr Pořízka (Univerzita Palackého)

Tento příspěvek metodologického charakteru se věnuje možnostem kvantitativních analýz dramát se zřetelem k roli vlastních jmen, ale i způsobům jejich využití pro interpretaci textu, jeho motiviky, struktury či užitých jazykových prostředků. Proprium chápeme jako klíčový prvek jak z hlediska vhodného zpracování zdrojového textu (oddělení textu a metatextu), tak následných analýz. Dramatická postava, jež je vlastními jmény reprezentována, je v tomto literárním žánru chápána jako ústřední prvek vyprávění (Todorov, 1977; Fořt, 2008). Propria se tak stávají nejen důležitým objektem, ale zároveň i prostředkem kvantitativně založené lingvistické analýzy: strukturace textu je na replikách jednotlivých (literárních) postav založena a reflexe této strukturace je i základem našich analýz, jež budeme demonstrovat na vybraných dramatech bratří Čapků (zejm. *R.U.R.*, *Bílá nemoc*, *Loupežník*, *Země mnoha jmen*). Konkrétně se zaměříme na analýzu klíčových slov (KWs, Scott – Tribble, 2006) umožňující sledovat motiviku her, a to prostřednictvím metody TF\*IDF (Rajaraman – Ullman, 2011), s následnou klasifikací extrahovaných výrazů do různých lexikálně-sémantických tříd (podle jejich jazykového charakteru v dané hře), vyjádřením jejich lexikální disperze na základě vzájemných vztahů literárních postav, jež tyto klíčové výrazy užívají. Tyto třídy KWs budou doplněny o analýzu sentimentu (Python, NLTK & vader\_lexicon, Hutto – Gilbert, 2014), jež místy přináší zajímavé výsledky – srov. Tab. 1, kdy má kladný hrdina (Galén) z *Bílé nemoci* celkové skóre negativní, zatímco hrdina záporný (Maršál) skóre kladné, což je mj. dáno tím, jaké klíčové výrazy tyto postavy ve svých narativních používají. Analýza sentimentu mj. dokládá fakt, že činy nebo záměry literární postavy z hlediska vyprávění nemusejí odpovídat emocionálnímu náboji slov, jež postava ve vyprávění pronáší (hodnocení sentimentu je výsledkem dílčích hodnot výrazů dané postavy). Role literární postavy ve vztahu k ostatním participantům má klíčový význam, i pokud jde o interpretaci „narativity“ postav na základě jejich interakcí a vztahů (Trilcke a kol., 2015). Tyto interakce lze z hlediska utváření vyprávění modelovat prostřednictvím sítí a uzlů, kde uzel reprezentuje postavu (proprium) a hustota sítě narativní komplexitu dramatu – srov. Obr. 1. Pokud se postavy objevují ve stejném aktu (scéně nebo jednání), jsou vzájemně propojené. Hustší síť znamená vyšší míru zapojení postav do vyprávění a vyšší počet spojení s ostatními postavami v inter-dialogu.

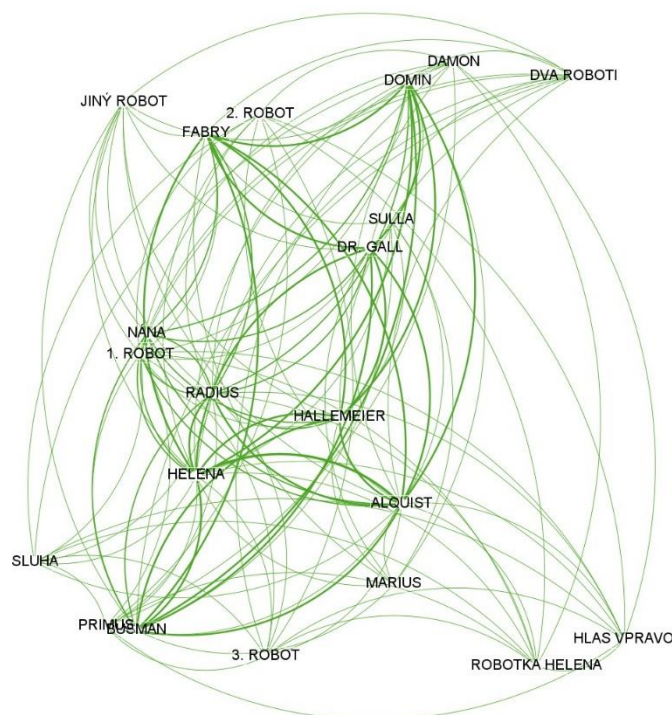
Využijeme i metodu vícerozměrného škálování dat, konkrétně faktorovou analýzu klastrů (Hendl, 2012), jež vyjadřuje podobnosti či rozdílnosti mezi porovnávanými entitami (např. autor, text, žánr, postava, gramatický jev atp.) mírou vzdálenosti objektů. Kromě tradičních jednotek jako slovní tvar či lemma, ukazuje ilustrativní analýza hry *R.U.R.* užití slovních druhů ve vztahu k postavě tendence k nominálnějšímu (fakta, substance) vyjadřování (Fabry, Domin) ve srovnání s dynamičtějším, verbálním jazykem (Alquist) atp. Vlastní jméno jako nositel/prostředek konstruující narativ reprezentuje i rozdílné jazykové prostředky a vyjevuje odlišnou strategii dialogů u různých autorů. Patrné je to u jazyka postav různých sociálních vrstev. Zatímco Josef Čapek tíhne až k jakési jazykové exkluzivitě, kdy promlouvají spisovným kódem, až knižním stylem všechny postavy (např. dialog metařů v *Země mnoha jmen*), bratr Karel užívá obecně autentičtější a hovorovější styl dialogů; využívá princip jazykového kontrastu a různé jazykové (stylistické) variety (zejm. kontrast spisovnosti a nespisovnosti) volí podle funkčního rozvrstvení z hlediska sociální příslušnosti. Pro jazyk nižších vrstev (služebnictvo apod.) volí nespisovný kód, hojně s prvky obecné češtiny, u postav společensky výše postavených či vzdělanějších užívá spisovnou češtinu, příp. zvýší kontrast poetickými dialogy hlavních postav ve srovnání s prozaickými replikami postav vedlejších (Mimi vs. Fanka v *Loupežníkovi*).

Závěrem: Proprium jako nástroj a centrální jednotka analýzy vede k vyšší analyticko-interpretaci granularitě či detailnosti. Je zároveň strukturálně anotačním prvkem i klíčovým objektem na různé úrovni textové či jazykové analýzy a interpretace.



postava	nemoc	specifika	abstrakta	konkrétna	celkové skóre
<b>Dvorní rada</b>	-0.027	0.0	0.089	0.009	<b>0.018</b>
<b>Maršál</b>	0.123	0.072	0.0	-0.071	<b>0.031</b>
<b>Galén</b>	-0.036	-0.115	0.046	0.0	<b>-0.026</b>
<b>Baron Krüg</b>	-0.183	0.014	0.106	0.293	<b>0.058</b>

**Tab. 1** Analýza sentimentu klíčových slov hlavních postav dramatu *Bílá nemoc* (K. Čapek); klíčová slova jsou rozdělena do tříd na abstrakta, konkrétna, specifické výrazy a výrazy vážící se k hlavnímu tématu – nemoci



**Obr. 1** Síť vztahů založená na inter-dialogu (interakci) mezi postavami hry *R.U.R.* (K. Čapek)

## Bibliografie

- Čermák, F. et al. (2007): *Slovník Karla Čapka*. Praha: NLN.
- Fořt, B. (2008): *Literární postava. Vývoj a aspekty naratologických zkoumání*. Praha.
- Hendl, J. (2012): *Přehled statistických metod. Analýza a metaanalýza dat*. Praha: Portál.
- Hutto, C. J. – Gilbert, E. E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014.
- NLTK :: Natural Language Toolkit* [software]. Dostupné z www: <https://www.nltk.org/>.
- Rajaraman, A. – Ullman, J. D. (2011): Data Mining. In Leskovec, J. et al., *Mining of Massive Datasets*. 1–17.
- Scott, M. – Tribble, Ch. (2006): *Textual Patterns. Key words and Corpus Analysis in Language Education*. Amsterdam – Philadelphia.
- Todorov, T. (1977): *The Poetics of Prose*. Ithaca – New York: Cornell university Press.
- Trilcke, P. – Fischer, F. – Kampkaspar, D. (2015): Digital Network Analysis of Dramatic Texts. In: *DH2015: »Global Digital Humanities«*. University of Western Sydney.
- The R Project for Statistical Computing: R* (version 4.0.4) [Software]. Dostupné z www: <https://www.r-project.org/>.
- Textometrie project: TXM* (version 0.8.1) [software]. Dostupné z www: <http://textometrie.enslyon.fr/>.

## Kapitulace na konci světa: kvantitativní analýza jmen v současných francouzských dystopiích

Karel Střelec, Michal Místecký (Ostravská univerzita)

Príspevek se zaměřuje na kvantitativnělingvistický rozbor jmen postav ve dvou současných francouzských dystopických románech – v textu *2084: Konec světa (2084 : La fin du monde)* z pera Boualema Sansala a *Podvolení (Soumission)* Michela Houellebecqa. Texty jsou analyzovány ve francouzském originále, přičemž z románů byly odstraněny paratexty (např. úvodní strana, titíž, reklamní texty, poděkování, epilog; v případě Sansalova textu však byly ponechány rozvinuté názvy kapitol). Metodologicky se výzkum opírá o frekvenční a kolokační analýzu (cf. Brezina, 2018; David et al., 2022) a usiluje o detekci rozdílů mezi oběma knihami v rovině práce s vlastními jmény. Texty jsou lemmatizovány v programu Python, protože dostupné zabudované lemmatizátory (např. v programu LancsBox) byly shledány poměrně nepřesnými; lemmatizace, k níž bylo využito knihovny SpaCy pro francouzský jazyk, byla však přesto manuálně korigována – separátně byly lemmatizovány elidované tvary (např. *d' > de*, *c' > ce*, *s' > se*, *puisqu' > puisque*) a kontrakce (např. *du > de + le*, *aux > à + les*). K samotnému výzkumu je využito víceúčelového lingvistického softwaru LancsBox (Brezina et al., 2015). Výsledky jsou prezentovány formou tabulek a kolokačních sítí (Brezina & Pořízka, 2021).

Na datech byly doposud provedeny tři předvýzkumy – frekvenční analýza vlastních jmen postav, kolokační analýza vybraných antroponym a analýza klíčových slov (cílovým korpusem byl Houellebecqův román, referenčním text Sansalův a statistikou Kilgariffův SMP – viz Davidová Glogarová & Kubát, 2020). Frekvenční analýza ukázala, že Sansal využívá jmen postav statisticky signifikantně více než Houellebecq ( $\chi^2 = 212$ ,  $df = 1$ ,  $p \lll 0,05$ ), což souvisí s narativními preferencemi obou autorů – Houellebecqův román volí *ich*-formu, kdežto Sansalův si od postav drží větší odstup pomocí *er*-formy. V *Podvolení* se navíc objevuje více postav, ale méně frekventovaných, je tak možno dovozovat, že vypravěč strhává pozornost spíše na sebe.

Kolokační analýza probíhala pouze na jménech postav, která se v daném textu objevovala s frekvencí vyšší než 100. Tuto podmínku splnily pouze postavy z románu Boualema Sansala, a to konkrétně Ati (hlavní postava, hledač pravdy), Abi (prorok), Yölah (uctívané božstvo) a Koa (Atiho přítel). Ve fázi předvýzkumu jsme se zaměřili na kolokáty, které se vyskytují v rozmezí -5/+5 (pět slov nalevo a pět slov napravo od daného antroponyma) s minimální frekvencí 5. Z analýzy vyplývá, že nejrozvinutější postavou je Ati (viz obr. 1), s nímž se pojí kognitivní a volní slovesa, která dokumentují jeho komplexní myšlení a touhu po poznání („comprendre“ = rozumět, „savoir“ = znát, „vouloir“ = chtít, „pouvoir“ = moci, „demander“ = požadovat), zatímco Abi a Yölah se objevují především ve fixovaných frázích (např. „Yölah est grand et Abi est son Délégué“ = Yölah je velký a Abi je jeho vyvolený), což poukazuje na rigidnost systému, s nímž Ati bojuje. Koa je z hlediska kolokační analýzy nejméně výrazný; asociuje se především s Atim a slovesy obecného významu („dire“ = říci, „faire“ = dělat).

Klíčová slova, která jsou předmětem třetí analýzy (pracovně jsme cut-off skóre stanovili na SMP = 5), dokládají, že Houellebecq se prominentně zaměřuje na protagonistu (lemmata „je“ = já, „mon“ = můj) a univerzitní prostředí (lemmata „intellectuel“ = intelektuální, „universitaire“ = univerzitní, „thèse“ = závěrečná práce), má tendenci k opatrnému vyjadřování (lemmata „presque“ = téměř, skoro, „probablement“ = pravděpodobně), explicitně operuje s lexikem navázaným na islám (lemmata „islam“ = islám, „musulmane“ = muslimský a postava s arabsky znějícím jménem Ben-Abbes, která se stává francouzským prezidentem) a doslovně také svůj román zasazuje (lemmata „Europe“ = Evropa, „France“ = Francie). Můžeme tedy dovozovat, že Sansalovo *2084* je textem implicitnějším a alegoričtějším. Nejsilnějším klíčovým slovem je však v *Podvolení* vlastní jméno Huysmans, které odkazuje na francouzského spisovatele Jorise-Karla Huysmanse, autora satirického románu *À rebours* (v češtině *Naruby*, 1884), na nějž se Houellebecqova hlavní postava specializuje. Je tak možné spekulovat, že Huysmans je skutečnou hlavní postavou *Podvolení*; a protože se v románu objevují i další spisovatelé, nabízí se interpretace, že prominentními charaktery jsou vlastně oni – text tak působí jako snaha o záchranu/konzervaci těchto osobností v kontextu postupující islamizace Evropy.

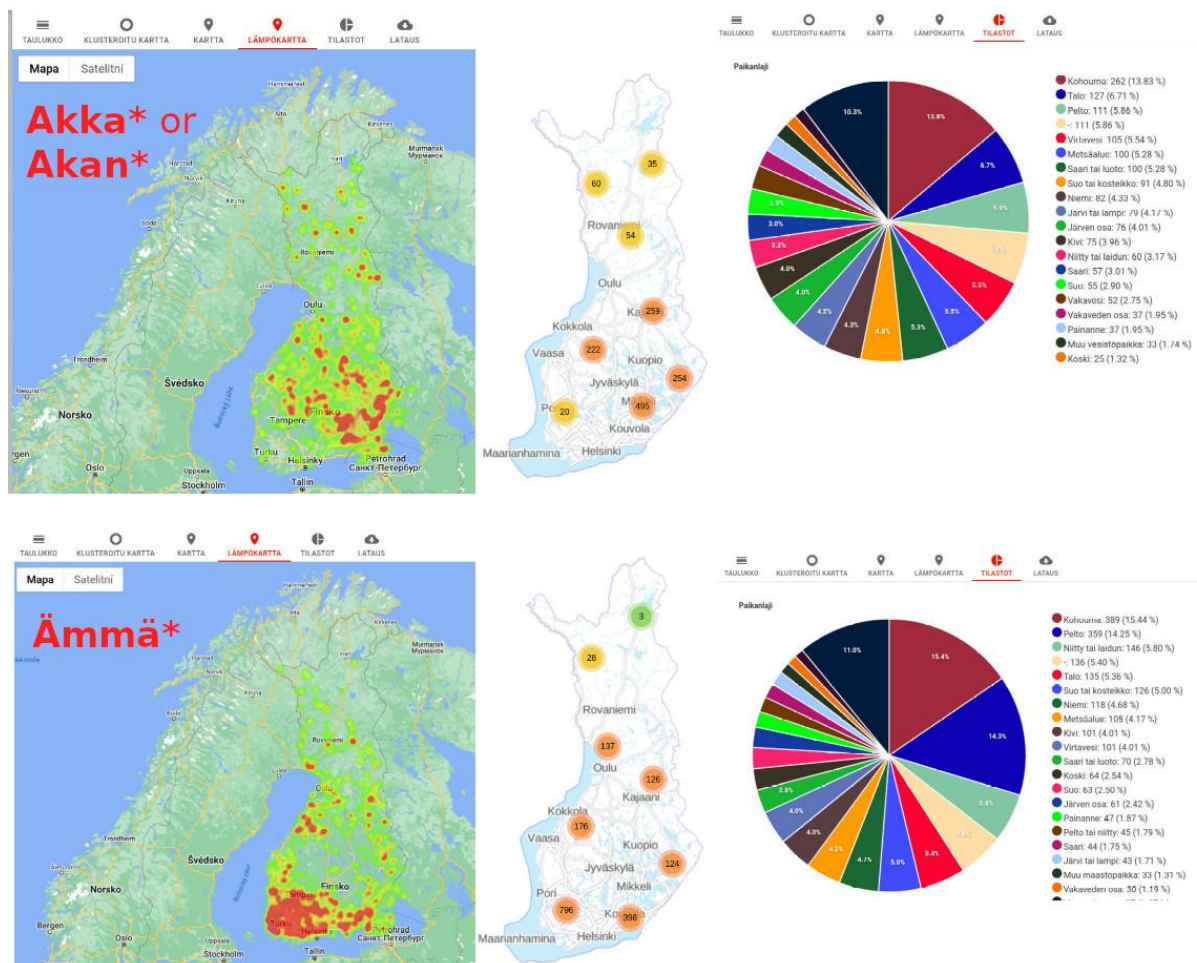
Výzkum bude dále rozvíjen několika směry; především se hodláme zaměřit na kompletní onymickou krajinu zkoumaných děl a neopomíjet další typy vlastních jmen (např. toponyma).



**NimiSampo: Databáze finských toponym online**

Soňa Wojnarová (ÚJČ AV ČR)

Cílem tohoto příspěvku je představit finskou online databázi toponym *NimiSampo* [1] a možnosti jejího využití při studiu materiálu v režimu otevřených dat. Databáze vznikla ve spolupráci mezi toponomastikou a počítačovými odborníky se záměrem vytvořit nástroj, který dokáže propojit, vizualizovat a kombinovat data z různých zdrojů s možností výsledky různě filtrovat, vytvářet mapy a grafy a statisticky je zpracovávat, případně exportovat pro další analýzu, např. v GIS [2]. Databáze kromě toho nabízí také možnost využívat současné i historické mapové podklady a také analyzovat strukturu toponym. Po technické stránce se jedná o sémantický web, který vychází ze standardů W3C [3]. Databáze byla zpřístupněna v roce 2019 a je určena onomastikům a dalším badatelům (genealogům, historikům, archeologům apod.) i laické veřejnosti. Materiál použitý pro tvorbu databáze vychází především z toponym a mapových podkladů shromážděných od 1. poloviny 20. stol. v Archivu jmen (Nimiarkisto, NA), který je součástí Centra pro domácí jazyky ve Finsku (KOTUS, Helsinky). Tento materiál, který je veřejnosti běžně přístupný, představuje korpus obsahující přibližně 2,7 mil. finských, švédských a sámských toponym z Finska a oblastí Karélie, která byla po 2. sv. válce odstoupena SSSR. Kromě toponym a map z NA byly zpracovány některé další sbírky (např. 0,8 mil. toponym a mapové podklady z registru Národního zeměměřičského úřadu, Maanmittauslaitoksen Paikannimirekisteri) [2, 3]. V roce 2017 byl materiál zdigitalizován a ve většině případů se podařilo zaznamenat i polohu místa na mapě. Údaje o toponymu jsou však zaznamenány i jako jednotlivé kategorie (např. typ krajinného prvku, lokalita, jednotlivé části jména a rok záznamu), podle kterých je možné informace dále analyzovat pomocí aplikací v databázi (statistiky či vizualizace dat např. pomocí tzv. heat map, viz Obr. 1). V databázi je možné vyhledávat buď konkrétní jména, nebo pracovat s větším množstvím dat (např. s toponymy obsahující ve své určující či základní části konkrétní výraz nebo zkoumat názvy pro konkrétní krajinné prvky). Dotazy na vyhledávání je možné zadávat pomocí symbolu „\*“ (nahrazující libovolný počet jakýchkoliv znaků před či za jménem) nebo výrazu „or“. Viz příklad v Obr. 1. kdy lze zadat dotaz na výskyt výrazu *akka* ‚baba‘ a jeho genitivního tvaru *akan* v určující části toponyma jako „akka\* or akan\*“.



**Obr. 1. Ukázka vizualizace dat pro výrazy *akka* a *ämmä* (obě s výrazem ‚baba, babizna‘). Pro výraz *akka* byl do vyhledávání zahrnut i genitivní tvar (*akan*). Data je v databázi možné zobrazit jako heat map, klustry nebo graf výskytu podle krajinného prvku. Z vizuálního zpracování je patrné, že výskyt výrazů *akka* a *ämmä* se geograficky liší (v závislosti na rozšířených nářečích). Oba výrazy se vyskytují nejčastěji v názvech vyvýšenin (fin. *kohouma*, viz grafy). (Zdroj: nimisampo.fi, 2024-02-15).**

## Bibliografie

- [1] Databáze *NimiSampo*: <https://nimisampo.fi>
- [2] IKKALA Esko, Jouni TUOMINEN, Jaakko RAUNAMAA, Tiina AALTO, Terhi AINIALA, Helinä UUSITALO, Eero HYVÖNEN: NameSampo: A Linked Open Data Infrastructure and Workbench for Toponomastic Research. *Proceedings of the 2nd ACM SIGSPATIAL Workshop on Geospatial Humanities*, GeoHumanities 18, pp. 2:1-2:9, ACM, Seattle, WA, USA, November, 2018. Dostupné online: <https://seco.cs.aalto.fi/publications/2018/ikkala-et-al-namesampo-2018.pdf> (2024-02-15).
- [3] *Nimisampo: nimistöntutkijan työpöytä ja linkitetyn paikkatiedon infrastruktuuri* [Nimisampo: onomastická databáze a infrastruktura vytvořená spojením dostupných dat o toponymech]. Webová stránka Aalto University (Helsinki): <https://seco.cs.aalto.fi/projects/nimisampo/> (2024-02-15).

5

*Sémantika, pragmatika  
a logika*

*Semantics, pragmatics,  
and logic*

organizují

Mojmír Dočekal

Matej Drobňák

Jiří Raclavský

(CZE, ENG)

Bienále české lingvistiky 2024  
Biennial of Czech Linguistics 2024

## Popis panelu / Workshop description

**(CZE)** Ve 20. století byla lingvistika zásadně ovlivněna logikou; ať už v okamžiku vzniku formální sémantiky (Montague 1974), tak v 70.–90. letech, kdy došlo k formálně-sémantickému inkorporování modální logiky (Kratzer 1977) a dynamickému obratu v sémantice a pragmatice (Kamp a Reyle 1993). Poté, co byly formální sémantika a pragmatika etablovány jako samostatné lingvistické disciplíny, došlo na začátku 90. let k experimentálnímu obratu, který do lingvistiky přinesl nové metody sběru dat a s tím spojené statistické modely nutné k jejich vyhodnocení (Baayen 2008). Tyto formálně-experimentální metody byly aplikovány i na slovanská data (vč. českých), viz řadu konferencí (následovanou sborníky) jako FASL a FDSL. Tato data mnohdy nabyly důležitosti i v současných teoretických lingvistických debatách, ať už v rámci tradičních dominantních témat, jako jsou význam určitých popisů (Šimík a Demian 2020), nebo tzv. teorie pluralit (TP). TP je v návaznosti na rozvoj logických nástrojů pro popis významu gramatického čísla a číslovek, počítatelných/nepočítatelných jmen (Link 1983) jednou z nejvlivnějších současných poddisciplín formální sémantiky (spojení TP a experimentů: Dotlačil 2010, Dočekal a Wągiel 2021). Tento panel navazuje jak na zmíněný trend, tak na českou lingvistiku, ať už jde o tradici výzkumu informační struktury (Sgall, Hajičová, a Panevová 1986) nebo o sémanticky zaměřená zkoumání vět a souvětí (Daneš 1968, Karlík 1995). Tím motivujeme i výběr témat (viz už zmíněné tradiční velké otázky až po témata TP). Zároveň vítáme moderní pragmatická témata, jako jsou výzkum společenského rozměru významu, Bayesiánských modelů interpretace, presupozic/implikatur, ale i formálních modelů osvojování jazyka. Tento výčet témat je nutně neúplný a zůstává otevřený všem příspěvkům z (experimentální) sémantiky, pragmatiky, logiky a filozofie jazyka. Společným jmenovatelem příspěvků budiž zaměření na vztah přirozeného jazyka k lingvistickým/logickým, případně analyticko-filozofickým problémům. Výzkumná otázka daného příspěvku by pak měla souviset s tím, jak zkoumaná jazyková data posouvají naše lingvistická/logická/jazykově-filozofická bádání a jak ukazují, kudy postupovat v místech, kde naše teorie stále ještě selhávají.

**(ENG)** In the 20th century, linguistics was fundamentally influenced by logic, first with the emergence of formal semantics (Montague 1974), then in the 1970s–1990s, with the formal-semantic incorporation of modal logic (Kratzer 1977), followed by the dynamic turn in pragmatics (Kamp and Reyle 1993) and semantics. Then, after formal semantics and pragmatics were established, the early 1990s saw the experimental turn that brought new methods of collecting data and the associated statistical models necessary to evaluate them (Baayen 2008). These formal-experimental methods were also applied to Slavic (and also Czech) data; see many conferences (followed by proceedings) such as FASL and FDSL. These data have then sometimes become important in contemporary theoretical linguistic debates, touching the traditional big topics, such as the meaning of definite descriptions (Šimík and Demian 2020) or the so-called theories of plurality (TP). TP followed the development of logical tools for describing the grammatical number and numerals meaning (Link 1983) and is one of the most influential contemporary subdisciplines of formal semantics (see Dotlačil 2010; Dočekal and Wągiel 2021, a.o.). This panel follows both the trend mentioned above and the Czech linguistics tradition of research on informational structure (Sgall, Hajičová, and Panevová 1986) or the semantic investigations of sentences (Daneš 1968, Karlík 1995). This motivates the choice of topics (see the already mentioned traditional big questions). At the same time, we welcome modern pragmatic topics such as the social dimensions of meaning, Bayesian models of interpretation, presuppositions/implicatures, and formal models of language acquisition. This list of topics is necessarily incomplete and remains open to all contributions from (experimental) semantics, pragmatics, logic, and philosophy of language. The common denominator of the papers should focus on the relation of natural language to linguistic/logical or analytical/philosophical problems. The research question of a given abstract should then address the linguistic data as a tool for moving our linguistic/logical/linguistic-philosophical investigations forward.

## Literatura / References

- Baayen, Harald. 2008. *Analyzing Linguistic Data*. Cambridge: Cambridge University Press.
- Daneš, František. 1968. „Some thoughts on the semantic structure of the sentence“. *Lingua* 21: 55–69.
- Dočekal, Mojmír, a Marcin Waگیel, eds. 2021. *Formal approaches to number in Slavic and beyond*. OSL 5. Berlin: LangSciPress. <https://doi.org/10.5281/zenodo.5082006>.
- Dotlačil, Jakub. 2010. *Anaphora and distributivity*. Netherlands Graduate School of Linguistics.
- Kamp, Hans, a Reyle, Uwe. 1993. *From Discourse to Logic*. Dordrecht: Kluwer Academic Publishers.
- Karlík, Petr. 1995. *Studie o českém souvětí*. Brno: Masarykova univerzita.
- Kratzer, Angelika. 1977. „What ‚Must‘ and ‚Can‘ Must and Can Mean“. *L&P* 1 (3): 337–55. <https://doi.org/10.1007/bf00353453>.
- Link, Godehard. 1983. „The Logical Analysis of Plurals and Mass Terms“. In *Formal Semantics*, editoval P. Portner a B. H. Partee, 127–47. Blackwell.
- Montague, Richard. 1974. „The proper treatment of quantification in ordinary English“. In *Formal Philosophy*, Thomason (ed.). Dordrecht: D. Reidel.
- Sgall, Petr, Eva Hajicová, a Jarmila Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects*. Dordrecht: Springer.
- Šimík, Radek, a Christoph Demian. 2020. „Definiteness, uniqueness, and maximality in languages with and without articles“. *JoS* 37 (3): 311–66.



# Obsah / Table of Contents

*Pluractionality and argument structure*

Petr Biskup (Universität Leipzig)

*The interpretations of the Japanese topic marker -wa from a functional sentence perspective*

Masanori Deguchi (Western Washington University)

*A dynamic semantic analysis of wh-questions*

Jakub Dotlačil (Utrecht University, Masarykova Univerzita)

*Onticky indukovaná sémantická vágnost*

Petr Dvořák (Filosofický ústav AV ČR, Praha)

*Reflexives, person constraints and distributivity in Czech: An experimental study*

Rhiana Horovská (Masaryk University), Radim Lacina (Osnabrück University)

*Constraints on sorting: Experimental evidence*

Aviv Schoenfeld (Ben-Gurion University of the Negev), Michaela Hulmanová (Masaryk University)

*Sémantika dispozičních predikátů*

Filip Tvrдый (Univerzita Komenského v Bratislavě)

*Exploring partitivity and space: Insights from spoken and sign languages*

Lucia Vlášková (Masaryk University in Brno), Marcin Wągiel (Masaryk University in Brno, University of Wrocław, Leibniz-Centre General Linguistics (ZAS))

*Adding individuals and numbers: A unified analysis of nominal conjunction and additive numerals*

Marcin Wągiel (Masaryk University in Brno, University of Wrocław, Leibniz-Centre General Linguistics (ZAS))

## Pluractionality and argument structure

Petr Biskup (Universität Leipzig)

**1. Introduction:** This paper deals with the interaction between pluractionality and the verbal argument structure. It discusses participial adjectives, motion verbs and impersonal constructions, mainly from Czech. The analysis is couched in the generative framework. It is argued that the specific verbal theme *-a-* realizes pluractionality and the agentive or expletive voice, which is responsible for changing the verb class behavior.

**2.** Slavic uses two specific types of **participial adjectives**: *past active*, with *-l-*, and *past passive*, with *-n/t-* (Schoorlemmer 1995, Kosta & Frasek 2004, Veselovská & Karlík 2004). Standardly, *l-* participles are derived from unaccusative stems and *-n/t-* participles from transitives. However, there are certain unaccusative stems co-occurring with *-n/t-*; see (1a). The stem has the plural interpretation (glossed PL), as shown by the contrast between the sg. and pl. patient object:

- (1) a. vy-pad-a-n-é vlas-y b. # vy-pad-a-n-ý vlas  
 out-fall-PL-n/t-pl hair-pl out-fall-PL-n/t-sg hair.sg  
 ‘fallen out hair’

**3.** It has been argued for Russian that directed **motion verbs** are unaccusative, whereas non-directed motion verbs unergative (agentive). An analogous test for Czech shows that the directed verb (2a) behaves unaccusatively and the ungrammatical non-directed verb ((2b), with *-a-*) unergatively.

- (2) Tolik lidí tam a. na-běh-l-o / b.\* na-běh-a-l-o!  
 so.many people there on-run-l-sg.n on-run-PL-l-sg.n  
 ‘So many people ran there!’

In addition, the directed *běží* has an episodic reading, the non-directed *běhá* the plural (generic) meaning:

- (3) Jirka a. běž-í / b. běh-á do školy.  
 Jirka run-TH run-PL to school  
 ‘Jirka is running / runs to school.’

**4.** Contrary to transitives and unergative, the formation of **impersonal constructions** from unaccusatives is restricted; compare the (im)perfective (4a), (4b) with (5a) (also Fehrmann *et al.* 2010).

- (4) a. Už se (do-)četlo. b. (Do-)pracovalo se.  
 already self to-read.l.sg.n to-work.l.sg.n self  
 ‘People already finished reading.’ ‘People finished working.’
- (5) a. \*Umře-l-o se na covid. b. Umír-a-lo se na covid.  
 at.die-l-sg.n self on covid at.die-PL-l-sg.n self on covid  
 ‘People died of covid.’
- b’.# Ale nikdo nezemřel.  
 ‘But nobody died.’

Contrary to the perfective/quantized (5a), the imperfective (5b, with *-a-*) is grammatical. (5b) can only have the plural (habitual) interpretation, but not the progressive one, as shown by the contradictory (5b’).

**5. Analysis:** All three phenomena (2,3,4) display an unexpected argument structure behavior and a plural interpretation. The theme *-a-* brings about plurality (see (1a), (3b), (5b)) and marks changes in the verb class behavior. For the pluralizing head, I assume (6), with the iteration set *E*:

$$(6) [[\mathbf{PL}]] = \lambda P \lambda E \exists e. P(e) \wedge e \in E \wedge |E| > 1 \wedge \forall e'. e' \in E \rightarrow P(e') \wedge \neg \tau(e') \supset \tau(e)$$

The specific *-a-* cannot be just a secondary imperfective or progressive marker; see (3b) and (5b,b’). *-a-* licenses *se*, introduced by Voice (which is expletive in the non-agentive (5b), e.g. Alexiadou *et al.* 2015). *-a-* also verbalizes roots; compare *běh* ‘the run’ with (3b). Hence, *-a-* spells out more heads (cf. Ramchand 2008): verbalizing *v*, pluralizing PL and Voice (agentive (3b) or expletive (1)). Participle head is spelled out as *-n/t-* if it merges with a complement containing Voice (expletive (1a) or agentive, as in *čtený* ‘being read’), otherwise as *-l-*.

## Bibliography

- Alexiadou, Artemis, Elena Anagnostopoulou and Florian Schäfer. 2015. *External Arguments in Transitivity Alternations: A Layering Approach*. Oxford: Oxford University Press.
- Fehrmann, Dorothee, Uwe Junghanns and Denisa Lenertová. 2010. Two reflexive markers in Slavic. *Russian Linguistics* 34, 203–238.
- Kosta, Peter and Jens Frasek. 2004. Neakuzativita (ergativita) vs. neergativita v češtině, polštině a jiných slovanských jazycích na rozhraní morfologie a syntaxe. In Zdeňka Hladká and Petr Karlík (eds.), *Čeština - univerzália a specifika 5*. Praha: Nakladatelství Lidové noviny, 189–212.
- Ramchand, Gillian C. 2008. *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge: Cambridge University Press.
- Schoorlemmer, Maaïke. 1995. *Participial Passive and Aspect in Russian*. (Ots dissertation series.) Utrecht University.
- Veselovská, Ludmila and Petr Karlík. 2004. Analytic Passives in Czech. *Zeitschrift für Slawistik* 49, 163–235.

The interpretations of the Japanese topic marker *-wa* from a functional sentence perspective

Masanori Deguchi (Western Washington University)

**Background** The so-called Japanese topic marker *-wa* lends itself to two distinct readings: thematic and contrastive readings [4,5 *inter alia*]. The elusive nature of the distribution of these readings is one of the classical problems in Japanese linguistics. In his seminal work on this topic, Kuno (1973), for example, approaches this issue from a semantic perspective. In the present study, I argue for an analysis rooted in discourse. Specifically, I demonstrate that, once two subtypes of contrastive readings are recognized, the distribution of the thematic and contrastive readings follows from the three types of thematic progression (TP) discussed in Daneš (1974).

Kuno (1973) observes that anaphoric (and generic) NPs induce the thematic reading whereas non-anaphoric NPs result in the contrastive reading as illustrated in (1a) and (1b), respectively. Kuno also observes that what is contrasted in (1b) is context dependent (e.g., it is raining, but it will stop soon).

- (1) a. John-**wa**   gakusee       desu  
       John-**TOP**   student       is  
       Thematic: ‘John is a student.’
- b. Ame-**wa**    hutte           imasu-(ga)  
       rain-**TOP**   falling        is-(but)  
       Contrastive 1: ‘It is raining, but ...’
- c. John-**wa**    hon-o        kaimashita-ga   Mary-**wa**    kaimasen   deshita  
       John-**TOP**   book-ACC   bought-but    Mary-**TOP**   buy.not    PAST  
       Contrastive 2: ‘John bought the book, but Mary didn’t.’

Kuno further notes that anaphoric NPs may also induce a contrastive reading as shown in (1c). I refer to the reading as in (1a) the ‘thematic’ reading, the one as in (1b) the ‘contrastive rheme’ reading, and the one as in (1c) the ‘contrastive theme’ reading.

**Proposal** I take the basic function of *-wa* to be marking of a theme (cf. [3]) and argue that the contrastive readings, both the contrastive theme and contrastive rheme readings, are derived from the thematic function of *-wa* due to the notion of “opposition” involved. As the contrast between (1c) and (2a) demonstrates, there needs to be opposition in meaning (e.g., affirmative vs. negative) when two clauses are marked with *-wa*. Example (2b) further confirms that an opposition licenses a second *-wa* in contrastive sentences.

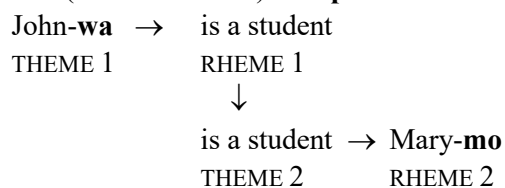
- (2) a. ??John-**wa**   hon-o        kaimashita-ga   Mary-**wa**    kaimashita.  
       John-**TOP**   book-ACC   bought-but    Mary-**TOP**   bought  
       ‘John bought a book, and Mary bought (one) too.’
- b. John-**wa**    hon-o        kaimasita-ga   Mary-**wa**    karimashita.  
       John-**TOP**   book-ACC   bought-but    Mary-**TOP**   borrowed  
       ‘John bought a book, and Mary borrowed (one).’

The main objective of the present paper is to demonstrate that the three distinct readings associated with *-wa* as in (1) follow from Daneš’ (1974) three types of thematic progression (TP). As illustrated in [Figure 1](#) below, the **thematic reading** follows **simple linear TP**, where the rheme of the first clause serves as the theme of the second clause. The **contrastive theme reading**, on the other hand, is reducible to **TP with derived themes**, where the derived themes have rhemes that are in opposition as shown in [Figure 2](#). The **contrastive rheme reading** results from **TP with a continuous theme** as shown in [Figure 3](#), where there is an opposition between rheme 2 and the implicature of rheme 1 (cf. [2]), inducing a contrastive meaning.

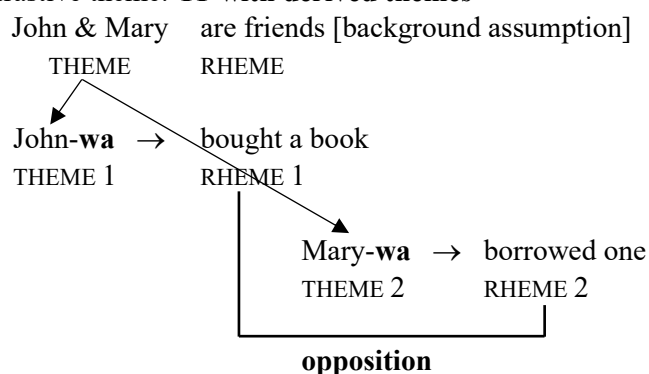
**Conclusion** I argue that *-wa* is thematic semantically and that the contrastive readings are derived in discourse. Specifically, I demonstrate that the three distinct readings associated with *-wa* follow from the three types of TP in the sense of Daneš’ (1974).

## Figures

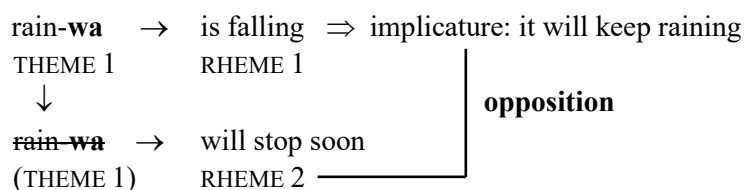
### 1. Thematic (non-contrastive): **Simple linear TP**



### 2. Contrastive theme: **TP with derived themes**



### 3. Contrastive rheme: **TP with a continuous theme**



## Bibliography

- [1] Daneš, František. 1974. Functional sentence perspective and organization of the text. In František Daneš (ed.), *Papers on functional sentence perspective*, 102–128. Prague: Academia.
- [2] Grice, H. P. 1975. Logic and conversation. In Peter Cole and Jerry L. Morgan (eds.), *Syntax and Semantics: Speech acts*, 41–58. New York: Academic Press.
- [3] Jorden, Eleanor Harz. 1987. *Japanese: The spoken language*. New Haven, Conn: Yale University Press.
- [4] Kuno, Susumu. 1973. *The structure of the Japanese language*. Cambridge, Mass.: MIT Press.
- [5] Kuroda, S.-Y. 1965. Generative grammatical studies in the Japanese language. Doctoral dissertation. MIT.

**A dynamic semantic analysis of wh-questions**

Jakub Dotlačil (Utrecht University, Masarykova Univerzita)

In this talk, I will present an analysis of wh-questions (i.e., questions such as *who left?*) that I and Floris Roelofsen developed in Dotlačil & Roelofsen (2019), Dotlačil & Roelofsen (2021) and Roelofsen & Dotlačil (2024). The goal of the analysis was to see how a dynamic framework, in which words and sentences are seen as updates on contexts, can be combined with inquisitive semantics, which analyzes question interpretations. As I will show, the result of the combination, dynamic inquisitive semantics, gives us an analysis that is superior in several respects to previous accounts of wh-questions. Namely, it allows for a uniform treatment of single-wh and multiple-wh questions, it derives mention-some and mention-all readings in a principled way, as well as an often neglected partial mention-some reading in multiple-wh questions. The analysis can also capture the effects of number marking on which-phrases across single-wh and multiple-wh questions.

## Onticky indukovaná sémantická vágnost

Petr Dvořák (Filosofický ústav AV ČR, Praha)

Tradičně byla vágnost považována buď za epistemický jev nebo fenomén reprezentace, tj. neurčitost v poznání nebo pojmová neurčitost, která se projevuje jako neúplně definovaná extenze vágních termínů. V posledních desetiletích však stále více lidí tvrdí, že je třeba do ontologie vpustit vágnost jako součást reality samotné, tzv. onticky vágní objekty. Kdokoli takto argumentuje, musí překonat slavný argument pocházející od G. Evanse, podle kterého je nekonzistentní samotný pojem onticky neurčité identity. Tato nekonzistence se přenáší i na onticky vágní objekty, protože, jak ukazuje Weathersonovský argument, připustíme-li, že existují onticky vágní objekty, musíme automaticky připustit i existenci neurčité identity. Otázkou tedy je, zda může existovat neurčitá identita. Přijmeme-li Evansův argument, pak tvrzení o neurčité identitě mohou být pravdivá, ale pouze taková, v nichž alespoň jeden ze dvou singulárních výrazů nerefereje určitě. Přesto není nutné vysvětlovat veškerou vágnost pouze sémanticky, tj. jako neurčenost významu, protože některá taková tvrzení mohou být pravdivá proto, že je neurčitá skutečnost. To se ukazuje zejména v Loweově kvantovém příkladu o identitě absorbovaného a emitovaného elektronu. Singulární výrazy v tvrzeních o identitě v tomto příkladu je třeba chápat tak, jak naznačuje Abasnezhad a jak to činí Barnesová a Williams v případě, když se tvrdí identita Kilimandžára s jedním přesným agregátem částic, který horu tvoří: jeden z výrazů referuje neurčitě, přičemž tato neurčitost je typem neurčitosti, s níž referuje neurčitá deskripce.

Teze příspěvku je tedy tato: dosavadní diskuse se až na výjimky soustředila na falešné dilema mezi dvěma krajními pozicemi. Zastánci názoru, že v některých případech je vágnost ontická a nikoliv referenční, musejí prokázat neplatnost (*invalidity*) či nepravdivost (*unsoundness*) Evansova argumentu. Navíc nejpřesvědčivější vyvrácení (van Inwagenovo) čelí Bělohradově námitce, že je nekonzistentní. Naproti tomu zastánci názoru, že vágnost není nikdy ontická, ale vždy je jen otázkou reference, čelí argumentům ve prospěch existence ontické vágnosti. Navíc závěry Evansova a Weathersonova argumentu nutně nevedou k podpoře této pozice, ale jsou slučitelné s vágností na ontické úrovni, nakolik je zároveň přítomna i neurčitost referenční. Proto se zdá rozumná určitá syntéza: v některých případech nejde jen o neurčitost reference, ale navíc je zde přítomna i neurčitost věcná, ontická.

## Bibliografie

Abasnezhad, Ali. "How (Not) To Argue Against Vague Object". *Metaphysica* 17, 2 (2016): 195–205.

Barnes, Elizabeth a Williams, John Robert Gareth. "A Theory of Metaphysical Indeterminacy". In *Oxford Studies in Metaphysics* 6, vyd. Karen Bennett a Dean W. Zimmerman, 103–148. Oxford: Oxford University Press, 2011.

Barnes, Elizabeth a Williams, John Robert Gareth. "Vague Parts and Vague Identity". *Pacific Philosophical Quarterly* 90 (2009): 176–187.

Evans, Gareth. "Can There Be Vague Objects?". *Analysis* 48 (1978): 130–134.

Fine, Kit. "Vagueness, Truth, and Logic". *Synthese* 30 (1975): 265–300.

R. Bělohrad, "On Three Attempts to Rebut the Evans Argument against Indeterminate Identity". *Croatian Journal of Philosophy* 19, 55 (2019): 137–152.

Curtis, Benjamin L. a Noonan, Harold W. "Castles Built on Clouds: Vague Identity and Vague Objects". In *Vague Objects and Vague Identity: New Essays on Ontic Vagueness*, vyd. Ken Akiba a Ali Abasnezhad, kap. 15, s. 305–326. Dordrecht: Springer, 2014.

Inwagen, van Peter. *Material Beings*. Ithaca: Cornell University Press, 1995.

Lewis, David. "Many, but almost one". In *Ontology, causality and mind: Essays on the philosophy of D. M. Armstrong*, vyd. K. Campbell a J. Bacon a L. Reinhardt, 23–38. Cambridge: Cambridge University Press, 1993.

Lewis, David. "Vague Identity: Evans Misunderstood". *Analysis* 48, 3 (1988): 128–130.

Lowe, Edward Jonathan. "Ontic Indeterminacy of Identity Unscathed". *Analysis* 61, 3 (2001): 241–245.



**Reflexives, person constraints and distributivity in Czech: An experimental study**

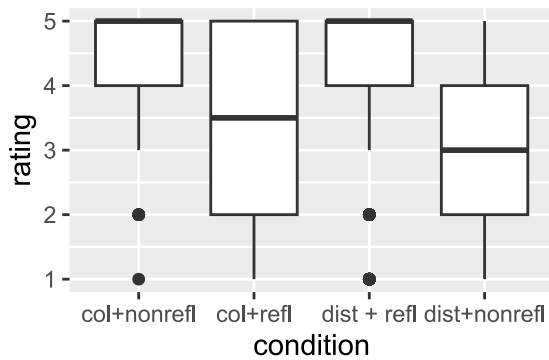
Rhiana Horovská (Masaryk University), Radim Lacina (Osnabrück University)

**Background.** Slavic languages have anaphors of both the argumental and the possessive type. Regarding locality, both kinds must be bound within their minimal finite clause, and split antecedents are never allowed, the tell-tale signs of non-logophoric expressions (Charnavel, 2019). Despite this, the pattern of anaphoric binding in Slavic languages is very intricate and underivable from the standard Binding Theory (Chomsky (1981) and subsequent work). The majority of the previous work on anaphoric binding in Slavic languages has been theoretical and focused on Russian (Padučeva (1983); Zubkov (2018); Reuland and Zubkov (2022) a.o. but see Mertins (2021)). The previous work discovered effects like animacy requirement on the antecedents of non-local binding, interpretive effects (distributivity and awareness), and decreased complementarity between anaphors and pronouns. We add to this research our experimental findings on anaphoric binding in Czech and bring new data and a first experimentally supported generalization concerning the distributive vs. collective predication and its effect on anaphoric binding.

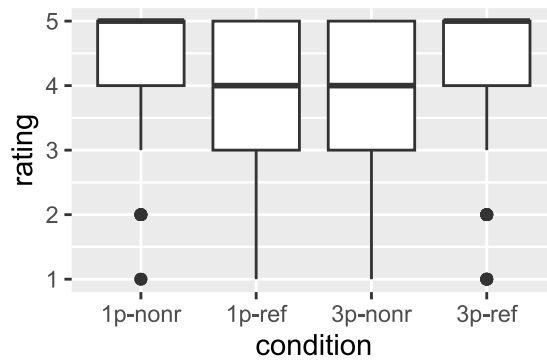
**Method.** We ran two experiments on separate samples of native speakers of Czech. They judged the acceptability of sentences on a scale from 1 (unacceptable) to 5 (acceptable). Both experiments were of the 2×2 factorial design, sharing the factor of +/- reflexivity, which meant that sentences either contained a reflexive or a non-reflexive possessive bound within a plural antecedent's domain, and differing in the other factor. Experiment A (N = 67) additionally examined how the factor of overt collectivity/distributivity affects native speakers' intuitions about the acceptability of both types of possessives. With collectivity, we expected higher ratings for non-reflexive possessives than reflexive ones, but for the distributive items the opposite. Experiment B (N = 65) used the first/third person as the second factor to test whether the acceptability of non-/reflexive binding is affected by person. We hypothesised that in the first person context, non-reflexives would be more accepted than reflexives.

**Results.** For each experiment, a sum contrast-coded linear mixed model was created with the conditions' factors and their interaction as fixed effects and participants and items as random effects. Random slopes were added, but they only converged in Experiment B. In Experiment A, we found two significant effects. Distributivity had a negative effect ( $\beta = -0.19$ ,  $SE = 0.04$ ,  $t = -4.24$ ,  $p < 0.001^{***}$ ). The interaction between distributivity and reflexivity had a positive effect ( $\beta = 0.62$ ,  $SE = 0.07$ ,  $t = 8.94$ ,  $p < 0.001^{***}$ ). In Experiment B, reflexivity had a negative effect ( $\beta = -0.12$ ,  $SE = 0.05$ ,  $t = -2.31$ ,  $p = 0.0256^*$ ). The interaction of first person with reflexivity showed a negative effect ( $\beta = -0.27$ ,  $SE = 0.05$ ,  $t = -5.94$ ,  $p = < 0.001^{***}$ ). Person did not have a significant effect on its own.

**Discussion.** Our study provides new data on Czech binding in plural contexts and comes to the following conclusions. Firstly, non-reflexive possessives are preferred to reflexive ones for in-domain binding, but secondly, the first does not stand if the antecedent is in the third person or if the sentence's proposition is overtly distributive. It follows that Czech binding patterns cannot be accounted for by traditional binding theory, which does not take grammatical person or the difference between collective and distributive interpretations into account. While a comprehensive analysis of these data is still missing, we tentatively sketch a proposal based on the morphosyntactic approach to binding (see Reuland and Zubkov (2022), R&Z furthermore). Following R&Z, we treat anaphor binding as a case of agreement: the anaphor (as a goal) is valued for its deficient  $\phi$ -features by the antecedent (via a probe). R&Z propose two probes: high # (CP level) and low # (vP level), the latter responsible for the obligatory distributive interpretation of dative antecedents in Russian (and other interpretive effects). Our experimental data are compatible with a proposal where the 1st/2nd person antecedents value the low # probe simultaneously valuing the anaphor (goal), resulting in binding (with high preference for distributive construal). Pronouns are specified for  $\phi$ -features, and the low # probe does not mediate their binding (or lack of it).



**Fig. 1:** Experiment A – condition ratings.



**Fig. 2:** Experiment B – condition ratings

**Example item from Experiment A:**

- (1) a. *Já a Petr jsme společně venčili našeho psa.*  
I and Petr AUX.1PL together walk.3PL.PST our.ACC dog.ACC  
'Petr and I walked our dog together.' collective & non-reflexive
- b. *Já a Petr jsme společně venčili svého psa.*  
I and Petr AUX.1PL together walk.3PL.PST our.ACC dog.ACC  
'Petr and I walked our dog together.' collective & reflexive
- c. *Já a Petr jsme každý venčili našeho psa.*  
I and Petr AUX.1PL each walk.3PL.PST our.ACC dog.ACC  
'Petr and I each walked our dog.' distributive & non-reflexive
- d. *Já a Petr jsme každý venčili svého psa.*  
I and Petr AUX.1PL each walk.3PL.PST our.ACC dog.ACC  
'Petr and I each walked our dog.' distributive & reflexive

**Example item from Experiment B:**

- (2) a. *Adama jsme zaměstnávali v naší malé firmě.*  
Adam.ACC AUX.1PL employ.3PL.PST in our.LOC small.LOC business.LOC  
'We employed Adam in our small business.' 1st-person & non-reflexive
- b. *Adama jsme zaměstnávali ve své malé firmě.*  
Adam.ACC AUX.1PL employ.3PL.PST in svůj.LOC small.LOC business.LOC  
'We employed Adam in our small business.' 1st-person & reflexive
- c. *Adama zaměstnávali v jejich malé firmě.*  
Adam.ACC employ.3PL.PST in their.LOC small.LOC business.LOC  
'They employed Adam in their small business.' 3rd-person & non-reflexive
- d. *Adama zaměstnávali ve své malé firmě.*  
Adam.ACC employ.3PL.PST in svůj.LOC small.LOC business.LOC  
'They employed Adam in their small business.' 3rd-person & reflexive

**Bibliography:**

- [1] Charnavel, I. (2019). *Locality and logophoricity: A theory of exempt anaphora*. Oxford University Press.
- [2] Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
- [3] Mertins, B. (2021). On the interpretation of possessives in Czech: An experimental approach. *Oslo Studies in Language* 12(2), 97–126.
- [4] Padučeva, E. V. (1983). Vozvratnoe mestoimenie s kosvennym antecedentom i semantika reflektivnosti. *Semiotika i informatika* 21, 3–33.
- [5] Čmejrková, S. (2003). Osudy zvrátneho posesivního zájmena *svůj*. *Naše řeč* 86, 181–205.
- [6] Reuland, E. and P. Zubkov (2022). Agreeing to bind: the case of Russian. *Glossa: a journal of general linguistics* 7(1).
- [7] Zubkov, P. (2018). *The grammar of binding: A study with reference to Russian*. LOT.
- [8] Perevozchikova, T. (2023). Reflexive or Not? Choosing a Possessive in Bulgarian, Czech, and Russian. *Scando-Slavica* 69/2, 244–263.

## Constraints on sorting: Experimental evidence

Aviv Schoenfeld (Ben-Gurion University of the Negev), Michaela Hulmanová (Masaryk University)

**Background.** In English, the sorted (aka taxonomic) reading of nouns is available in both episodic and generic sentences. For example, *every tree* can mean ‘every kind of tree’ in the episodic *Sharon photographed every tree*, and in the generic *Every tree exchanges carbon dioxide for oxygen* (Carlson, 2006). However, constraints on sorting are not well understood cross-linguistically, and Slavic languages can offer valuable insight. We present new experimental evidence from Slovak regarding constraints on sorting which seem different from English, and which challenge existing analyses.

In addition to the sorted reading being available to c(OUNT) nouns like *tree*, it is available to (M)ASS nouns, e.g. *fuels* can mean ‘kinds of fuel’ (Bunt, 1985). However, o(BJECT) m(ASS) n(OUNS) in English resist sorting, e.g. *furniture* cannot mean ‘kinds of furniture’ (Sutton and Filip, 2018). Additionally, sorting is constrained in Czech by episodicity (Grimm and Dočekal, 2021), e.g. *tři paliva* ‘three fuels’ is better in (1) with the non-episodic *prodává* ‘sell.IMPERF’ compared to the episodic *prodala* ‘sell.PFV’. We present experimental evidence from Slovak which challenges the two preceding citations on the grounds that (i) sorting is better with OMN compared to c(OUNT), and (ii) episodicity imposes the strongest restriction on sorting with c(OUNT).

**Experiment.** The design was 3x2 factorial: (i) The NP-TYPE was c, m or omn, and (ii) the PRED(ICATE) was e(PISODIC) or g(ENERIC). There were 18 items in total in 6 conditions. (2) shows a c item in the e and g conditions. Each participant saw 18 items in a random order. Using an Acceptability Judgment Task (1 to 5 Likert scale: 1 = unacceptable, 5 = fully acceptable), 239 native speakers of Slovak completed an online questionnaire (L-Rex: (Starchenko and Wierzba (2023)) consisting of 18 filler items and 18 experimental items. 211 participants passed the fillers, and their data were analyzed.

**Results.** Using a Bayesian random-effects hierarchical model (R-package: Goodrich et al. (2023), default priors), with slope and intercept both for items and participants), we found that m(ASS) nouns were judged better than the reference level of c(OUNT) and e(PISODIC): Intercept of the model  $\hat{\mu} = 2.05$ , 95% C(redible) Interval [1.87, 2.24]: the effect is positive and certainly existing ( $\hat{\mu} = 0.42$ , CI = [0.22, 0.62], p(robability of) d(irection) 100%, 0% in R(egion) O(f) P(ractical) E(quivalence)). Also, g(ENERIC) predication is a positive and certainly existing effect (g:  $\hat{\mu} = 0.32$ , CI = [0.16, 0.48], pd 100%, 0% in ROPE). Next, o(BJECT) m(ASS) n(OUNS) were judged better than the reference level; the weak positive effect is probably existing, but ROPE suggests that its significance is undecided (omn:  $\hat{\mu} = 0.20$ , CI = [0.08, 0.33], pd 99.88%, 14% in ROPE). Next, g does not improve sorting with m as much as with the reference level: The weak negative interaction effect of m by g is likely existing, but ROPE is  $\geq 2.5$  (m:g:  $\hat{\mu} = -0.19$ , CI = [-0.37, -0.01], pd 98.32%, 26% in ROPE). Similarly, the weak and possibly existing interaction effect of omn by g is negative (omn:g:  $\hat{\mu} = -0.16$ , CI = [-0.33, 0.02], pd 95.95%, 43% in ROPE). This suggests that for omn, g also plays a much smaller role than for c. See figure (1-a) for the descriptive barplot, and Figure (1-b) for the main effects and the interactions.

**Discussion.** For Sutton & Filip (2018), c and m can be sorted, but omn cannot, because only the former are consistent with well-formed taxonomies. However, we find a scale in Slovak: m is best for sorting, followed by omn and c. This suggests that the resistance to sorting of omn in English is more language-specific than cross-linguistic (Erbach and Schoenfeld, 2022).

For Grimm and Dočekal (2021), m nouns like *palivo* ‘fuel’ require an appropriate context to shift to count, which is not provided by e(PISODIC) predication, hence the restriction in (1). This predicts that episodicity should not constrain sorted c(OUNT) nouns, which start as count and thus do not shift to count. We find instead that e imposes the strongest constraint on c. This suggests that the restriction in (1) is for a different reason than *paliva* ‘fuels’ shifting to count.

1. *Naše benzínka {prodá -vá, ??včera Proda -la } tři paliva.*  
 our gas.station {sell.PFV IMPERF, yesterday sell.PFV PTCP} three fuels  
 ‘Our gas station {<sub>G</sub> sells, <sub>E</sub> sold yesterday} three fuels.’
2. *Obchod {predá -va, včera vy- predá -l } tri zemiaky: sladké, nové a červené.*  
 shop.N.SG {sell.PFV IMPERF, yesterday out sell.PFV PTCP} three potato.ACC.PL sweet.ACC.PL  
 new.ACC.PL and red.ACC.PL  
 ‘The shop {<sub>G</sub> sells, <sub>E</sub> yesterday sold out of} three potatoes: Sweet, young, and red.’

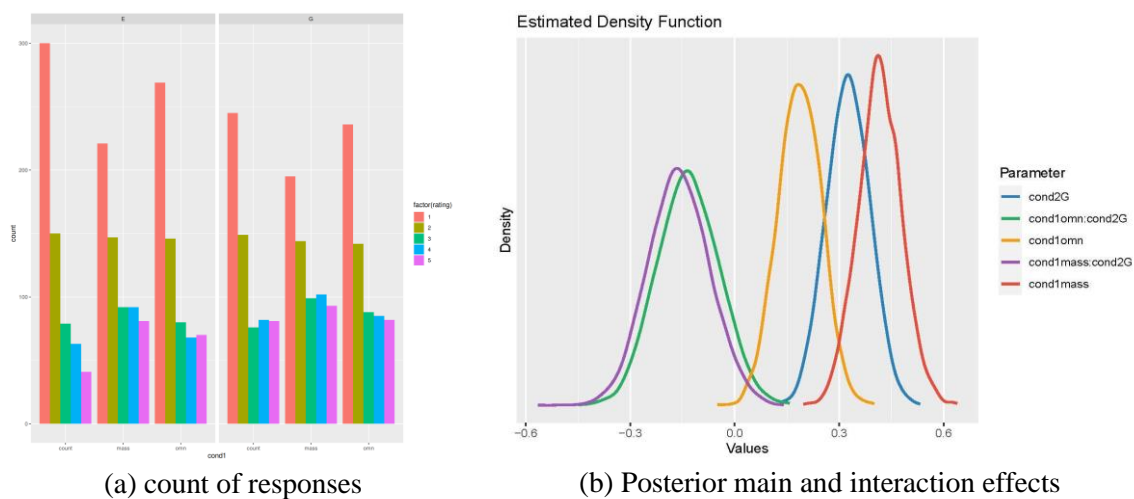


Figure 1: Descriptive stats barplot and the posterior effects graph.

## Bibliography

- Bunt, H. C. (1985). *Mass Terms and Model-Theoretic Semantics*. Cambridge: Cambridge University Press.
- Carlson, G. (2006). Generic reference. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics (Second Edition)*, pp. 14–18. Oxford: Elsevier.
- Erbach, K., and A. Schoenfeld. (2022). Object mass nouns and subkind countability. *Glossa* 7.
- Goodrich, B., J. Gabry, I. Ali, and S. Brilleman (2023). rstanarm: Bayesian applied regression modeling via Stan. R package version 2.26.1.
- Grimm, S. and M. Dočekal (2021). Counting aggregates, groups and kinds: Countability from the perspective of a morphologically complex language. In H. Filip (Ed.), *Countability in Natural Language*, 85–120. Cambridge: Cambridge University Press.
- Starschenko, A. and M. Wierzba (2023). L-Rex Linguistic rating experiments [software], version 1.0.3. GNU General Public License v3.0.
- Sutton, P. R. and H. Filip (2018). Restrictions on subkind coercion in superordinate object mass nouns. In R. Truswell, C. Cummins, C. Heycock, B. Rabern, and H. Rohde (Eds.), *Proceedings of Sinn und Bedeutung* 21, pp. 1195–1123.

## Sémantika dispozičních predikátů

Filip Tvrđý (Univerzita Komenského v Bratislavě)

Podmínková souvětí se stala předmětem zájmu analytické filozofie ve 30. letech 20. století, protože v mnohých případech vzdorovala snahám o jejich zachycení pomocí prostředků formální logiky. Vídeňští novopozitivisté se navíc domnívali, že se filozofové mají podílet na vytvoření vědeckého obrazu světa, a kontrafaktuální podmínky nejsou s tímto úsilím zcela slučitelné: jejich antecedenty a obvykle i konsekventy bývají považovány za nepravdivé. Diskuse se zaměřila na dispoziční predikáty označující nepozorované fyzikální kvality objektů, které se mohou, ale nemusejí projevit (např. „cukr je rozpustný ve vodě“). Představitelé logického empirismu a analytického behaviorismu začali tyto vlastnosti popisovat pomocí podmínkových souvětí. První analýzu dispozičních predikátů provedl Rudolf Carnap, jenž si vystačil se souvětími obsahujícími *indicative mood* a otevřené podmínky:

1. ‘*x* is soluble in water’ = ‘whenever *x* is put into water, *x* dissolves’ (Carnap 1936, 440)

S obdobným problémem se potýkal i Gilbert Ryle, jemuž šlo o analýzu mentálních výroků, které chtěl převést do behaviorálních. Dispoziční predikáty ovšem na rozdíl od Carnapa parafrázoval souvětími se *subjunctive mood* a kontrafaktuálními podmínkami:

2. ‘sugar is soluble’ = ‘sugar would dissolve, or would have dissolved, if immersed in water’ (Ryle 2009, 31)

Takřka stejné řešení můžeme najít i u Nelsona Goodmana, tentokrát na příkladu ohebnosti:

3. ‘*k* was flexible at time *t*’ = ‘if *k* had been under suitable pressure at time *t*, then *k* would have bent’ (Goodman 1983, 34–35)

Kontrafaktuální forma měla zajistit obecnou platnost výroků, která by nevyžadovala aktuální plnění výchozích podmínek pro aktivaci dispoziční vlastnosti. Například – i kdyby nebyla konkrétní kostka cukru (nebo žádná kostka cukru na světě) ponořena do vody, přesto by byla ve vodě rozpustná. Všichni zmínění myslitelé směřovali k nahrazení efemerních dispozic robustnějšími výroky o vnitřní struktuře popisovaných objektů, jejichž pomocí bude možné vysvětlit všechny fyzikální i mentální predikáty (srov. Quine 2013, 205). Další vývoj analytické filozofie tento ideál avšak opustil a nahradil uvažováním o logické a metafyzické modalitě, jež nabyla nejvýraznější podobu v sémantice možných světů.

Cílem mého příspěvku je ukázat, že čeština pro zachycení dispozičních predikátů aparát kontrafaktuálních podmínek nepotřebuje – vystačí si s těmi otevřenými. Čeští logici a filozofové jazyka si to málo uvědomují, protože až příliš spoléhají na anglofonní literaturu. Mechanické přejímání cizojazyčných pramenů vede k terminologické konfúzi, jež pramení z neznalosti českého lingvistického pojmosloví a gramatiky českého jazyka obecně (Tvrđý 2024). Ukazuje se, že jazykové intuice ohledně dispozičních predikátů se u anglických a českých mluvčích výrazně liší. V češtině zavádíme dispoziční termíny pomocí minulého i přítomného kondicionálu, přičemž obě formy lze považovat za gramatické a sémanticky rovnocenné:

4. Křehký objekt je ten, který by se rozbil, kdyby (byl) spadl.
5. Křehký objekt je ten, který se rozbije, když spadne.

Rodilými mluvčími angličtiny je naopak preferována verze používající *subjunctive mood* (6), protože indikativní formulace (7) není obvykle považována za výstižnou (viz Montague 2006, 4):

6. A fragile object is one that would break if it were dropped.
7. ?? A fragile object is one that breaks if it is dropped.

Svou hypotézu doložím příklady z textů českých logiků a filozofů, kteří v použití minulého a přítomného kondicionálu u dispozičních predikátů chybují (např. Berka 1961; Vlasáková 1997; Peregrin 2014; Kolman a Punčochář 2015). Popisovaný problém nelze považovat za marginální, protože může mít závažné metodologické a metafyzické důsledky. Pokud uznáme skutečnost, že se zavádění dispozičních predikátů v různých jazycích liší, budeme nuceni opustit univerzalistickou ambici filozofie jazyka, a snad se také budeme moci vyhnout sémantice možných světů.

## **Bibliografie:**

- Berka, Karel. 1961. „O některých aplikacích moderní formální logiky v jazykovědě“. *Slovo a slovesnost* 22 (3): 198–203.
- Carnap, Rudolf. 1936. „Testability and Meaning“. *Philosophy of Science* 3 (4): 419–471.
- Goodman, Nelson. 1983. *Fact, Fiction, and Forecast*. Cambridge: Harvard University Press.
- Kolman, Vojtěch a Vít Punčochář. 2015. *Formy jazyka: Úvod do logiky a její filosofie*. Praha: Filosofia.
- Montague, Michelle. 2006. „Counterfactuals“. In *Encyclopedia of Philosophy, Vol. 10*, editoval Donald M. Borchert, 4–10. Farmington Hills: Thomson Gale.
- Peregrin, Jaroslav. 2014. *Inferentialism: Why Rules Matter*. New York: Palgrave Macmillan.
- Quine, W. V. O. 2013. *Word and Object*. Cambridge: MIT Press.
- Ryle, Gilbert. 2009. *The Concept of Mind*. London: Routledge.
- Tvrđý, Filip. 2024. „Nesoulad mezi morfosyntaxí a sémantikou podmínkových souvětí“. *Filozofia* 79 (2): 150–167.
- Vlasáková, Marta. 1997. „Dispoziční vlastnosti“. *Filosofický časopis* 45 (1): 37–49.

## Exploring partitivity and space: Insights from spoken and sign languages

Lucia Vlášková (Masaryk University in Brno), Marcin Wągiel (Masaryk University in Brno, University of Wrocław, Leibniz-Centre General Linguistics (ZAS))

**Introduction.** Cross-linguistic research identified three main marking strategies employed in order to express partitivity: (i) the possessive (POSS), (ii) the locative (LOC) and (iii) the ablative (ABL) (e.g., Hoeksema 1996, Koptjevskaja-Tamm 2009, Luraghi & Kittilä 2014, von Heusinger & Kornfilt 2017, Seržant 2021). While the POSS is based on the internal-possession metaphor and exploits possessive morphology to mark the part-whole relation, in this paper we focus on the LOC and ABL, which employ topological notions for the purpose in question. In particular, the LOC is based on the metaphor that the whole is a container in which parts are spatially located. Thus, the part-whole relation is marked by locative prepositions such as ‘in’ or by the locative case (1). In contrast, the ABL employs a spatial periphrasis based on the metaphor that the whole is the origin from which a part is separated. Formally, this is expressed either via the use of the directional (dir.) preposition ‘from’ or the ablative case (2). Our aim is (i) to contribute by extending the typology of partitive constructions by including evidence from sign languages (SLs) and (ii) to examine what semantic component allows for LOC and ABL markers to serve both as spatial and partitivity expressions.

**Sign language strategies.** Studies on partitive constructions in SLs are scarce and all attest the same type of strategy exploiting the SL visual-spatial modality: the higher position in the signing space indicates the non-specific indefinite interpretation of the sign (3), whereas the lower position indicates the specific and partitive interpretation (4) (LSC; Barberà 2015). This strategy has been attested in LSC, ASL (Bahan 1996) and HKSL (Tang & Sze 2002).

**Method and data.** Our focus lies on Czech Sign Language (ČZJ), an understudied SL with approx. 10 000 signers primarily in the Czech Republic. We conducted free and guided elicitations with three native ČZJ informants (2F:1M, 35–52yo., 1st–3rd generation). The high vs. low position strategy was expected to be found also in ČZJ, due to the shared modality with other SLs. However, these constructions were not produced at all, and instead a new, ABL strategy was found. The minimal pairs are shown in (5) and (6), where the partitive constructions in (5a) and (6a) are differentiated from the non-partitive ones in (5b) and (6b) by the sign FROM<sup>IX</sup>-a. FROM<sup>IX</sup>-a is a compound of FROM (Dictio) and the common index sign IX pointing to a specific location in the signing space (a) which has been previously associated with the denotation of the restrictor. FROM itself is borrowed from the surrounding spoken language (Czech) and functions as a dir. preposition, which is unique within ČZJ (the only one identified so far) and SLs in general, since the prevailing strategy of encoding relationships between referents (which is in the semantic domain of prepositions) is carried out by grammatical (as opposed to lexical) means in SLs, e.g., referent positioning in the signing space or classifier constructions.

**Discussion.** The SL data further indicate the tight semantic relationship between the spatial use of locative and dir. expressions and their use as partitive markers. The SL modality can iconically exploit physical space for LOC/ABL. Yet, ČZJ still chooses to employ a dir. preposition as a lexical means for marking partitivity. This suggests a robust link between the manner in which humans conceptualize space and partitivity. Importantly, certain semantic properties of spatial prepositions are often preserved on their partitive use, e.g., Polish has two partitive markers: *z* ‘from’ and the compositionally transparent *pośród* ‘from among’ (*z* + *pośród* ‘among’). On the spatial use *pośród* and *spośród* require the complement nominal to denote a plurality of entities and are infelicitous with expressions referring to singular objects (7). The same restriction is preserved on the partitive use of *spośród*, which makes it incompatible with entity partitives. This distinguishes it from the simplex dir. preposition *z*, which is felicitous with both entity and set partitives (8)–(9). The discussed phenomena parallel the relationship between spatial expressions and number words since directional aspects of spatial prepositions are active when applied to the domain of numerals in modified expressions such as *up to fifty* (Nouwen 2008, 2016, Blok 2015). This in turn invites thinking about the concept of part-whole also in relation to topological notions (Grimm 2012).

- (1) *aan̩kutt̩ika[-il raŋə peer vannu.*  
 male.children-LOC two person come.PST  
 ‘Two of the boys came.’ (Malayalam; Seržant 2021)
- (2) *Ali kadın-lar-dan iki-sin-i tanı-yor-du.*  
 Ali woman-PL- two-3.SG-ACC know-PROG-  
 ABL PST  
 ‘Ali knew two of the women.’ (Turkish; Enç 1991)
- (3) a. HOUSE SOME<sub>[ipsi\_up]</sub> ‘some houses’  
 b. HOUSE ONE<sub>[ipsi\_up]</sub> ‘one houses’  
 c. HOUSE ANY<sub>[ipsi\_up]</sub> ‘any houses’
- (4) a. HOUSE SOME<sub>[ipsi\_down]</sub> ‘some of the houses’  
 b. HOUSE ONE<sub>[ipsi\_down]</sub> ‘one of the houses’  
 c. HOUSE ANY<sub>[ipsi\_down]</sub> ‘any of the houses’
- (5) TABLE BE/HAVE<sub>a</sub> CUBE<sup>++</sup> SIX BLUE  
 ‘There were six blue cubes on the table.’  
 a. THREE FROM<sup>^IX-a</sup> BLUE <sub>a</sub>TAKE<sub>b</sub>  
 ‘Three of the blue ones were taken.’  
 b. THREE BLUE <sub>a</sub>TAKE<sub>b</sub>  
 ‘Three blue ones were taken.’
- (6) TABLE BE/HAVE CAKE STRAWBERRY<sub>a</sub> BLUEBERRY<sub>b</sub>  
 ‘There was a strawberry and a blueberry cake on the table.’  
 a. HALF FROM<sup>^IX-a</sup> STRAWBERRY EAT  
 ‘Half of the strawberry one was eaten.’  
 b. HALF STRAWBERRY EAT  
 ‘Half of the strawberry one was eaten.’
- (7) *Hania wyciągnęła kartkę s-pośród { #książki / książek}.*  
 Hania pulled.out page.ACC from-among book.GEN books.GEN  
 ‘Hania pulled out a page from among the { #book / books }.’ (Polish)
- (8) *Hania przeczytała połowę { z / #s-pośród } książki.*  
 Hania read half.ACC from from-among book.GEN  
 ‘Hania read half of the book.’ (Polish)
- (8) *Hania przeczytała połowę { z / s-pośród } książek.*  
 Hania read half.ACC from from-among books.GEN  
 ‘Hania read half of the books.’ (Polish)

**Bibliografie.** Bahan (1996) Non-manual realization of agreement in American Sign Language • Barberà (2015) The Meaning of Space in Sign Language: Reference, Specificity and Structure in Catalan Sign Language Discourse • Blok (2015) The semantics and pragmatics of directional numeral modifiers • Dictio: Multilingual Online Dictionary (2023) • Enç (1991) The semantics of specificity • Grimm (2012) Number and individuation • Hoeksema (1996) Partitives: Studies on the syntax and semantics of partitive and related constructions • Koptjevskaja-Tamm (2009) A lot of grammar with a good portion of lexicon: Towards a typology of partitive and pseudopartitive nominal constructions • Luraghi & Kittilä (2014) Typology and diachrony of partitive case markers • Nouwen (2008) Directional numeral quantifiers • Nouwen (2016) Making sense of the spatial metaphor for number in natural language • Seržant (2021) Typology of partitives • Tang & Sze (2002) Nominal expressions in Hong Kong Sign Language: Does modality make a difference? • von Heusinger & Kornfilt (2017) Partitivity and case marking in Turkish and related languages



**Adding individuals and numbers:  
A unified analysis of nominal conjunction and additive numerals**

Marcin Wągiel

(Masaryk University in Brno, University of Wrocław, Leibniz-Centre General Linguistics (ZAS))

Cross-linguistically, additive numerals (ANs) such as *two hundred and three* often involve a conjunction marker (Hurford 1975), which indicates a deep relationship between the expression of arithmetical addition (1)–(2) and mereological sum formation (3)–(4). However, applying a non-Boolean analysis of ‘and’ to (1)–(2) in a straightforward manner yields a wrong result. Assuming that the conjoined DPs in (3)–(4) refer to the plural individual  $H \sqcup A$ , (1)–(2) would then refer to the plurality of two entities  $3 \sqcup 20$  while what we want is not a plurality of two objects but a singular object, namely 23.

- |                                                                              |                                                                                                                              |
|------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------|
| (1) drei- <b>und</b> -zwanzig<br>three-and-twenty<br>‘twenty three’ (German) | (3) Heidi <b>und</b> Alma haben sich getroffen.<br>Heidi and Alma AUX REFL met<br>‘Heidi and Alma met.’ (German)             |
| (2) bist <b>o</b> se<br>twenty and three<br>‘twenty three’ (Farsi)           | (4) Heidi <b>o</b> Alma ham-digar-o dide-and.<br>Heidi and Alma each-other-RA seen. PL<br>‘Heidi and Alma have met.’ (Farsi) |

Ionin & Matushansky (2018) argue that phrases involving ANs like *twenty-three cats* have the syntactic structure in (5) and show how Heycock & Zamparelli’s (2005) set-product and Champollion’s (2016) intersective analysis of ‘and’ can derive the desired additive interpretation for (5). What is left unexplored is the role of conjunction in arithmetical uses of ANs in which they do not quantify over entities but rather function as names of abstract numbers (6) (Rothstein 2017). In this paper, I will show that I&M’s approach fails to derive this meaning of ANs and demonstrate how its modified version generates incorrect predictions concerning number agreement and the behavior of classifiers.

- (5) [<sub>&P</sub> [<sub>NP</sub> twenty [<sub>NP</sub> **eats** ] ] and [<sub>NP</sub> three [<sub>NP</sub> cats ] ]    (6) (Twenty-)three is a natural number.

As an alternative, I propose that a unified non-Boolean analysis of ‘and’ in ANs and in nominal conjunction is possible without running into the problems that I&M’s approach faces in arithmetical environments if we adopt an analysis of numerals in the spirit of Wągiel & Caha (2021) coupled with a few additional ontological assumptions. W&C take numerals to be morphosemantically complex objects that underlyingly refer to intervals on the number line (see also Nouwen 2016) represented for *three* by the structure in (8). (9) provides the relevant interval, (10) yields the greatest number in that interval and (11) is a shift from a number to a counting device. W&C’s account is couched in Nanosyntax (Starke 2009), which allows for utilizing different parts of (8) to spell out different aspects of the meaning of *three*. Thus, the entire CLP in (8) represents the quantifying meaning, e.g., in *three cats*, whereas the [NUM SCALE<sub>3</sub>] subtree expresses the arithmetical meaning in (6).

- (8) *three*  $\leftrightarrow$  [CL [NUM SCALE<sub>3</sub>]]                      (9) [[SCALE<sub>3</sub>]] = [0, 3]                      (10) [[NUM]] =  $\lambda P$  [MAX(P)]  
 (11) [[CL]] =  $\lambda n \lambda P \lambda x [ *P(x) \wedge \#(P)(x) = n ]$

To account for ANs let us make the following assumptions. First, I assume that (i) ‘and’ is analyzed in terms of non-Boolean conjunction, i.e., as mereological sum formation ( $\sqcup$ ), but it does not involve idempotence. This is supported by the behavior of nominal conjunction, e.g., (12) is infelicitous if both occurrences of *Kim* pick the same referent. Rather, it must refer to distinct entities. This effect can be pragmatic in nature in (12) but if so, then I assume that it is grammaticalized in ANs.

- (12) #Yesterday at 20:13 pm, Amy met Kim<sub>i</sub> and { Kim<sub>i</sub> / Kim<sub>i</sub> and Tim }.

Furthermore, (ii) intervals corresponding to SCALES are in fact vector-like objects in 1-dimensional space (the number line), which does not involve gaps relative to the relevant level of granularity (Pianesi & Varzi 1996, Mazzola 2019); specifically, they are mereotopological integrated objects that are maximally strongly self-connected (Casati & Varzi 1999, Grimm 2012) 1-dimensional entities (Wągiel 2023) that have a specific length and no determined starting point but start at 0 as a default. Combining (i) and (ii) has the following consequence for capturing (1). When ‘and’ sums up the two intervals corresponding to  $SCALE_3$  and  $SCALE_{20}$ , they cannot overlap. Due to the gapless nature of the space SCALES occupy we get  $[0,3]$  and  $[3,23]$ , respectively. The result is a new mereotopological object that we can represent as the interval  $[0,23]$ . In terms of composition, NUM applies only after the intervals are combined, and thus yields the maximal (rightmost) value of  $[0,23]$  (13).

(13)  $[[\text{drei-und-zwanzig}]] = \text{MAX}([0, 3] \sqcup [3, 23]) = 23$

## References

- Casati, R. & Varzi, A. C. (1999). *Parts and places: The structures of spatial representation*. Cambridge, MA: MIT Press.
- Champollion, L. (2016). Ten men and women got married today: Noun coordination and the intersective theory of conjunction. *Journal of Semantics* 33(3): 561–622.
- Grimm, S. 2012. *Number and individuation*. Stanford University, CA PhD dissertation.
- Heycock, C. & Zamparelli, R. (2005). Friends and colleagues: Plurality, coordination, and the structure of DP. *Natural Language Semantics* 13: 201–270.
- Hurford, J. (1975) *The linguistic theory of numerals*. Cambridge: Cambridge University Press.
- Ionin, T. & Matushansky, O. (2018). *Cardinals: The syntax and semantics of cardinal-containing expressions*. Cambridge, MA: MIT Press.
- Mazzola, C. (2019). The mereotopology of time. *Notre Dame Journal of Formal Logic* 60(2): 215–252.
- Nouwen, R. (2016). *Making sense of the spatial metaphor for number in natural language*. Unpublished manuscript available on LingBuzz: [lingbuzz/003100](https://lingbuzz.org/003100).
- Pianesi, F. & Varzi, A. C. (1996). Events, topology and temporal relations. *The Monist* 79(1): 89-116.
- Rothstein, S. (2017). *Semantics for counting and measuring*. Cambridge: Cambridge University Press.
- Starke, M. (2009). *Nanosyntax: A short primer to a new approach to language*. *Nordlyd* 36(1): 1–6.
- Wągiel, M. (2023). Acts, occasions and multiplicatives: A mereotopological account. *Semantics and Linguistic Theory* 33: 276–297.
- Wągiel, M. & Caha, P. (2021). Complex simplex numerals. *Acta Linguistica Academica* 68(4): 470–515.

6

*Conversion/Zero-derivation:  
Aspects of the theory*

organizers

Cristina Fernández-Alcaina

Salvador Valera

(ENG)

# Workshop description

This workshop is intended to discuss some of the main aspects of the theory of the word-formation process described as conversion or, less frequently, zero-derivation, e.g. as in Czech *výběr* ‘choice’ vs. *vybírat* ‘to choose’, English *bottle*<sup>N</sup> vs. *to bottle*<sup>V</sup>, German *Arbeit*<sup>N</sup> ‘work’ vs. *arbeiten*<sup>V</sup> ‘to work’ or Spanish *cita*<sup>N</sup> ‘appointment’ vs. *citar*<sup>V</sup> ‘to make an appointment’, to name some of the many examples available in a range of languages and from within a range of word-classes that may be involved in the process.

Despite the objective conditions of formal identity and multiple word-class properties under which it is defined (cf. Bauer 1983, Tournier 2010), conversion raises more questions than answers both in specific languages and cross-linguistically (Štekauer et al. 2012, Valera 2015). This is not just as a result of the interpretation of the abovementioned conditions by various schools of thought, but also because the morphological model of each specific language imposes specific conditions on the formalizations of the process (e.g. root-/stem-/word-based conversion). The uncertainty resulting therefrom adds to the inherent difficulty in the resolution of specific questions of the process, starting with the very nature of the process, i.e. whether it is a word-formation process of its own or merely an instance of affixation by a zero allomorph.

A number of recent workshops at major international conferences (Iordăchioaia & Melloni, SLE 2020, Fernández-Alcaina & Valera, Košice 2022, and Vasile & Valera, Bucharest 2022) and several publications in preparation by the workshop convenors (e.g. Bauer & Valera, in preparation) bear witness to all the above. This workshop develops issues that previous events discussed from a different point of view, for different languages, or to a lesser degree of detail:

1. directionality, where the little research published in the past 20 years has evidenced that, whereas substantial progress is made in this regard for several languages (cf., among others, Don 2005 for Dutch or Ševčíková 2021 for Czech), in others the question is very much at the same point as it has been for decades (e.g. English),
2. the semantics of conversion, which has been discussed from a range of perspectives for the polysemanticity inherent in conversion and which sets it apart from other word-formation issues (cf., among others, Clark & Clark 1979, Štekauer 1996), or
3. the relevance and the specificities of conversion as a productive word-formation process (compared or not with other processes) in particular languages where the process has not been discussed in as much detail as, e.g. English, Lithuanian and Latvian within the Baltic family, or Portuguese within the Romance family.

This list of topics is neither exhaustive, nor restrictive. Papers approaching conversion and related issues from a cross-linguistic or typological perspective are welcome.

## References

- Bauer, L. & Valera, S. In preparation. *Conversion: Theory, Typology*.
- Bauer, L. 1983. *English word-formation*. Cambridge: Cambridge University Press.
- Clark, E.V. & Clark, H.H. 1979. When nouns surface as verbs. *Language*, 55(4): 767–811.
- Don, J. 2005. On conversion, relisting and zero-derivation. A comment on Rochelle Lieber *English Word-formation Processes*. *SKASE Journal of Theoretical Linguistics*, 2(2): 2–16.
- Fernández-Alcaina, C. & Valera, S. 2022. *Theoretical and empirical descriptions of conversion/zero-affixation*. Workshop at the *Word-Formation Theories VI & Typology and Universals in Word-Formation V Conference*, Pavol Jozef Šafárik University in Košice, 23–26 June 2022.
- Iordăchioaia, G. & Melloni, C. 2020. *Derivational zero affixes*. Workshop at the *53rd Annual Meeting of the Societas Linguistica Europaea*, 26 August–1 September 2020.
- Ševčíková, M. 2021. Action nouns vs. nouns as bases for denominal verbs in Czech: A case study on directionality in derivation. *Word Structure*, 14: 97–128.
- Štekauer, P. & Valera, S. & Körtvélyessy, L. 2012. *Word-formation in the world's languages*. Cambridge: Cambridge University Press.
- Štekauer, P. 1996. *A theory of conversion in English*. Frankfurt am Main: Peter Lang.
- Tournier, J. [1985] 2010. *Introduction descriptive à la lexicogénétique de l'anglais contemporain*, 2nd edition. Paris/Genève: Champion-Slatkine.
- Valera, S. 2015. Conversion. In P. O. Müller et al. (eds.), *Word-formation: An international handbook of the languages of Europe*, Vol. 1. Berlin: de Gruyter; 322–340.
- Vasile, C. M. & Valera, S. 2022. *Conversion/zero-affixation*. Workshop at the *International Conference, Institute of Linguistics Iorgu Iordan–Al. Rosetti*, University of Bucharest, 18–19 November 2022.

# Table of Contents

*What is conversion? Upping the ante and downing the enemy*

Laurie Bauer (Victoria University of Wellington)

*Conversion from the word-formation and cognitive points of view (evidence from Czech)*

Božena Bednaříková (Palacký University)

*The structural complexity of zero vs. overt nominalizations: An experimental study*

Gianina Iordăchioaia (University of Graz), Jeannique Darby (University College Volda)

*The status of conversion nouns in the Word family. Evidence from priming experiments in German*

Andreas Opitz<sup>1</sup>, Denisa Bordag<sup>1,2</sup>, Alberto Furgoni<sup>1</sup> (<sup>1</sup>Leipzig University, <sup>2</sup>Haifa University)

*Defining conversion in Baltic*

Jurgis Pakerys (Vilnius University)

*Conversion, paradigms and semantic (non-)compositionality*

Gergana Popova (Goldsmiths, University of London)

*The semantics of conversion vs. affixation: Insights from nominalizations in French*

Justine Salvadori, Rossella Varvara, Richard Huyghe (University of Fribourg)

*Non-affixed deverbal nouns, compounds, and conversion degrees*

Alexandra Soares Rodrigues (Instituto Politécnico de Bragança)

## What is conversion? Upping the ante and downing the enemy

Laurie Bauer (Victoria University of Wellington)

English shows patterns of homophony between prepositions and verbs of the following kind:

- (1) *She walked **down** the hill. He **downed** his beer, and got up to leave.*  
 (2) *They ran **up** the stairs. They **upped** the price of covid masks by 200%.*

The question is whether such a relationship is a matter of conversion, and if it is, how do we know and what might be the constraints on it. If it is not conversion, what is it?

The first point to consider is the nature of the prepositions that can be subject to this type of homophony, and whether the label ‘preposition’ is appropriate. It turns out that only a subset of prepositions provide a suitable input, and even then, only a small subset of relevant forms seem to be used. The same prepositions are available for homophony with nouns, although there is no semantic match between nouns and verbs, or even between antonymous pairs of prepositions.

The relevant verbs and nouns are stylistically marked: some are jocular, some are slangy, some are technical. Directionality is not clear. Although new forms can be created on the pattern (e.g. *to out a person* ‘disclose their non-binary sexuality’ from the 1990s), productivity is low and unsystematic. Despite all this, they do not fit the criteria for expressive morphology as defined by Zwicky & Pullum.

When we look at the expected criteria for classing something as conversion, these formations do not fit well, but neither do they fit criteria for coercion. They do appear to work by metonymy, just as conversion has been argued to do, but not entirely in the same way.

We cannot use them to help us choose a model for the description of conversion: they seem to fit relisting rather than zero-derivation, but on a wider scale it is not clear what relevant evidence might be.

As theorists we should be worried by the fact that we cannot tell whether this is conversion or not, or what it might be if it isn’t conversion.

## **Bibliography**

Zwicky, Arnold M. & Pullum, Geoffrey K. 1987. Plain morphology and expressive morphology. In Jon Aske, Beery, Natasha, Michaelis, Laura & Filip, Hana (eds.), *Proceedings of the thirteenth annual meeting of the Berkeley Linguistics Society, general session and parasession on grammar and cognition*, 330–340. Berkeley: Berkeley Linguistics Society.



## Conversion from the word-formation and cognitive points of view (evidence from Czech)

Božena Bednaříková (Palacký University)

**Background** Slavonic languages belong to languages that are heavily inflected (Štekauer and Lieber, 2006). Hence the affixation plays a crucial role in their formal morphology. The same is true in derivational morphology that deals with deriving words from other words using affixes (Pala and Hlaváčková, 2007). Nevertheless, as inflectional languages they are assumed to make use chiefly of conversion (Skalička 1935, Bednaříková 2009, 2020). It is not a usual problem to search for the boundary between inflection and derivation in the respective languages, but rather the boundary between derivation and conversion. Generally conversion is a morphological, formally non-additive proces with onomasiological function whose central formant is a change in morphological characteristics (Bednaříková, 2017).

**Contribution** The basic goal of the project is to develop word-formation networks with regard to both derivation and conversion, namely in Czech (and partly in Russian and Ukrainian). The related goal is to identify individual naming units by semantic categories and find their interrelations. The semantic categories have the character of comparative cognitive categories that are able to capture the meaning introduced by word-formational formant and can be used cross-linguistically. The research sample consists of 10 verbs taken from the so-called core vocabulary. The chosen verbs are primarily underived and unconverted words, but can serve as word-forming bases for other word-formation processes bringing an added onomasiological value.

**Implications** In the end of the study there is a conclusion summarizing the main findings. What appears to be the most interesting result is the frequency of the verbal actions. There may be two sources for explanation: a) employing of conversion as an important word-formational process, b) the power of the so-called deverbatives to retain the meaning of action, including the ability to express some of the typical verbal categories, such as the verbal aspect and the relative time. Further off, the analysis of the research sample suggests that the interrelation between the semantic categories seems to point to trends in mental concepts of the speakers.

## Bibliography

- [1] Bednaříková, Božena, 2020. Conversion and Its Onomasiological Value Cross-linguistically. In: *Lingvistika-korpus-empirie* (Jana Bílková and Ivana Kolářová and Miloslav Vondráček, eds.). Praha: Ústav pro jazyk český. 135–141.
- [2] Bednaříková, Božena, 2017. *Slovo a jeho struktura*. Olomouc: Univerzita Palackého v Olomouci. Goodhue, Daniel. 2022.
- [3] Bednaříková, Božena, 2009. Is Czech ideal representative of the inflectional language type? *Romanoslavica XLV*, Bucharest: Universitatea din Bucuresti, 41–50.
- [4] Pala, Karel, Hlaváčková, Dana, 2007. Derivational relations in Czech Wordnet. In: *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing: Information Extraction and Enabling Technologies*. Association for Computational Linguistics. Prague. 75–81.
- [5] Skalička, Vladimír, 1935. *Zur ungarischen Grammatik*. Prague: Nakl. Filosofické fakulty University Karlovy.
- [6] Štekauer, Pavol, Lieber, Rochelle (eds.), 2005. *Handbook of Word-Formation* (Vol. 64). Springer Science & Business Media.

## The structural complexity of zero vs. overt nominalizations: An experimental study

Gianina Iordăchioaia (University of Graz), Jeannique Darby (University College Volda)

**Aims and motivation.** In this study we experimentally evaluate three patterns of deverbal nominalizations derived by means of zero, *-ing* and Latinate suffixes (i.e. *-(at)ion*, *-ment*, and *-ance*) in terms of their ability to express causative and inchoative readings of their base verbs. The literature on alternating change-of-state verbs such as *crash* in (1) below agrees that the causative/transitive reading is semantically and syntactically more complex than the inchoative/intransitive one (Rappaport Hovav & Levin 1998ff, Schäfer 2009, Alexiadou et al. 2015, Beavers & Koontz-Garboden 2020, a.o.). Assuming that deverbal event nominalizations inherit the event structure of their base verbs (see Grimshaw's 1990 complex event nominals), a question that arises is whether zero nominalizations (ZeroNs) might differ from overtly marked ones in their ability to derive causative and inchoative readings. Although the literature on this topic is limited, it has been argued that causative/transitive readings require an overt nominalizing suffix, and thus ZeroNs can only realize inchoative/intransitive readings: see *the beating of prisoners* vs. *the beat of your heart* (Cetnarowska 1993: 113) and *John's raising/\*raise of the glass* vs. *The elevator's rise to the top floor* (Marantz 1997). This would suggest that ZeroNs are structurally simpler than overt nominalizations (see also Borer 2013: ch. 7). However, natural text corpora reveal plenty of examples of ZeroNs that exhibit causative readings: e.g. *a continuous raise of salaries*, *the deliberate crash of a Germanwings passenger jet*.

**Methodology.** We test this claim experimentally, hypothesizing based on the corpus data that ZeroNs may in fact realize *both* causative and inchoative readings. Since *-ing* is the most productive suffix, we test pairs of nominalizations that employ *-ing* vs. Latinate suffixes (4 base verbs), as well as *-ing* vs. zero (4 base verbs). Native speakers will evaluate the naturalness of each nominalization in describing a causative vs. inchoative event previously expressed with the base verb, as illustrated in (1) for an *-ing*-zero pair. Each of the 8 base verbs will yield 4 stimuli as illustrated below (causative + *ing*; causative + zero; inchoative + *ing*; inchoative + zero), distributed across 4 experimental lists:

(1) **Causative context:** Shop robberies are quite frequent in this poor neighborhood of the city and often go unnoticed. However, after plundering a jewelry store last week, the thieves hijacked a bus and **crashed** it into a public building, while running away from the police. Several people were injured but nobody's life was in danger.

**Inchoative context:** Shop robberies are quite frequent in this poor neighborhood of the city and often go unnoticed. However, after plundering a jewelry store last week, a thief hijacked a bus and managed to escape the police by jumping out of the bus and grabbing a motorcycle. Without a driver, the bus **crashed** into a public building.

**Target sentence:** The {**crash/crashing**} of the bus happened after the robbery.

In order to assess the validity of the task, each list will also include causative-only and inchoative-only verbs, as well as filler items. In terms of our timeline, we are currently finishing up the design before making the experiment available online to paid participants. By the time of the conference in September, we expect to have concrete results and answers to our hypotheses.

**Contribution to the study of conversion.** Following the data attestations in corpora, we predict that ZeroNs will be capable of receiving causative readings, and thus do not necessarily have simpler event structures than overtly marked forms. Should this hypothesis be confirmed, it will have an important impact on the discussion about the syntactic-semantic complexity of zero vs. overt morphology, showing conversion to be more systematic than usually assumed. In addition, we expect suffix competition for individual items to also play a role in interpretation, and for this we will evaluate additional factors such as noun/base verb frequency and Germanic/Latinate etymology of the base.

## **Bibliography**

- Alexiadou, A., E. Anagnostopoulou & F. Schäfer. 2015. External arguments in transitivity alternations: A layering approach. Oxford: Oxford University Press.
- Beavers, J. & A. Koontz-Garboden. 2020. The roots of verbal meaning. Oxford: Oxford University Press.
- Borer, H. 2013. Taking form. Structuring sense, vol. 3. Oxford: Oxford University Press.
- Cetnarowska, B. 1993. The syntax, semantics and derivation of bare nominalizations in English. Katowice: Uniwersytet Śląski.
- Grimshaw, J. 1990. Argument structure. Cambridge, MA: MIT Press.
- Marantz, A. 1997. No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania Working Papers in Linguistics* 4.2: 201-225.
- Rappaport Hovav, M. & B. Levin, 1998. Building verb meanings. In Butt, M. & W. Geuder, eds., *The projection of arguments*, 97-134. Stanford: CSLI Publications.
- Schäfer, F. 2009. The causative alternation. *Language and Linguistics Compass* 3.2: 641-681.

**The status of conversion nouns in the word family.  
Evidence from priming experiments in German**

Andreas Opitz<sup>1</sup>, Denisa Bordag<sup>1,2</sup>, Alberto Furgoni<sup>1</sup> (<sup>1</sup>Leipzig University, <sup>2</sup>Haifa University)

In a series of overt priming experiments, we investigated how German word families are organized and structured in the adult human mind, i.e., what are the relations between the word family members and how they affect each other. Previous research has shown that morphologically related word forms typically speed up each other's recognition indicating that their representations are not independent. At the same time, priming is not equally strong between the members of a word family (e.g., Schriefers et al., 1992) implying that a word family is not a simple list but has a complex structure (cf. van de Vijver & Baer-Henney, 2019). Surprisingly, relatively little research exists that directly addresses the question of word family mental representation (cf. Schmidtke & Kuperman, 2019), though some topics that are subsumed under it have been explored rather extensively, such as the research on inflection and derivation suggesting that the degree to which word family members affect each other depends on the type of morphological relation between them. *In our study, we focus on the mental relationships of the so-called non-finites (in particular infinitives, participles, and conversion nouns) between each other and also on their relationship to other members of their morphological family.*

In our experiments, we integrated both primes and targets into short syntactic contexts that enable their morphosyntactic disambiguation in cases, where the forms were overtly homonymous. The relationship between the primes and the targets was manipulated along the way demonstrated as an example in **Table 1**. for an experiment, where the target form was a conversion noun. We designed further experiments of this type in which we varied the target form (e.g., infinitive form, inflected verb, countable noun) and the morphological function of the primes.

The results of this series of experiments revealed a complex pattern of different priming strengths among the members of a word family (see **Figure 1**). Inflected forms fully primed (i.e., to the same degree as the identical condition) infinitives and participles, but participle and infinitive forms only partially primed finite forms. Conversion nouns differed from both infinitives and participles on the one hand and finite forms on the other. They were partially primed by both the infinite forms and finite forms, and they also only partially primed them when presented as primes. In addition, we also observed that the presence or absence of an orthographic cue (capitalization of initial letter) influenced the priming potential of conversion nouns.

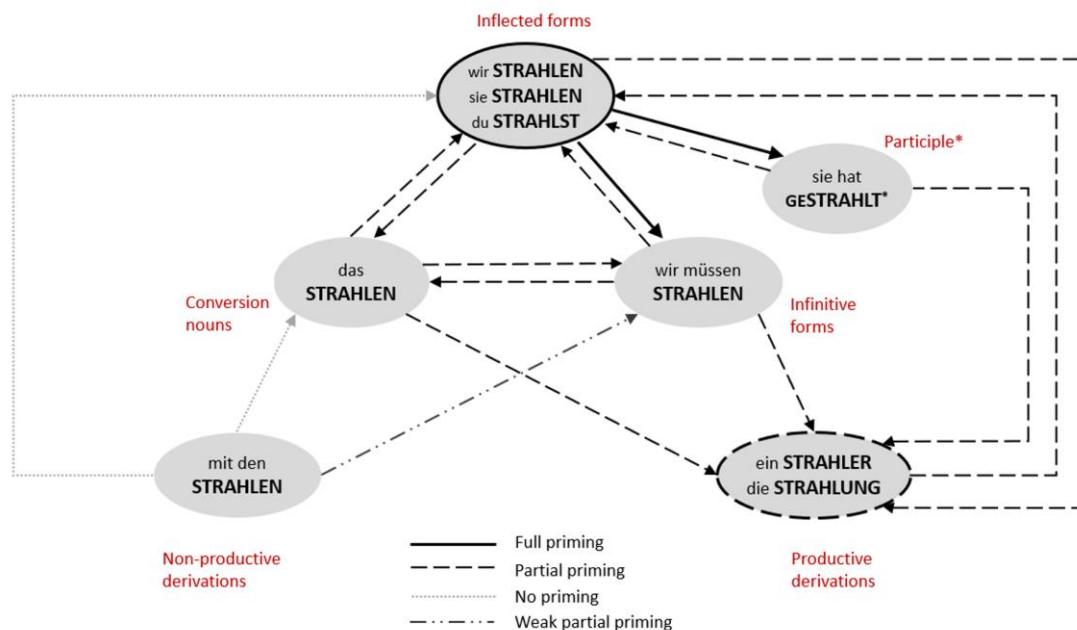
The pattern of results allows to draw several conclusions about the representational status of the explored forms. First, the assumption presented in typological literature that there is a class of the so-called non-finites (see Ylikoski, 2003, for an overview) in addition to the classes of verbs and nouns is not fully supported by the psycholinguistic evidence: While infinitives and participles manifest the same properties regarding the representation and processing within their morphological family, conversion nouns diverge from that pattern. However, their priming pattern also differs from that of derived countable nouns indicating that they are more strongly tied to/ closer to the verbal forms in the mental structure of morphological families.

Second, the results indicate that infinitives and participles have the same status in the mental representation of the morphological family, but both differ from conversion forms. The asymmetrical priming results between them and finite forms indicate a hierarchical structure in which underspecified infinite forms represent (or are linked with) (a) higher node(s) which are also accessed when representations of finite forms are retrieved. Thus, full priming is observed when an infinite form is accessed after the finite form, but only partial priming is observed when the forms are accessed in the reversed order. In our talk we will elaborate on the theoretical implications of the findings and outline the directions of further research.

**Table 1.** Examples of conditions of an experiment in which the target was a conversion noun.

Condition	Prime Phrase	Target Phrase	
		Part 1	Part 2
1 identical	<i>das</i>	<i>SPIELEN</i>	
	‘the’	‘playing’	
2 inflected	<i>wir</i>	<i>SPIELEN</i>	
	‘we’	‘play’	
3 infinitive	<i>wir wollen</i>	<i>SPIELEN</i>	
	‘we want to’	‘play’	<i>das SPIELEN</i>
4 conversion 2	<i>beim</i>	<i>SPIELEN</i>	‘the’ ‘playing’
	‘while’	‘playing’	
5 countable noun	<i>mit den zwei</i>	<i>SPIELEN</i>	
	‘with the two’	‘plays’ (pl.)	
6 unrelated	<i>sie</i>	<i>WEINEN</i>	
	‘they’	‘cry’	

**Figure 1.** Pattern of priming relations, summarized across experiments (arrows indicate direction of tested priming, “full priming” = priming of the same size as identity priming, “no priming” = priming of the same size as unrelated prime)



## Bibliography

- Schmidtke, D., & Kuperman, V. (2019). A paradox of apparent brainless behavior: The time-course of compound word recognition. *Cortex*, 116, 250-267.
- Schriefers, H., Friederici, A., & Graetz, P. (1992). Inflectional and derivational morphology in the mental lexicon: Symmetries and asymmetries in repetition priming. *The Quarterly Journal of Experimental Psychology*, 44(2), 373-390.
- van De Vijver, R., & Baer-Henney, D. (2019). Paradigms in the mental lexicon: Evidence from German. *Frontiers in Communication*, 3, 65.
- Ylikoski, J. (2003). Defining non-finites: action nominals, converbs and infinitives. *SKY Journal of linguistics*, 16(1), 185-237.

## Defining conversion in Baltic

Jurgis Pakerys (Vilnius University)

Baltic languages (Indo-European, Balto-Slavic) have rich morphology and some of their derivational patterns are clear candidates for conversion, see (Stundžia 2016: 3091, 3102–3104) on Lithuanian and (Navickaitė-Klišauskienė 2016: 3120–3121), (Kalnača & Lokmane 2021: 66, 73, 137–140) on Latvian. In order to capture conversion in Baltic, one may begin with two requirements frequently mentioned in the definitions of canonical conversion, see (Bauer 1983: 32), (Bauer & Valera 2005), (Valera 2015: 322), etc.: **(1)** word-class change, **(2)** identity of form.

In the basic case, stem-based derivation is found in Baltic when the inflectional material of the base is substituted with that of the derivative while the base stem itself remains unaltered, e.g., Lithuanian *apklaus-ti* ‘(to) interview’ → *apklaus-a* ‘(a) survey’, Latvian *teik-t* ‘tell’ → *teik-a* ‘tale’. Here, the change of inflectional markers need not be seen as a violation of (2), see discussion in (Valera 2015: 330–331). In (Kalnača & Lokmane 2021: 73), however, the added inflections are interpreted as overt derivational markers and formations of this type are excluded from the description of conversion, see (Valera 2014: 160) on this point of view. If (1) and (2) are accepted as sufficient attributes of conversion, the derivational function ascribed to the inflectional endings appears to be unnecessary limitation for the definition of conversion in Baltic. To address this aspect explicitly, one may add yet another requirement allowing further differentiation of conversion from prefixal and suffixal derivation: **(3)** no dedicated derivational markers.

This requirement is also helpful in the case when the identity of form is not maintained due to morphophonological alternations or base subtraction, e.g., Lithuanian *atrink-ti* ‘select’ → *atrank-a* ‘selection’ (root vowel alternation), Latvian *apgād-ā-t* ‘provide’ → *apgād-e* ‘provision’ (suffix *-ā-* of the base is subtracted), Lithuanian *aukšt-as* ‘high’ → *aūkšt-is* ‘height’ (syllable accent alternation), etc. Such cases may still count as conversion because they fulfill requirements (1) and (3); see further discussion of conversion without the identity of form in (Valera 2015: 328–329).

Moreover, if word-class change is not seen as an obligatory requirement of conversion (Valera 2015: 331–332), requirements (2) and (3) would be still fulfilled and the following cases could be listed under conversion, as in (Stundžia 2016: 3102), but not in (Navickaitė-Klišauskienė 2016: 3120): Lithuanian *knyg-a* ‘book’ → *knyg-ius* ‘book lover, etc.’, *generol-as* ‘(army) general (male)’ → *generol-ė* ‘(army) general (female)’. There are formations of this type when the identity of form is not maintained and only requirement (3) is fulfilled, e.g., Lithuanian *puod-as* ‘pot’ → *puodž-ius* ‘potter’ (syllable accent alternation; stem-final consonant alternation is required by the inflection class), Latvian *cilvēk-s* ‘human’ → *cilvēc-e* ‘humanity’ (consonant alternation required by the inflection class).

Finally, we reach the cases when conversion occurs without the exchange of inflection and these are the only instances listed under conversion in (Kalnača & Lokmane 2021: 73), e.g., Latvian *vecāk-i* ‘older’ (nom. pl. masc. comparative grade of *vec-s*, *-a* ‘old’) → *vecāk-i* ‘parents’ (masc., plurale tantum), Lithuanian *jaunoj-i* ‘(the) young’ (nom. sg. fem. definite form of *jaun-as*, *-a* ‘young’) → *jaunoj-i* ‘bride’ (fem.). If we consider the requirements listed above, we see that all of them are fulfilled. The crucial difference, however, is that we are not dealing with the stem-based derivation mentioned earlier where the inflectional material is changed; here, part of the original paradigm of the base is transferred to the derivative and the remaining part of the paradigm is blocked due to the word-class change. Another important aspect is the interpretation of such cases as lexical derivation rather than syntactic phenomena, see (Valera 2014: 157–158), (Valera 2015: 325–326).

If all aspects discussed above are taken into account, we arrive at the following tentative definition: conversion in Baltic is a derivational process where no dedicated derivational markers are employed and word-class, base stem, and inflectional markers may change. The observed phenomena form a continuum in respect to canonical conversion which can be still usefully defined as a derivational process characterized by word-class change and identity of (non-inflectional) form.

## Bibliography

Bauer, Laurie. 1983. *English Word-Formation*. Cambridge: Cambridge University Press. (doi:10.1017/CBO9781139165846)

Bauer, Laurie & Valera, Salvador. 2005. Conversion or zero-derivation: an introduction. In Bauer, Laurie & Valera, Salvador (eds.), *Approaches to conversion/zero-derivation*, 7–17. Münster, New York, Munich, and Berlin: Waxmann.

Kalnača, Andra & Lokmane, Ilze. 2021. *Latvian Grammar*. Rīga: University of Latvia Press. (doi:10.22364/latgram.2021)

Navickaitė-Klišauskienė, Agnė. 2016. Latvian. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe*, vol. 5, 3107–3123. Berlin, Boston: De Gruyter. (doi:10.1515/9783110424942-002)

Stundžia, Bonifacas. 2016. Lithuanian. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe*, vol. 5, 3089–3106. Berlin, Boston: De Gruyter. (doi:10.1515/9783110424942-001)

Valera, Salvador. 2014. Conversion. In Lieber, Rochelle & Štekauer, Pavol (eds.), *The Oxford Handbook of Derivational Morphology*, 154–168. Oxford: Oxford University Press. (doi:10.1093/oxfordhb/9780199641642.013.0010)

Valera, Salvador. 2015. Conversion. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe*, vol. 1, 322–339. Berlin, Boston: De Gruyter. (doi:10.1515/9783110246254-019)



## Conversion, paradigms and semantic (non-)compositionality

Gergana Popova (Goldsmiths, University of London)

As is well known, in English and other languages there are pairs of words that are semantically related and have the same form, but belong to two different word classes, e.g. *table* (n) and *table* (v), *snitch* (v) and *snitch* (n), *brown* (adj) and *brown* (v), and so on. It is often argued that in most cases one of these words is derived from the other, a derivational relationship known as conversion (e.g. Martsa 2013, Valera 2014, a.o.).

Conversion has a number of properties that make it resemble other derivational processes, as well as some that make it stand out. As in the case of some of the lexemes derived via affixation, the two forms related by conversion belong to different word classes. The semantic changes attributed to conversion are in many cases very similar to those brought about by derivational affixation. And, in some cases, conversion seems to be in competition with affixal derivation (see Bauer et al. 2013 and references therein). For this reason, and controversially, conversion has frequently been modelled via zero-derivation, i.e. derivation by zero affixes. At the same time, conversion can be associated with far greater polysemy than derivational affixation (Bauer et al. 2013).

From a theoretical point of view, conversion has been accounted for in several different ways, amongst them as rule-governed or based on analogy (Martsa 2020). More recently, conversion has been placed in a different approach to derivational morphology, namely a paradigmatic one (Štekauer 2014, Rodrigues 2022). Paradigmatic approaches give prominence to semantic alignment between forms in (delimited) information spaces (Bonami and Strnadová 2019). In paradigmatic approaches to inflectional morphology the semantic relationship is seen as primary, and formal relatedness as its reflex, or exponence.

This paper argues that this recent resurgence of interest in derivational paradigms allows us to draw another parallel between inflectional and derivational paradigms. In particular, conversion can be conceived as similar to the ‘identity function default’ in inflectional paradigms (see Stump 2001 for a definition). Identity function defaults apply when other morphological rules don’t or can’t. Focusing in particular on noun to verb conversion in English, this paper asks the question whether conceiving of conversion in such terms could help account for some of its properties without recourse to zero-derivation. Identity function defaults operate when a morphological process applies but is not accompanied by a formal exponent, so in that respect the lack of formal changes typical of conversion falls out automatically. Given the additional assumption that its application is governed by an elsewhere condition, it can also account for the observation that conversion is largely complementary to the derivation of denominal verbs via affixation (Marchand 1969, Plag 1999). However, such a paradigmatic account is challenged by the semantic flexibility of noun to verb conversion and by the existence of doublets of affixed and converted forms. Semantic flexibility, in particular, demands a more careful look at the differences between inflectional and derivational paradigms, since it means we can’t assume a pre-defined semantic space. The paper argues that although this semantic flexibility sets this particular type of conversion apart, it also lends itself to a view of semantic patterns structured within an elsewhere principle.

## Bibliography

- Bauer, L., Lieber, R., and Plag, I. (2013). *The Oxford Reference Guide to English Morphology*. Oxford University Press, Oxford.
- Bonami, O. and Strnadová, J. (2019). Paradigm structure and predictability in derivational morphology. *Morphology*, 29:167–197.
- Lieber, R. and Štekauer, P., editors (2014). *The Oxford Handbook of Derivational Morphology*. Oxford University Press, Oxford.
- Marchand, H. (1969). *The Categories and Types of Present-Day English Word-Formation: A Synchronic Diachronic Approach*. C. H. Beck'sche Verlagsbuchhandlung, München, 2nd edition.
- Martsa, S. (2013). *Conversion in English: A Cognitive Semantic Approach*. Cambridge Scholars Publishing, Cambridge.
- Martsa, S. (2020). Conversion in morphology. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press, Oxford.
- Plag, I. (1999). *Morphological Productivity: Structural Constraints on English Derivation*. Mouton de Gruyter, Berlin.
- Rodrigues, A. S. (2022). Conversion in a paradigmatic framework of word formation. In Ruz, A. E., Fernández-Alcaina, C., and Lara-Clares, C., editors, *Paradigms in Word Formation: Theory and applications*, pages 215–248. John Benjamins, Amsterdam/Philadelphia.
- Štekauer, P. (2014). Derivational paradigms. In Lieber and Štekauer (2014), pages 354–369.
- Stump, G. (2001). *Paradigm Function Morphology. A Theory of Inflectional Form*. Cambridge University Press, Cambridge.
- Valera, S. (2014). Conversion. In Lieber and Štekauer (2014), pages 154–168.

## The semantics of conversion vs. affixation: Insights from nominalizations in French

Justine Salvadori, Rossella Varvara, Richard Huyghe (University of Fribourg)

Semantic properties of morphological patterns is a long-standing topic of discussion in the word-formation literature. The present study focuses on conversion as opposed to affixation, which differ in terms of absence vs. presence of overt derivational material. It specifically addresses two common (but somewhat conflicting) assertions regarding their semantics: on the one hand, that converted words are semantically more diverse than affixed ones (e.g., Plag, 1999: 221); on the other hand, that converted and affixed words exhibit similar semantic categories (e.g., Don, 2005: 7).

Despite some recent studies (e.g., Missud & Villoing, 2021; Valera, 2023; Barbu Mititelu et al., 2023), quantitative comparisons of the semantics of conversion and affixation are not plentiful. Moreover, depending on the language studied, the focus is often only on a selection of affixes vs. conversion considered as a global phenomenon, or on a handful of affixes vs. a few forms of conversion (when conversion can take several forms as in French for instance, cf. Tribout, 2010).

In light of all this, we propose a comprehensive quantitative study of the semantics of converted and suffixed nouns in French in the context of verb-noun conversion and verb-to-noun suffixation. Abstracting away from the issue of directionality, we systematically analyze and compare the semantic diversity of 4 forms of conversions (e.g., conversion based on past participles as in *flamber<sub>V</sub>* ‘blaze’ ~ *flambée<sub>N</sub>* ‘blaze’, or on nominalized infinitives as in *déjeuner<sub>V</sub>* ‘have lunch’ ~ *déjeuner<sub>N</sub>* ‘lunch’) and 42 nominalizing suffixes (e.g., *-age*, *-ment*, *-eur*). Our goal is twofold: (i) determine whether converted nouns are indeed semantically more diverse than affixed ones; (ii) determine to what extent the semantics of converted nouns and the semantics of suffixed ones are (dis)similar.

The present study is based on a large dataset of 5,120 nouns that were extracted from the FRCOW16A corpus (Schäfer & Bildhauer, 2012) and analyzed manually according to three criteria: (i) related verb, assuming that derivational processes apply to semantically specified items (Fradin & Kerleroux, 2003); (ii) ontological type (e.g., Event, Artefact), which depends on the nature of the noun referent; (iii) relational type (e.g., Agent, Instrument), which describes the relation of the nominal lexeme with the related verb. Ontological and relational types were assigned to nouns on the basis of linguistic tests taken from the literature on nominal semantics (e.g., Flaux & Van de Velde, 2000). Given its pervasiveness in nominalization, lexical ambiguity was also taken into account: different meanings were postulated for any change of related verb, ontological type or relational type (e.g., *couver<sub>VI</sub>* ‘brood’ ~ *couveuse<sub>NI</sub>* ‘sitter’, *couver<sub>VI</sub>* ‘brood’ ~ *couveuse<sub>N2</sub>* ‘incubator’).

In total, 7,724 nominal meanings were identified, which represent 76 combined semantic types (i.e., comprising an ontological type and a relational type) such as Artefact-Instrument (e.g., *raserv* ‘shave’ ~ *rasoir<sub>N</sub>* ‘razor’) and Event-Transposition (e.g., *se suiciderv* ‘commit suicide’ ~ *suicide<sub>N</sub>* ‘suicide’). In terms of semantic diversity, our results indicate that 3 forms of conversion are indeed more semantically diverse than most suffixes according to the Shannon diversity index, which takes into account both the number of combined semantic types each morphological process realizes and the evenness of their abundance (see Figure 1 on next page). As far as the (dis)similarity between converted and suffixed words is concerned, the results of a hierarchical clustering analysis based on the Bray-Curtis dissimilarity coefficient show that the 4 forms of conversion are only semantically close to so-called “eventive” suffixes (e.g. *-age*, *-ment*, *-ion*).

Overall, the analysis reveals that the various forms of conversion we examined do not constitute a homogeneous semantic group. Moreover, although semantically close to certain eventive suffixes, these forms of conversion generally tend to realize a greater variety of semantic functions, and often in a more balanced manner in terms of frequency.

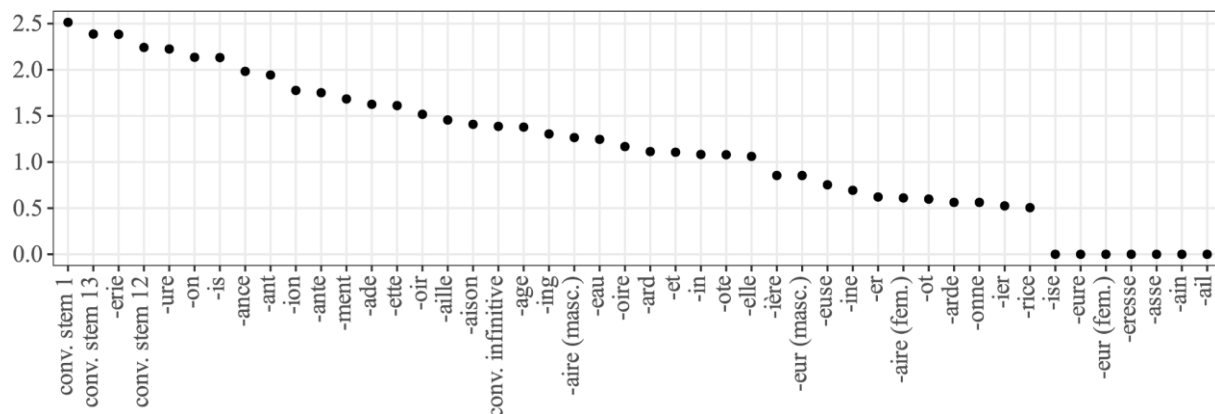


Figure 1. Shannon diversity score per derivational process.

## Bibliography

- Barbu Mititelu, V., Leseva, S., & Stoyanova, I. (2023). Semantic analysis of verb–noun zero derivation in Princeton WordNet. *Zeitschrift für Sprachwissenschaft*, 42(1), 181-207.
- Don, J. (2005). On conversion, relisting and zero-derivation. *SKASE Journal of Theoretical Linguistics*, 2(2), 2-16.
- Flaux, N., & Van de Velde, D. (2000). *Les noms en français : esquisse de classement*. Ophrys.
- Fradin, B., & Kerleroux, F. (2003). Quelle base pour les procédés de la morphologie constructionnelle ? *Sillexicales*, 3, 76-84.
- Missud, A., & Villoing, F. (2021). Investigating the distributional properties of rival *-age* suffixation and verb to noun conversion in French. *Verbum*, 43(1), 41-68.
- Plag, I. (1999). *Morphological productivity: Structural constraints in English derivation*. Mouton de Gruyter.
- Schäfer, R., & Bildhauer, F. (2012). Building large corpora from the web using a new efficient tool chain. In N. Calzolari, K. Choukri, T. Declerck, M. Uğur Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odiijk, & S. Piperidis (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 486-493). European Language Resources Association (ELRA).
- Tribout, D. (2010). *Les conversions de nom à verbe et de verbe à nom en français* [Doctoral dissertation, Université Paris Diderot (Paris 7)].
- Valera, S. (2023). The semantics of noun-to-verb zero-derivation in English and Spanish. *Zeitschrift für Sprachwissenschaft*, 42(1), 153-180.

## Non-affixed deverbal nouns, compounds, and conversion degrees

Alexandra Soares Rodrigues (Instituto Politécnico de Bragança)

One main topic about conversion concerns the extension of cases it may apply to (cf. Valera 2014, 2015; Martsa 2020), depending on the conditions that outline the concept of conversion itself and the formalisation of the process between bases and derivatives (e.g. root-/ stem- or word-based; cf. Manova and Dressler 2005). This paper aims to ponder on the boundaries of conversion based on the assumption that conversion is a class-shifting word-formation process preserving the phonological form of the base (Sweet 1891). The paper questions whether every word-formation process fulfilling those requirements is conversion, or if there are conditions that lead to conceiving either degrees of conversion or a different concept to describe some of those processes. The reflection is based on Portuguese data.

Portuguese presents deverbal nouns lacking affixation, showing class shift and preserving the phonological form of the base described as conversion (Rodrigues 2016). However, two groups within those nouns display different syntactic and semantic behaviours.

For example, the same verb (e.g. *escachar* ‘to crack’; *matar* ‘to kill’) is the base of a deverbal noun lacking derivational affixation with the meaning of ‘event’ or ‘result of event’ (*escacha*<sub>FEM</sub> ‘crack’, *mata*<sub>FEM</sub> ‘injury made by harness in horse cattle’) and of a deverbal noun also lacking derivational affixation with the meaning of ‘agent; instrument; cause’ (*escacha*<sub>MASC</sub> ‘one who cracks everything’, *mata*<sub>MASC</sub> ‘kind of dodgeball; knockout match’). Some aspects separate both nouns from each other. The latter i) are masculine or allow for gender alternation; ii) mean ‘agent’ / ‘instrument’ / ‘cause’ (*mata* is either a match that ‘kills’, i.e. knocks out the adversary, or, as a kind of dodgeball, a game where the one who is hit with the ball is knocked out’); iii) do not allow for the indefinite article; iv) designate something or someone characterised by a particular behaviour or feature’; v) do not theta-mark via preposition.

The same construction seems to appear in compounds (*mata-aula* ‘general absenteeism to classes’, *mata-bicho* ‘breakfast’, *mata-boi* ‘harness of bovine cattle’). The syntactic behaviour of those ‘agent; instrument; cause’ deverbal nouns differs from the behaviour of suffixed deverbal nouns with the same meanings (*escachador* ‘cracker’) and is more similar to the behaviour of the first element of the compounds (*o escachador de nozes* ‘the nutcracker’; literally, ‘the cracker of nuts’ vs. *o escacha-nozes* ‘the nutcracker’, literally, ‘the cracker nuts’).

These data lead us to several questions which the paper aims to answer:

- i) Is the first element of those compounds a noun created by conversion (as it is the noun *mata* in *Vamos jogar ao mata* ‘let us play *mata*’), or is it a verb?
- ii) Are the deriving processes of *escacha* ‘agent; instrument’ and *escacha* ‘event’ formalised similarly, or are there differences in root- /stem- /word-based formalisations?
- iii) Although both *escacha* ‘agent; instrument’ and *escacha* ‘event’ are deverbal nouns lacking affixation and preserving the phonological form of the base, may both words be interpreted as created by conversion, or do the differences between them allow us to conceptualise conversion degrees?
- iv) Are those degrees close enough to consider those phenomena within conversion, or are they better interpreted as transposition?
- v) Are there sufficient and necessary features that permit the distinction of those phenomena?

## Bibliography

- Manova, Stela and Wolfgang Dressler (2005). 'The morphological technique of conversion in the inflecting-fusional type', in Laurie Bauer and Salvador Valera (eds.), *Approaches to conversion/zero-derivation*. Münster: Waxmann Verlag, 67–101.
- Martsa, Sándor (2020). 'Conversion in morphology', In Rochelle Lieber (ed.), *The Oxford encyclopedia of morphology*, Vol. 1. Oxford: Oxford University Press, 779–802.
- Rodrigues, Alexandra Soares (2016). 'Noções basilares sobre a morfologia e o léxico', in Graça Rio-Torto, Alexandra Soares Rodrigues, Isabel Pereira, Rui Pereira, and Sílvia Ribeiro, *Gramática derivacional do português*. Coimbra: Imprensa da Universidade de Coimbra, 2<sup>nd</sup> ed., 35–133.
- Sweet, Henry. (1891). *A new English grammar*. Vol. I. Oxford: Clarendon Press.
- Valera, Salvador. (2014). 'Conversion', in Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford handbook of derivational morphology*. Oxford: Oxford University Press, 154–168.
- Valera, Salvador (2015). 'Conversion', in Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, and Franz Rainer (eds.), *Word-formation. An international handbook of the languages of Europe*, vol. 1. De Gruyter Mouton, 322–339.

7

# Aktuální výzvy (nejen) české dialektologie

organizují

Martina Ireinová

Marta Šimečková

(CZE)

# Popis panelu

V souvislosti s technickým pokrokem pronikají i do dialektologie nové metody, které jsou aplikovány na zpracování různého typu jazykových dat. Na akademických i univerzitních pracovištích jsou zakládány nářeční korpusy, jsou vyvíjeny softwary specializované na automatickou transkripci nářečních audiozáznamů nebo detekci dialektů metodami strojového učení, výrazný pokrok zaznamenala i geolingvistika uplatněním postupů geoinformatiky, též nářeční lexikografie využívající při tvorbě nářečních slovníků již téměř výhradně editační aplikační softwary (výběrově viz Goláňová–Waclawiczová, 2019; Šimečková et al., 2022; Ireinová, 2021; Voženílek, 2022). Nové metodologické otázky vyvstávají už při pořizování dialektologických dat využívajících digitální technologie, zejm. mobilní aplikace nebo elektronické dotazníky, jak se již děje v zahraničí (např. Vaux–Golder, 2003; Möller–Elspaß, 2008; Goldman et al., 2014). Různě se řeší též zachování existujících dat (obvykle audiální/audiovizuální povahy) ve snaze o ochranu před jejich poškozením nebo úplným zničením.

Primárním cílem panelu tak bude sdílení zkušeností s moderními metodami a technikami využívanými v dialektologii a areálové lingvistice při sběru, analýze a prezentaci jazykových dat, též komparace nových metod s postupy tradičními a diskuze o zpřístupňování výsledků výzkumu veřejnosti (i s ohledem na problematiku ochrany osobních údajů a autorského práva).

Volba konkrétních metod je (resp. měla by být) podřízena výzkumným cílům. Tradiční (nejen) česká dialektologie byla dříve zaměřena na popis regionálního rozrůznění češtiny užívané příslušníky starších generací, přičemž se vycházelo z představy ideálních nositelů dialektu z řad příslušníků starších generací. Bylo by však vhodné přejít k detailnějšímu výzkumu, pochopitelně vyžadujícímu nové metody, jenž by zohlednil proměny regionálně rozrůzněné jazykové situace z různých perspektiv a v němž by byl nově kladen důraz na vedení dialektologického výzkumu ve vertikální i horizontální linii (viz např. Chromý, 2014). Sekundárním cílem panelu tak bude klást si otázky, jakými směry vést dialektologický výzkum v budoucnosti a jaké inovační (mj. interdisciplinární) přístupy v něm aplikovat (výběrově viz Chambers–Trudgill, 1998; Preston, 1999; Long–Preston, 2002; Kristiansen, 2018).

Z popsaného plynou dvě komplexní otázky (s vydělením otázek dílčích):

- **Jaké moderní metody lze použít pro získání dialektologických dat, jejich analýzu a zpřístupnění výsledků? Diskutovány budou přednosti a nevýhody jednotlivých postupů.**
- 1. Jaké softwary byly vyvinuty (jsou vyvíjeny) pro tvorbu nářečních korpusů a slovníků? Jakou funkcionalitou disponují?
- 2. Jaké softwary byly vyvinuty (jsou vyvíjeny) pro kartografické zobrazení dialektologických dat?
- 3. Lze na dialektologická data uplatnit metody strojového učení? A může znamenat zapojení AI do zpracování dat nějaký přínos?
- 4. Jak katalogizovat a uchovávat audiální/audiovizuální data, jak je optimalizovat pro budoucí zpracování a jak je zpřístupňovat? Jaké funkce mají (měly by mít) nářeční fonotéky v zahraničí a jaké strategie ohledně zpřístupňování audiálních dat jsou uplatňovány v ČR?
- 5. Jaké aplikace byly (jsou) vytvořeny pro elektronický sběr materiálu? A nakolik je tento materiál validní?
- 6. Jak je v konkrétních projektech/výzkumech řešena ochrana osobních práv u zapojených respondentů a jakými údaji jsou data opatřována (např. pro budoucí využití v sociolingvistických výzkumech)?



- **Jakým směrem je žádoucí orientovat dialektologické výzkumy nadále?**
- 1. Jaké specifické postupy vyžaduje speciálně zaměřený dialektologický výzkum, orientovaný na sociodialektologii, percepční dialektologii, etnodialektologii aj.?
- 2. Jaké přínosy mají tyto přístupy oproti tradiční dialektologii?

## Literatura

- Goláňová, Hana – Waclawičová, Martina (2019): The DIALEKT corpus and its possibilities. *Jazykovedný časopis* 7, 2, s. 336–344.
- Goldman, Jean-Philippe et al. (2014): A Crowdsourcing Smartphone Application for Swiss German: Putting Language Documentation in the Hands of the Users. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik, Iceland, s. 3444–3447.
- Chambers, Jack – Trudgill, Peter (1998): *Dialectology*. Cambridge: Cambridge University Press.
- Chromý, Jan (2014): Přístupy ke zkoumání jazykové variace a česká jazyková situace. *Naše řeč* 97, 1, s. 1–12.
- Ireinová, Martina (2021): Úkoly české dialektologie na počátku 21. století. *Gwary Dziś* 14, s. 11–16.
- Kristiansen, Tore (2018): Sociodialectology. In: Boberg, Ch. – Nerbonne, J. – Watt, D. (eds.), *The Handbook of Dialectology*. Hoboken: J. Wiley & Sons, s. 107–122.
- Long, Daniel – Preston, Dennis R. (eds.) (2023): *Handbook of perceptual dialectology. Volume 2*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Möller, Robert – Elspaß, Stephan (2008): Erhebung dialektgeographischer Daten per Internet: ein Atlasprojekt zur deutschen Alltagssprache. In: Elspaß, S. – König, W. (eds.), *Sprachgeographie digital. Die neue Generation der Sprachatlanten (mit 80 Karten)*. Hildesheim: Olms, s. 115–132.
- Preston, Dennis R. (ed.) (1999): *Handbook of perceptual dialectology. Volume 1*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Šimečková, Marta et al. (2022–): *Jazyková paměť regionů České republiky. Metody strojového učení pro uchování, dokumentaci a prezentaci nářečí českého jazyka*. [Webová prezentace projektu NAKI III.] Dostupné z: <https://www.jamap.cz/>.
- Vaux, Bert – Golder, Scott (2003): *The Harvard Dialect Survey*. Cambridge: Harvard University Press.
- Voženílek, Vít et al. (2023): Tvorba nářečních map s využitím ArcGIS Maps for Adobe Creative Cloud. *ArcRevue*, 1, s. 6–13.
- Voženílek, Vít et al. (2022): Mapping, synthesis and visualization of Czech dialects. *International Journal of Cartography* 8, 1, s. 148–163.

# Obsah

*Zpracování nářečních dat v mluvených a psaných korpusech ČNK*

Hana Goláňová (Univerzita Hradec Králové)

*Dialektologie a vlivy jazykových standardů a koineisace na rozrůžňování topolektálních kontinuí*

Jiří Jančík (Univerzita Karlova)

*Automatické převody nářečních přepisů metodami strojového učení*

Martin Karafiát (Vysoké učení technické v Brně), Marta Šimečková (Ústav pro jazyk český AV ČR)

*Interdisciplinární přístupy v dialektologickém výzkumu*

Adéla Kereková (Ústav českého jazyka, Filozofická fakulta Masarykovy univerzity v Brně)

*Nářečová diverzita v mikroregione z pohľadu percepčnej dialektológie*

Mária Mikolajová (Jazykovedný ústav Ľudovíta Štúra SAV)

*Problem prezentowania historycznego materiału gwarowego w świetle geografii lingwistycznej*

Dorota Krystyna Rembiszewska (Instytut Slawistyki Polskiej Akademii Nauk, Warszawa)

*Geoinformatika a mapové zpracování nářečního materiálu*

Vít Voženílek (Univerzita Paleckého, Olomouc), Martina Ireinová (Ústav pro jazyk český AV ČR)

*Dialekty v Českém národním korpusu – současný stav a nové výzvy*

Martina Waclawičová (Ústav Českého národního korpusu FF UK)

## Zpracování nářečních dat v mluvených a psaných korpusech ČNK

Hana Goláňová (Univerzita Hradec Králové)

Nářeční rysy tradičních teritoriálních dialektů na území České republiky je možné zkoumat také prostřednictvím korpusů mluveného a psaného jazyka, vytvořených v rámci projektu Český národní korpus (ČNK), kde jsou jazykové jevy v jednotlivých korpusech vyhledávány přes jednotné rozhraní KonText. Korpusy mluvené češtiny řady ORAL (v KonTextu spojená verze: oral v1, 2017) a ORTOFON (v KonTextu: ortofon v3, 2024) se zaměřují na současnou podobu mluveného jazyka, zachycenou v celé dospělé populaci a v různých komunikačních situacích. Také v těchto synchronních mluvených korpusech se vyskytují některé regionální a nářeční prvky, protože reprezentují mluvu z území celého Česka.

Teritoriální dialekty a jejich rysy jsou však zachyceny zejména v nářečním korpusu DIALEKT, který uživatelům předkládá nářeční data od nejstarší generace mluvčích. V tomto příspěvku se zaměříme především na představení korpusu DIALEKT, obsahující ve své druhé verzi (v KonTextu: dialekt v2, 2021) nahrávky autentických projevů nářečních mluvčích a jejich přepisy v rozsahu 223 281 textových slov. Zvukové záznamy jsou přepsány ve dvou úrovních; základ tvoří dialektologický přepis (dialekt v2 – dial) a doplňkově je k dispozici i ortografický přepis (dialekt v2 – ort). Rozsahem se jedná sice o malý korpus, ale s poměrně přesnými transkripty zvukových záznamů pořízenými prostřednictvím prepisovacího programu ELAN. Tento korpus je možné jako jediný z korpusů mluveného jazyka využít ke studiu teritoriálních nářečních hláskových jevů a archaických nářečních rysů (Goláňová, Waclawičová 2019b).

Korpusy mluveného jazyka řady ORAL (spojená verze korpusů – oral v1: 5 368 392 textových slov) a ORTOFON (ortofon v3: 2 445 792 textových slov) jsou už svým rozsahem větší a lze je za určitých podmínek využít pouze ke studiu morfologických a lexikálních regionálních jevů, které jsou většinou interdialektické povahy.

Regionální a nářeční data se dostávají rovněž do synchronních psaných korpusů řady SYN (spojená verze celé řady – v KonTextu – syn v12, 2023, obsahuje 5 175 milionů textových slov), a to především prostřednictvím beletrie a oblastní publicistiky. V tomto případě je důležité si uvědomit, že jazyková data v psaných korpusech nejsou autentickými nářečními projevy a mohou být tedy stylizována.

Jako doplněk ke korpusům mluvené češtiny slouží mapová aplikace MAPKA, prezentující nářeční rysy českých dialektů a mluvenou češtinu na území celé ČR. Ve své druhé verzi (2023) je rozšířena o další nářeční ukázky z korpusu DIALEKT a nově se zde nacházejí ukázky současné mluvy z korpusů ORAL a ORTOFON. Podobně slouží jako doplněk k mluveným korpusům, především ke korpusu DIALEKT, Archiv diferenčních hlásek nářečí českého jazyka (2019), veřejně dostupný na stránkách ČNK.

Všechny výše uvedené korpusy a nástroje vzniklé v ÚČNK FF UK jsou prakticky použitelné jak pro odbornou veřejnost (např. pro badatele provádějící výzkumy zacílené na dialekt, mluvený jazyk, výzkum spisovnosti a nespisovnosti), ale i pro neobornou veřejnost (např. pro pedagogy jako praktické pomůcky při výuce na všech typech škol, Waclawičová, Goláňová 2018–19, Nachlingerová 2022). Uživatelům se touto cestou nabízejí prostředky, umožňující jim uvažovat o jazyce z různých úhlů pohledu, mnohdy novým způsobem.

## Bibliografie

- Goláňová, Hana (2015): A new dialect corpus: DIALEKT. [W:] *Proceedings of the Eight International Conference Slovko 2015 (Natural Language Processing, Corpus Linguistics, Lexicography)*, (red. K. Gajdošová et al.), Lüdenscheid, s. 36–44.
- Goláňová, Hana – Kupka, Karel (2022): *Mapa nářečí českého jazyka* [online]. Verze ze dne 4. 1. 2022, Ústav Českého národního korpusu FF UK, Praha. [cit. 2024-24-03]. <http://www.korpus.cz>
- Goláňová, Hana – Waclawičová, Martina (2019a): The DIALEKT corpus and its possibilities. *Jazykovedný časopis*, 2/2019, s. 336–344.
- Goláňová, Hana – Waclawičová, Martina (2019b): Vybrané hláskové jevy severovýchodočeské nářeční oblasti v korpusu DIALEKT. *Korpus – Gramatika – Axiologie*, 20/2019, s. 3–15.
- Goláňová, Hana – Waclawičová, Martina (2019c): *Archiv diferenčních hlásek nářečí českého jazyka* [online]. Verze ze dne 28. 11. 2019, Ústav Českého národního korpusu FF UK, Praha. [cit. 2024-24-03]. <http://www.korpus.cz>
- Goláňová, Hana – Waclawičová, Martina (2021a): Mapka: A Map Application for Working with Corpora of Spoken Czech. *Jazykovedný časopis*, 2/2012, s. 502–509.
- Goláňová, Hana – Waclawičová, Martina et al. (2021b): *DIALEKT: nářeční korpus, verze 2 z 23. 12. 2021* [online]. Ústav Českého národního korpusu FF UK, Praha. [cit. 2024-24-03]. <http://www.korpus.cz>
- Goláňová, Hana – Waclawičová, Martina et al. (2023): *Mapka: mapová aplikace pro korpusy mluvené češtiny* [online]. Verze 2.0. Ústav Českého národního korpusu FF UK, Praha. [cit. 2024-24-03]. <http://korpus.cz/mapka>.
- Komrsková, Zuzana – Kopřivová, Marie et al. (2017): New spoken corpora of Czech: ORTOFON and DIALEKT. *Jazykovedný časopis*, 2/2017, s. 219–228.
- Kopřivová, Marie – Laubeová, Zuzana et al. (2020): *ORTOFON v2: Korpus neformální mluvené češtiny s víceúrovňovým přepisem* [online]. Ústav Českého národního korpusu FF UK, Praha. [cit. 2024-24-03]. <http://www.korpus.cz>
- Kopřivová, Marie – Lukeš, David et al. (2017): *ORAL: korpus neformální mluvené češtiny, verze 1 z 2. 6. 2017* [online]. Ústav Českého národního korpusu FF UK, Praha. [cit. 2024-24-03]. <http://www.korpus.cz>
- Křen, Michal – Cvrček, Václav et al. (2023): *Korpus SYN, verze 12 ze 29. 12. 2023.* [online]. Ústav Českého národního korpusu FF UK, Praha. [cit. 2024-24-03]. <http://www.korpus.cz>
- Lukeš, David – Kopřivová, Marie et al. (2024): *ORTOFON v3: Korpus neformální mluvené češtiny s víceúrovňovým přepisem* [online]. Ústav Českého národního korpusu FF UK, Praha. [cit. 2024-26-06] <http://www.korpus.cz>
- Nachlingerová, Kristýna (2022): *Vybrané nářeční jevy v korpusu DIALEKT a jejich prezentace ve výuce češtiny*. Diplomová práce, Jihočeská univerzita v Českých Budějovicích, Filozofická fakulta.
- Waclawičová, Martina – Goláňová, Hana (2018–19): Nářeční korpus DIALEKT a jeho použití ve výuce češtiny. *Český jazyk a literatura*, 3/2018–19, s. 127–133.

## Dialektologie a vlivy jazykových standardů a koineisace na rozrůžňování topolektálních kontinuí

Jiří Jančík (Univerzita Karlova)

Ve svém příspěvku pojednáváme o jedné z velikých výzev areálové dialektologie, stavu, jenž může svou podstatou být pohnutkou k přehodnocování podob dalšího pokračování disciplíny. Je jím metodologická otázka, co přesně má být předmětem jejího výzkumu.

Pokud je rámcovou definicí areálové dialektologie to, že je oborem zabývající se studiem nářečí na území národního jazyka (cf. například Kloferová, 2017a), pak je zkoumáním komplexních kategorií lingvistických, sociologických a historických (Chloupek, 1971; Chloupek, 1986 a Kloferová, 2017b) jeho mluvčích prováděno ideálně na osobách, jež by bylo lze označit jako rodilé mluvčí s plnou jazykovou jistotou (tak, jak se ji snaží vymezit například Hymes, 1982; Kerbrat-Orecchioni, 1994; Calvet, 1999; Coste, 2001 a Bretegnier – Ledegen, 2008). Takovíto informanti jsou autentickými uživateli topolektu, a poskytují o jazykové varietě autentickou informaci s plnou jazykovou jistotou.

V současné době je ovšem vlivem zvýšené interregionální (a internacionální) migrace, vyššího celospolečenského vystavení, v rámci publika masmedií a v prostředí národního školství, efektům standardizovaných jazykových variet celonárodních a koineizací již od úrovně interregionů sousedících regionů pro nářečí původní ideální kontinuum rodilých mluvčích rozrůžněno tak, že je možné jej v logice jeho gradientu popsat koncepty definovanými původně kreolisty a sociolingvisty (cf. například Stewart, 1965; Bickerton, 1973 a Bickerton, 1975), *bazilekt*, *mezolekt* a *akrolekt*, které ovšem chápeme ve smyslu variací stupně jazykové kompetence (generativistická *linguistic competence*) nebo spíše komunikační / komunikativní kompetence (komunikační a sociolingvistické *communicative competence*) tak, jak to činí například romanisté Blampain – Goosse – Klinkenberg – Wilmet (1997: 322-324). Takto vzniklou škálu nízkého – středního – vysokého stupně komunikační / komunikativní kompetence u mluvčích nářečí je lze ještě zleva doplnit stupněm mluvčích s pouze naprosto rudimentální, střípkovou, kompetencí, odpovídající situaci mluvčích, již si zejména z rodinného prostředí osvojili pouze několik málo prostředků, často na úrovni komunikační kuriozity.

V podobném stavu rozrůžnění původního ideálního kontinua stojí ovšem areálová dialektologie před metodologickým dilematem přijetí, či odmítnutí informantů ze zmíněných tří nižších stupňů. Pokud tak, jako tomu bylo například v okamžiku, kdy v druhé polovině XIX. století lingvistické kolegium zjistilo, že již (románské) dalmátštině hrozí vyhynutí (cf. například Bartoli, 1906; Bartoli, 2000 a Muljačić, 2000), či v současnosti pro celou řadu nářečí sudetoněmecké a východopruské skupiny, se disciplína rozhoduje pro přijetí i takovýchto informantů, musí ovšem změnit svou metodologii a doplnit ji široce o analýzy sémiotické a sémiologické. V takových případech jsou pak i informanti nižších stupňů komunikační / komunikativní kompetence vhodným zdrojem poznání nářečí, pouze však za předpokladu, že jejich svědectví budou pečlivě analyzována v parametrizaci sémiotiky a sémiologie, to jest v termínech *extenze*, *intenze*, *diferenční hodnoty* a podobně.

## Bibliografie

- Bartoli, M. G. (2000): *Il Dalmatico. Resti di un'antica lingua romanza parlata da Veglia a Ragusa e sua collocazione nella Romània appennino-balcanica*. Roma: Istituto della Enciclopedia Italiana Fondata da Giovanni Treccani.
- Bartoli, M. G. (1906): *Das Dalmatische*. Bd. 1, 2. Wien
- Bickerton, D. (1975): *Dynamics of a Creole System*. London: Cambridge University Press.
- Bickerton, D. (1973): The Nature of a Creole Continuum. *Language*. 49. str. 640-669.
- Blampain, D., Goosse, A., Klinkenberg, J.-M., Wilmet, M. (eds.) (1997): *Le français en Belgique. Une langue, une communauté*. Louvain-la-Neuve: Duculot.
- Bretnegnier, A., Ledegen, G. (2008): *L'insécurité linguistique dans le contexte francophone périphérique*, Cahiers de l'Institut de linguistique de Louvain-la-Neuve. vol. 1, 2. Louvain-la-Neuve.
- Calvet, L.-J. (1999): *Pour une écologie des langues du monde*. Paris: Plon.
- Coste, D. (2001): Compétence bi/plurilingue et insécurité linguistique. In: Atti del Convegno Valle d'Aosta regione d'Europa: *l'Educazione bi/plurilingue, ponte verso la cittadinanza europea*. L'Ecole Valdotaïne, (supplément au n. 54).
- Hymes, D. H. (1982): *Vers la compétence de communication*. Paris: Hatier-Didier.
- Chloupek, J. *Aspekty dialektu*, 1971.
- Chloupek, J. *Dichotomie spisovnosti a nespisovnosti*, 1986.
- Kerbrat-Orecchioni, C. (1994): *Les interactions verbales*. Paris: Armand Colin.
- Kloferová, S. (2017a): DIALEKTOLOGIE. In: Petr Karlík, Marek Nekula, Jana Pleskalová (eds.), *CzechEncy - Nový encyklopedický slovník češtiny*.  
URL: <https://www.czechency.org/slovník/DIALEKTOLOGIE> (poslední přístup: 25. 3. 2024)
- Kloferová, S. (2017b): DIALEKT. In: Petr Karlík, Marek Nekula, Jana Pleskalová (eds.), *CzechEncy - Nový encyklopedický slovník češtiny*.  
URL: <https://www.czechency.org/slovník/DIALEKT> (poslední přístup: 25. 3. 2024)
- Muljačić, Ž. (2000): *Das Dalmatische. Studien zu einer untergegangenen Sprache*. Köln / Weimar / Wien: Böhlau Verlag.
- Stewart, W. A. (1965): "Urban Negro speech: Sociolinguistic factors affecting English teaching", In: Shuy, R., Davis, A., Hogan, R. (eds.) (1965): *Social dialects and language learning*. Champaign (Illinois): National Council of Teachers of English. str. 10-18.

**Automatické převody nářečních přepisů metodami strojového učení**

Martin Karafiát (Vysoké učení technické v Brně), Marta Šimečková (Ústav pro jazyk český AV ČR)

Součástí sbírek dialektologického oddělení Ústavu pro jazyk český AV ČR, v. v. i., je i Archiv zvukových záznamů nářečních promluv. Je budován od r. 1952 a v současnosti obsahuje přes 2 000 záznamů souvislých mluvených projevů z různých nářečních oblastí. K daným nahrávkám jsou postupně opatřovány přepisy, nezbytné pro vyhodnocení jazykového materiálu formou studií, slovníků či atlasů. Tyto přepisy jsou pořizovány dialektologickou transkripcí, jejíž základy byly stanoveny *Pravidly pro vědecký přepis dialektických zápisů* (1943, rozšířená verze 1951; drobné revize *České nářeční texty*, 1976, s. 7–9).

Bohatým zdrojem nářečních dat jsou však i záznamy etnografického, historiografického nebo literárního rázu zapsané tzv. folklorní transkripcí (Šimečková, v tisku). V ní se přihlíží k fonetické stránce řeči jen částečně – není reflektován výslovnostní rozdíl mezi *i* a *y*, nejsou zachycovány asimilace a měkkosti u konsonantů, neužívají se speciální znaky pro diferenční hlásky, např. pro tvrdé či obalované *l* nebo pro široké vokály.

Práce s takovými zdroji je problematická – kvůli neexistenci jednotné normy jsou záznamy různě variovány a v případě jejich zapojení do dialektologických výstupů je nutné sjednocení daného materiálu po formální stránce, tedy převod do dialektologické transkripce. Žádoucí je však i převod druhým směrem, tedy z transkripce dialektologické do transkripce folklorní, a to při prezentaci nářečního materiálu laickým čtenářům, pro které by mohla být tradiční dialektologická transkripce málo srozumitelná. Z toho důvodu je potřebné vytvořit pravidla pro obousměrný převod těchto transkripcí. A následně i software, který by umožňoval automatizaci takovýchto převodů, tedy i zefektivnění prací, které jsou nyní prováděny ručně.

K vyvinutí takového softwaru, na kterém se v současnosti s brněnskými dialektology podílejí odborníci z Vysokého učení technického, budou využity metody strojového učení. V práci budou srovnány jak přístupy statistické (např. Bisani–Hermann, 2008), tak neurální (Graves et al., 2005; Raffel et al., 2019). Půjde tak o první automatický převaděč tohoto druhu specializovaný právě na česká nářeční data. V příspěvku budou podrobněji představeny jak zmíněné transkripční typy a problémy spojené s prací s nimi, tak příprava textových dat pro strojové učení. A dále budou analyzovány a srovnány výsledky jednotlivých přístupů pro automatický převod.

## Bibliografie

BISANI, M. a N. HERMANN. Joint-Sequence Models for Grapheme-to-Phoneme Conversion. *Speech Communication* 50, č. 5, 2008, s. 434–451.

GRAVES, A., S. FERNÁNDEZ a J. SCHMIDHUBER. Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition. In: DUCH, W., J. KACPRZYK, E. OJA a S. ZADROŽNY, eds. *Artificial Neural Networks: Formal Models and Their Applications – ICANN 2005 (LNTCS, volume 3697)*. Springer Verlag: Berlin – Heidelberg, 2005, s. 799–804.

LAMPRECHT, A. et al. *České nářeční texty*. Praha: Státní pedagogické nakladatelství, 1976.

*Pravidla pro vědecký přepis dialektologických zápisů českých a slovenských*. Praha: Česká akademie věd a umění, 1943.

*Pravidla pro vědecký přepis dialektologických zápisů českých a slovenských*. Praha: Česká akademie věd a umění, 1951. [2., revid. vyd.]

RAFFEL, C. et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research* 21, 2020, s. 1–67.

ŠIMEČKOVÁ, Marta. Audiální data: sběr, archivace, katalogizace a jejich příprava pro strojové učení. In: *Metodika pro převod strukturovaných znalostí z oboru dialektologie do strojového učení*. V tisku.



### Interdisciplinární přístupy v dialektologickém výzkumu

Adéla Kereková (Ústav českého jazyka, Filozofická fakulta Masarykovy univerzity v Brně)

Současné výzkumné metody v dialektologii se neobejdou bez inovativních, mnohdy i interdisciplinárních přístupů. Ty mohou odhalit mj. dříve netušené sociolingvistické kontexty, týkající se například generačních posunů v užívání jazyka (Chromý, 2017, s. 118, 149–159; Milroy-Gordon, 2012, s. 47, 49–50; Coulmas, 2005, s. 52–66) či náhledu na jazyk mluvčími (Meyerhoff, 2011, s. 69–83). Percepce vlastního jazyka mluvčími je jednou z oblastí, která by neměla být v moderním nářečním výzkumu opomenuta (Chromý, 2014, s. 7). Mluvčí se tak může stát středobodem výzkumu, a nikoliv jen kategoriálně vymezeným vehikulem jazyka. Spojení bádání o jazykovém systému na jedné straně a náhledu mluvčích na jeho vlastní užívání na straně druhé dává vzniknout jedinečnému (multidimenzionálnímu) pohledu na nářečí, který zachycuje nejen individuální charakteristiky, ale i mezigenerační a sociálně-lingvistickou dynamiku odrážející vývoj mluveného jazyka v jeho celistvosti (Berruto, 2010, s. 227, 236–237; Chromý, 2009, s. 258–259).

Vhodným materiálem pro takový nářeční výzkum je mluvený jazyk, jenž je komplementárně opatřen podklady o vědomé jazykové volbě, informacemi o mluvčím a názory na percepci vlastního jazyka. V případě tohoto příspěvku jsme použili materiál získaný terénním výzkumem ve městě Uničově a jeho nejbližším okolí mezi lety 2022 až 2023. Mluvčí v rozmezí od 21 do 93 let odpovídali v dotazníkovém výzkumném šetření na otázky týkající se jazyka ve městě a okolí (teritoriální hledisko), na otázky týkající se sociálního ukotvení (sociolingvistické hledisko) a na otázky na percepci vlastního jazyka (percepční hledisko). Prolnutí vyjmenovaných oblastí umožňuje zaměřit se na otázku, jak souvisí vlastní pohled na hodnocení mluvy jako nářeční, příznakové (vědomá variantnost) a skutečné užití nářečních variant v mluveném jazyce a jak se mění toto hodnocení, resp. užívání variant mezi generacemi.

Pro zodpovězení položených otázek jsme se zaměřili na realizaci protetického *v-*, na výskyt střidnic *é*, *ó*, resp. *ej*, *ou* (na místě stč. *y/i*, *ú*) a okrajově též na úžení *é* > *í*. Rozdíly v užití uvedených jevů v mluveném jazyce (nahrávaný materiál) a při vědomé volbě (odpovědi z dotazníkového šetření) mohou ukazovat na korelaci mezi percepcí vlastního jazyka a ukotvením variant jevů v jazykovém systému a jejich užití.

## Bibliografie

BERRUTO, G. Identifying dimensions of linguistic variation in a language space. In Auer, P. & Schmidt, J. E., eds., *Language and Space: An International Handbook of Linguistic Variation*. Volume 1. Berlin: Mouton de Gruyter, 2010, s. 226–241.

COULMAS, F. *Sociolinguistics: the study of speakers' choices*. Cambridge: Cambridge University Press, 2005. ISBN 0-521-83606-9.

CHROMÝ, J. Postoje k jazyku, technika spojitých masek a čeština. *Naše řeč*. 2009, č. 5, s. 252–262

CHROMÝ, J. Přístupy ke zkoumání jazykové variace a česká jazyková situace. *Naše řeč*. 2014, č. 1, s. 1–12.

CHROMÝ, J. *Protetické v- v češtině*. Trivium. Praha: Filozofická fakulta Univerzity Karlovy, 2017. ISBN 978-80-7308-717-3.

MEYERHOFF, M. *Introducing sociolinguistics*. Second edition. London: Routledge, 2011. ISBN 978-0-415-55006-2.

MILROY, L. a GORDON, M. J. *Sociolingvistika: metody a interpretace*. Praha: Karolinum, 2012. ISBN 978-80-246-2125-8.

**Nárečová diverzita v mikroregióne z pohľadu percepčnej dialektológie**

Mária Mikolajová (Jazykovedný ústav Ľudovíta Štúra SAV)

Percepčná dialektológia predstavuje v kontexte slovenskej dialektológie isté novum, ktoré prináša do terénnych výskumov zaujímavé metódy a do teoreticko-metodologického záberu zmenu perspektívy na skúmanie dialektov a jazykov. Nejde však o izolovanú disciplínu. Naopak, percepčnú dialektológiu vnímame ako prirodzené pokračovanie a rozširovanie pozn(áv)ania nárečí a jazykov a ich fungovania na slovenskom území. Pokladáme za dôležité vsadiť zahraničné inšpirácie do slovenského dialektologického kontextu a inovatívne metódy výskumu aplikovať v našom jazykovom a nárečovom priestore. Na základe vlastného výskumu tak chceme predstaviť opodstatnenosť percepčného aspektu v lingvistickom skúmaní jazykových postojov či ideológií a tiež poukázať na prínos percepčnej dialektológie k explanácii jazykových zmien. Výskum metódou kreslenej mapy a riadeného rozhovoru bol vykonaný v okrese Stará Ľubovňa, v ktorom bezprostredne interagujú rusínske, nemecké, šarišské, rómske a goralské nárečové a jazykové komunity na pozadí štandardného slovenského jazyka s prvkami jednotlivých nárečí. Hodnotenia jazykov a nárečí používaných na malom geografickom území z pohľadu samotných používateľov upriamuje pozornosť lingvistov na to, ako nárečovú a jazykovú diverzitu reflektujú ľudia bez špecifického lingvistického vzdelania a ako práve cez jazyk identifikujú seba či svoju kultúru, identitu a komunitu v opozícii k ostatným nárečovým a jazykovým skupinám.

## **Bibliografie**

CRAMER, Jennifer a Chris MONTGOMERY (eds). *Cityspaces and Perceptual Dialectology. Global perspectives on non-linguists' knowledge of the dialect landscapes*. Boston/Berlín: Walter de Gruyter. 2016. ISBN 978-1-61451-477-0.

GAVUROVÁ, Miroslava. *Dialekt ako prejav identity*. Fintice: FACE – Fórum alternatívnej kultúry a vzdelávania, 2021. ISBN 978-80-89763-62-7.

LONG, Daniel a D. R. PRESTON. *Handbook of Perceptual Dialectology*. Vol. 2. Amsterdam/Philadelphia: John Benjamins Publishing Company. 2002.

MÚCSKOVÁ, Gabriela. *Regionálne dialekty a ľudová dialektológia*. Holub, Zbyněk (ed.), *Dialektologie a geolingvistika v současné střední Evropě III*. Opava: Slezská univerzita v Opavě. 2018, s. 191–211.

PRESTON, D. R. *Handbook of Perceptual Dialectology*. Vol. 1. Amsterdam/Philadelphia: John Benjamins Publishing Company. 1999.

SAICOVÁ ŘÍMALOVÁ, Lucie. *Laická jazykověda*. *Časopis pro moderní filologii* 102/1. 2020, s. 24–35.

**Problem prezentowania historycznego materiału gwarowego w świetle geografii lingwistycznej**

Dorota Krystyna Rembiszewska (Instytut Sławistyki Polskiej Akademii Nauk, Warszawa)

Historyczne materiały gwarowe stanowią z jednej strony nieprzecenione źródło do badań poszczególnych dialektów, z drugiej zaś niosą problem wiarygodności utrwalonego materiału. Bez wątpienia jednak umożliwiają one w dużym stopniu rekonstrukcję gwar i wyznaczenie dawnych zasięgów wyrazów o charakterze regionalnym.

Ankiety do *Niemieckiego atlasu językowego (Deutscher Sprachatlas)*, z czterdziestoma zdaniem w języku niemieckim do przetłumaczenia, stanowią specyficzne źródło (Wenker – Lameli i in. 2013a, 2013b). Ta swoistość wynika ze sposobu gromadzenia materiału – metoda korespondencyjna obciążona jest często przypadkowością w doborze informatorów, niekiedy nierzetelnością lub niezrozumieniem wypełniającego kwestionariusz. Formuła badawcza, jaką są ankietowe badania korespondencyjne, daje szansę na kartograficzne przedstawienie zasięgów danych jednostek leksykalnych, ukazanie geograficznego przenikania poszczególnych wyrazów pochodzących z różnych języków, które odnoszą się do tych samych desygnatów.

Przydatność tych materiałów do badania dialektów słowiańskich wykazali badacze opracowujący zbiory ankiet z Kaszub (Taborska – Rzetelska-Feleszko 2009), Śląska Opolskiego (Feleszko 1999a, 1999b), Łużyc (Stone 2003).

Ankiety G. Wenkera dzięki gęstej siatce punktów są cennym źródłem do ustalania zasięgów zjawisk językowych. Zastosowanie metod geografii lingwistycznej stwarza możliwość porównania ich z innymi areałami gwarowymi i jednocześnie pokazuje zmiany (przez porównanie z materiałami późniejszymi), jakie dokonują się na danym terenie.

Ankiety z obszaru polskiego od dawna cieszą się zainteresowaniem badaczy (Stone 1990, Feleszko – Siatkowski 1994, Rembiszewska 2010), ale dotychczas nie opublikowano całościowego opracowania. Swego rodzaju próbą ukazania danych Wenkerowskich metodą geografii lingwistycznej jest monografia dotycząca niewielkiego obszaru dawnych Prus Wschodnich – dzisiejszych wschodnich Mazur (Rembiszewska 2020).

Zatem pozostaje do rozstrzygnięcia ważny problem badawczy, w jaki sposób, z wykorzystaniem najnowszych metod kartograficznych, przedstawić wyraziście i kompletnie dane historyczne, by stanowiły paralele do najnowszych danych i dawały jak najszerszy obraz dawnego stanu gwary.

## Bibliografie

- Feleszko, Kazimierz (1999a), *Sześć tekstów prudnickich z archiwum Wenkera*, [w:] *O kształcie języka. Studia i rozprawy*, red. Bogusław Wyderka, Opole, s. 65–73.
- Feleszko, Kazimierz (1999b), *Charakterystyka punktów prudnickich w archiwum Deutscher Sprachatlas*. *Prace Filologiczne*, t. 44, s. 173–178.
- Feleszko, Kazimierz – Siatkowski, Janusz (1994): *Polskie XIX-wieczne materiały gwarowe w Marburgu*. *Rocznik Towarzystwa Naukowego Warszawskiego*, 45, 52–57.
- Popowska-Taborska, Hanna – Rzetelska-Feleszko, Ewa (2009): *Dialekty kaszubskie w świetle dziewiętnastowiecznych materiałów archiwalnych: Prezentacja i opracowanie kaszubskich materiałów językowych zebranych przez Georga Wenkera w latach 1879–1887*, Warszawa.
- Rembiszewska, Dorota K. (2010): *Wartość XIX-wiecznych kwestionariuszy Georga Wenkera do badania przeszłości ludności mazurskiej*. *Studia Łomżyńskie*, 21, 121–127.
- Rembiszewska, Dorota K. (2020): *Gwary Mazur wschodnich w XIX wieku (na podstawie ankiet Georga Wenkera do Niemieckiego atlasu językowego)*, Warszawa.
- Stone, Gerald (1990): *Polskie materiały dialektyczne w archiwum Niemieckiego atlasu językowego w Marburgu*. *Język Polski*, 70 (3–4), s. 124–127.
- Stone, Gerald (2003): *Der erste Beitrag zur sorbischen Sprachgeographie: Aus dem Archiv des Deutschen Sprachatlas*. Bautzen.
- Wenker, Georg – Lameli, Alfred – Heil, Johanna – Wellendorf, Constanze (2013a): *Schriften zum Sprachatlas des Deutschen Reichs, Gesamtausgabe, Band 1: Handschriften: Allgemeine Texte. Kartenkommentare 1889–1897*, Hildesheim, Zürich, New York.
- Wenker, Georg – Lameli, Alfred – Heil, Johanna – Wellendorf, Constanze (2013b). *Schriften zum Sprachatlas des Deutschen Reichs, Band 2: Kartenkommentare 1898–1911, Druckschriften: Veröffentlichungen 1877–1895*, Hildesheim, Zürich, New York.

## Geoinformatika a mapové zpracování nářečního materiálu

Vít Voženílek (Univerzita Palackého, Olomouc), Martina Ireinová (Ústav pro jazyk český AV ČR)

V příspěvku budou představeny výsledky spolupráce Katedry geoinformatiky Přírodovědecké fakulty Univerzity Palackého v Olomouci a dialektologického oddělení Ústavu pro jazyk český AV ČR.

Během více než čtyřleté kooperace vznikla řada nástrojů, které umožňují dialektologický materiál zpracovávat efektivněji, než tomu bylo v předcházejících obdobích, a vytvářet kvalitní výstupy prezentující tento materiál novými, moderními formami, zajímavými jak pro odborné, tak pro laické publikum. Mezi zásadní výstupy patří softwary ProPONAR a ProMAP, jež dovolují získávat, snadno spravovat a vizualizovat data v mapách. Dále byl vytvořen webový nářeční geoportál DiaMa, který poskytuje pokročilou vyhledávací a komparační funkcionalitu pro nářeční mapy. Prostřednictvím bodových znaků v mapách je dostupná databáze elektronických dokumentů DiDa s unikátními záznamy nářečí z více než tisíce různých archivních (písemných i zvukových) zdrojů. Databáze InteGra obsahuje nářeční data z lokalit, v nichž se uskutečnil výzkum, a z těchto záznamů je dopočítává pro okolní místa, kde výzkum neproběhl. Tyto nástroje jsou dostupné na webu [ceskanareci.cz](http://ceskanareci.cz) a jsou využívány při tvorbě nářečních map jednotlivých svazků *Atlasu nářečí českého jazyka*. V současné době dialektologicko-geoinformatický tým připravuje nový celouzemní výzkum nářečí českého jazyka (poslední se uskutečnil zhruba před 50 lety) a vytváří nástroje, které by měly usnadnit získávání nářečního materiálu, jeho organizaci a zpracování.

## Dialekty v Českém národním korpusu – současný stav a nové výzvy

Martina Waclawičová (Ústav Českého národního korpusu FF UK)

V kontextu současné lingvistiky představují jazykové korpusy nezbytný nástroj pro výzkum, a to i v oblasti dialektologie. Tento příspěvek se zaměřuje na přehled datových základů a nástrojů určených pro analýzu a výzkum českých dialektů, které vycházejí z Českého národního korpusu (ČNK), a zároveň na nové možnosti dialektologického výzkumu, které korpusové metody poskytují.

Hlavním zdrojem dat pro studium českých nářečí v rámci ČNK je korpus DIALEKT, který kombinuje dialektologický přepis s ortografickým a propojuje ho se zvukovou stopou. Tento korpus se snaží zachytit nejstarší zjištělý stav nářečí, a to prostřednictvím monologických projevů nejstarší generace mluvčích z různých nářečních oblastí na území České republiky a v českých jazykových ostrovech v Polsku. Z korpusu DIALEKT vychází Archiv diferenčních hlásek nářečí českého jazyka, který obsahuje zvukové ukázky všech typických nářečních hlásek v jejich tradiční podobě.

Svůj význam mají pro dialektologii i obecné korpusy neformální mluvené češtiny ORAL a ORTOFON, které poskytují širší obraz variantnosti mluveného jazyka napříč generacemi a regionálními oblastmi České republiky. ORAL zachycuje ve své jediné rovině přepisu typické rysy mluvenosti a pravidelnou variantnost ve výslovnosti. ORTOFON obsahuje vedle ortografického i fonetický přepis. V obou korpusech je přepis propojen se zvukovou stopou a nechybí ani lemmatizace a morfologické značkování.

Korpusy ORAL, ORTOFON a DIALEKT reprezentuje aplikace MAPKA, která zachycuje mj. různé druhy územního členění, především nářečního, přehledy nářečních rysů a nabízí i 61 ukávek mluvených projevů vybraných z uvedených korpusů. Pro dialektologický výzkum je určena funkce zakreslování vyhledaných obcí na mapu pomocí různobarevných markerů, která je vhodná pro tvorbu vlastní mapy. Pokud se týká technického pozadí, aplikace Mapka pracuje s body a polygony uloženými v databázi MariaDB. Server zasílá prohlížeči data ve formátu GeoJSON a ta jsou knihovnou Leaflet vykreslena do mapy jako body a oblasti.

Primární metodou, která se nabízí pro zkoumání nářečních prvků, je frekvenční analýza. Přístupné jsou jí především rysy morfologické a hláskoslovné (s určitým omezením vzhledem k různé interpretaci hlásek na rovině dialektologické a fonetické). Zkoumání lexikální stránky poněkud omezuje velikost korpusů, ale není to na překážku v případě frekventovaných jednotek slovní zásoby, např. kluk, hoch, ogar, synek. Pomocí frekvenční analýzy můžeme sledovat proměny nářečních jevů v čase, a to pomocí srovnávání dvou časových vrstev v nářečním korpusu nebo mluvčích různých generací v obecných korpusech. Je možné kupříkladu sledovat postupující nivelizaci, kdy se z původních prvků nářečí stávají regionální jevy, které mají své místo i v současné mluvené češtině, jsou stále produktivní (tak to musím jít k ortopédoj – severovýchodočeská nářeční oblast) nebo se dokonce šíří mimo původní území (pronikání nářečních jevů do oblastí pohraničí: přišel ten Jardovo vnuk – české pohraničí). Díky frekvenční analýze se prokazuje i postupující lexikalizace nářečních hláskoslovných a morfologických prvků u určitých jednotek slovní zásoby (*zrouna, klucí, s tátem*).

Výzvou může být data-driven přístup k materiálu (Szmrecsanyi, Benedikt – Anderwald, Lieselotte, 2017), který může pomoci při odhalování shod mezi nářečím (nejen sousedními, ale i geograficky vzdálenými – např. okrajová území se zachovanými archaickými prvky) nebo při vymezení nářečních oblastí na základě dat z korpusů (Embleton, Sheila – Uritescu, Dorin – Wheeler, Eric S., 2009). V tomto směru se nabízejí inovativní přístupy, jako je multidimenzionální analýza (MDA) a metoda t-distributed stochastic neighbor embedding (t-SNE), které již dosáhly pozoruhodných výsledků v jiných lingvistických studiích a mohou být aplikovány i na výzkum českých dialektů založený na korpusových datech.



## Bibliografie

- Anderwald, Lieselotte – Wagner, Susanne (2007): FRED – The Freiburg English Dialect Corpus: Applying Corpus-Linguistic Research Tools to the Analysis of Dialect Data. In: Beal, J.C., Corrigan, K.P., Moisl, H.L. (eds) *Creating and Digitizing Language Corpora*. London: Palgrave Macmillan.
- Embleton, Sheila – Uritescu, Dorin – Wheeler, Eric S. (2009): Identifying Dialect Regions: Specific Features vs. Overall Measures Using the Romanian Online Dialect Atlas and Multidimensional Scaling. Leeds, UK: Methods XIII Conference, August 2008. In B. Heselwood & C. Upton (eds.). 2009, *Proceedings of Methods XIII. Papers from the Thirteenth International Conference on Methods in Dialectology, 2008*. Frankfurt am Main: Peter Lang, s. 79–90.
- Goláňová, Hana – Waclawičová, Martina (2019): The DIALEKT corpus and its possibilities. *Jazykovedný časopis* 70, 2, s. 336–344.
- Goláňová, Hana – Waclawičová, Martina (2019): Vybrané hláskové jevy severovýchodočeské nářeční oblasti v korpusu DIALEKT. *KGA* 20, s. 3–15.
- Goláňová, Hana – Waclawičová, Martina (2018): Archiv diferenčních hlásek nářečí českého jazyka. Ústav Českého národního korpusu FF UK, Praha. Dostupné z WWW: <http://www.korpus.cz>
- Goláňová, Hana – Waclawičová, Martina – Lukeš, David (2021): DIALEKT: nářeční korpus, verze 2 z 23. 12. 2021. Ústav Českého národního korpusu FF UK, Praha. Dostupný z WWW: <http://www.korpus.cz>
- Goláňová, Hana – Waclawičová, Martina – Pejcha, Jakub – Čapka, Tomáš – Benešová, Lucie (2023): *Mapka: mapová aplikace pro korpusy mluvené češtiny. Verze 2.0*. ÚČNK FF UK, Praha. Dostupné z WWW: <http://korpus.cz/mapka>.
- Kopřivová, Marie – Laubeová, Zuzana – Lukeš, David – Poukarová, Petra – Škarpová, Marie (2020): ORTOFON v2: Korpus neformální mluvené češtiny s víceúrovňovým přepisem. Ústav Českého národního korpusu FF UK, Praha. Dostupný z WWW: <http://www.korpus.cz>
- Kopřivová, Marie – Lukeš, David – Komrsková, Zuzana – Poukarová, Petra – Waclawičová, Martina – Benešová, Lucie – Křen, Michal (2017): ORAL: korpus neformální mluvené češtiny, verze 1 z 2. 6. 2017. Ústav Českého národního korpusu FF UK, Praha. Dostupný z WWW: <http://www.korpus.cz>
- Szmrecsanyi, Benedikt – Anderwald, Lieselotte (2017): Corpus-Based Approaches to Dialect Study. In *The Handbook of Dialectology* (eds C. Boberg, J. Nerbonne and D. Watt), s. 300–313.
- Waclawičová, Martina (2011): Jazykovězeměpisné aspekty mluvených korpusů Českého národního korpusu. In: *Dialektologie a geolingvistika v současné střední Evropě*. Opava, SIU, s. 144–155.
- Waclawičová, Martina (2009): Regionální mluva v korpusu mluvené češtiny ORAL2006 se zaměřením na situaci v českém pohraničí. *NŘ* 92, 2, s. 72–86.

8

*Empirický výzkum jazyka  
v sociální interakci*

*Empirical research on language  
in social interaction*

organizují / organizers

Petr Kaderka

Martin Havlík

(CZE, ENG)

## Popis panelu / Workshop description

**(CZE)** Interakční přístupy k výzkumu jazyka přinesly v posledních desetiletích řadu inovativních impulzů pro poznání jazyka. Etablovaly se na vědeckém poli pod různými názvy a s různými, byť v mnohém podobnými výzkumnými programy. K nejdůležitějším přístupům patří konverzační analýza (Clift, 2016), etnometodologická analýza interakce (vom Lehn, 2018), interakční sociolingvistika (Rampton, 2020), interakční lingvistika (Couper-Kuhlen, Selting, 2018), výzkum jazykové interakce (Auer, 2014) a výzkum dialogu (Linell, 2001). Kromě klíčového pojmu sociální interakce spojuje tyto přístupy opatrný až skeptický postoj vůči poznatkům z oblasti systémových přístupů k jazyku a vůči nereflektovanému přenášení těchto poznatků do oblasti výzkumu interakce. Hlavním poznávacím znamením je silný důraz na empirický výzkum a detailní analýzu dat.

Cílem panelu je vytvořit prostor pro konstruktivní dialog mezi interakčně založenými badateli různé teoretické konfese. Podmínkou účasti v panelu je detailní představení empirických dat a metodologicky transparentní analýza vybraného příkladu (nebo několika málo příkladů). Panel není tematicky omezen, účastníci panelu se mohou zaměřit na jakékoli jazykové či multimodální jevy, které mají sociálněinterakční charakter a které ukazují, jak „obýváme jednání toho druhého“ (Goodwin, 2018). Přestože panel nesměruje k teoretické syntéze, je namísto, aby se účastníci panelu na základě představené empirické analýzy zamysleli nad vztahy mezi jazykovými (multimodálními) strukturami a interakčními praktikami. Vítány jsou příspěvky založené na datech z různých kulturních prostředí, z různých typů aktivit a z různých jazyků.

Příspěvky v panelu se mohou zaměřit na následující témata:

- střídání mluvčích a komunikační dynamismus
- multimodální konstrukce replik
- členská kategorizace a sekvenční organizace sociální interakce
- konstruování vědění v interakci
- interakce ve výuce
- interakce v sociálních médiích
- interakce v mediálních dialogických sítích
- interakce s umělou inteligencí

**(ENG)** Interactional approaches to language research have provided a number of innovative impulses for the study of language in recent decades. They have emerged in the scientific field under various names and with distinct, though in many ways similar, research programs. Among the most important approaches are conversation analysis (Clift, 2016), ethnomethodological analysis of interaction (vom Lehn, 2018), interactional sociolinguistics (Rampton, 2020), interactional linguistics (Couper-Kuhlen, Selting, 2018), linguistic interaction research (Auer, 2014) and dialogue research (Linell, 2001). Beyond the key notion of social interaction, what unites these approaches is a cautious or even sceptical attitude towards findings from systemic approaches to language and towards the unreflected translation of these findings into the field of interaction research. A major hallmark is the strong emphasis on empirical research and detailed data analysis.

The aim of the panel is to create a space for constructive dialogue between interactionist scholars of varying theoretical persuasions. Participation in the panel requires a detailed presentation of empirical data and a methodologically transparent analysis of a selected single example (or a few examples). The

panel is not restricted thematically; panelists can focus on any linguistic or multimodal phenomena that are social-interactional in nature and that show how we “inhabit each other’s actions” (Goodwin, 2018). Although the panel does not aim at a theoretical synthesis, it is appropriate for panelists to reflect on the relations between linguistic (multimodal) structures and interactional practices on the basis of the presented empirical analysis. Presentations based on data from different cultural settings, different types of activities and different languages are welcome.

Submissions to the panel may focus on the following topics:

- turn taking and communicative dynamism
- multimodal turn construction
- members’ categorizations and sequential organization of social interaction
- constructing knowledge in interaction
- classroom interaction
- interaction on social media
- interaction in media dialogical networks
- interaction with artificial intelligence

### **Literatura / References**

Auer, P. 2014. *Jazyková interakce* [Linguistic Interaction]. Praha: NLN.

Clift, R. 2016. *Conversation Analysis*. Cambridge: Cambridge University Press.

Couper-Kuhlen, E., Selting, M. 2018. *Interactional Linguistics: Studying Language in Social Interaction*. Cambridge: Cambridge University Press.

Goodwin, Ch. 2018. *Co-Operative Action*. Cambridge: Cambridge University Press.

Linell, P. 2001. *Approaching Dialogue: Talk, Interaction and Contexts in Dialogical Perspectives*. Amsterdam: John Benjamins.

Rampton, B. 2020. Interactional sociolinguistics. In: K. Tusting (ed.), *The Routledge Handbook of Linguistic Ethnography*. London: Routledge, 13–27.

vom Lehn, D. 2018. *Ethnomethodologische Interaktionsanalyse: Videodaten analysieren und die Organisation von Handlungen darstellen*. Weinheim: Beltz Juventa.

## Obsah / Table of Contents

*„Podívám se do chytré knížky“: jak pracovníci jazykové poradny pracují se zdroji při zodpovídání telefonických dotazů*

Klára Dvořáková (Ústav pro jazyk český AV ČR)

*Smích v sociální interakci: Čemu se smějeme v pracovních rozhovorech a čeho smíchem dosahujeme?*

Martin Havlík, Petr Kaderka (Ústav pro jazyk český AV ČR)

*Interpretace doprovodného rozhovoru prostřednictvím následného interview*

Jana Hoffmannová (Ústav pro jazyk český AV ČR)

*Sekvenční struktura diskuse po online přednášce*

Jakub Kopecký (Ústav pro jazyk český AV ČR)

*Procedural and substantive clarifications in medical interaction with an AI-based platform*

Jakub Mlynář (HES-SO Valais-Wallis University of Applied Sciences Western Switzerland, Sierre)

*Assessing what can be heard and what can be seen: The multimodal construction of assessments in smartphone-based showing sequences*

Florence Oloff (Leibniz-Institute for the German Language, Mannheim)

*Jak studenti při ústních zkouškách z angličtiny zařazují výrazy z tištěných materiálů do probíhající interakce*

David Ryška (Masaryk University), František Tůma (Vienna University of Economics and Business)

*Mediální dialog v podmínkách autoritářského režimu*

Marián Sloboda (Univerzita Karlova)

*The effect of chosen turn-taking strategies on the perception of the politician's public image*

Alexey Tymbay (Technical University in Liberec)

**„Podívám se do chytré knížky“:  
jak pracovníci jazykové poradny pracují se zdroji při zodpovídání telefonických dotazů**

Klára Dvořáková (Ústav pro jazyk český AV ČR)

Jazyková poradna Ústavu pro jazyk český (JP) nabízí široké veřejnosti jako jednu z forem jazykového poradenství telefonické konzultace. Kdokoliv může zavolat na linku JP a požádat o pomoc s řešením jazykového problému. S obsahem dotazů pokládaných JP (a to nejen telefonicky) se lze seznámit v různých přehledových člancích (např. Polívková, 1972; Svobodová, 1988; Uhlířová, 1999; Smejkalová, 2020), průběhem telefonických rozhovorů se pak zabývá několik článků publikovaných v posledních letech (Dvořáková, 2020; Kopecký, 2022; Beneš, 2022; Dvořáková, 2024). Cílem tohoto příspěvku je rozšířit současné poznání jazykověporadenských interakcí přivedením pozornosti k tomu, jak pracovníci JP reflektují práci s odbornými zdroji při zodpovídání jazykových dotazů.

Když volající položí jazykový dotaz, pracovník JP by mu měl poskytnout pokud možno okamžitou odpověď. Některé dotazy však pracovníci JP nejsou s to zodpovědět ihned, například když jim nejsou známy potřebné informace nebo když si nejsou odpovědi zcela jisti. V takových případech se uchylují k práci s různými odbornými zdroji a danou aktivitu někdy tematizují ve svých promluvách.

Pro zodpovězení otázky, jak pracovníci JP reflektují práci s odbornými zdroji během rozhovoru, bylo analyzováno 102 jejich odpovědí, zaznamenaných na 63 nahrávkách autentických telefonátů do JP, jež byly pořízeny v období od února 2017 do února 2018. Rozhovorů se zúčastnilo celkem osm zaměstnanců JP s různou délkou jazykověporadenské praxe (v řádu měsíců či let) a převážně různí tazatelé (někteří se na nahrávkách objevili prokazatelně vícekrát, u jiných to nelze tvrdit s jistotou, jelikož se někteří volající nepředstavili). Pro rozbor rozhovorů byla využita primárně konverzační analýza (Sacks, 1992; Hutchby, Wooffitt, 2002; Liddicoat, 2007; Sidnell, Stivers, 2013; aj.), s ohledem na téma výzkumu však byly její hranice záměrně překročeny.

Analýza ukázala, že pracovníci JP explicitně signalizovali využití odborných zdrojů ve čtvrtině svých odpovědí. Zprvė popisovali svou činnost (např. *podívám se do chytré knížky; já už sem si nalistoval a už to hledám*)<sup>1</sup>. Zadruhé sdělovali předjímaný výsledek hledání (např. *ale pochybuju o tom, že tam něco takovéhohle bude napsáno*) a zatřetí konečný výsledek (např. *my to máme tady v příručce, máme to tu*). Začtvrté produkovali komentáře ke zdroji, v nichž k němu sdělovali podrobnější informace (např. když mluvčí specifikoval, že bude hledat řešení v knihách, *kteřé vysvětlují původ přímení*). Je však nutno dodat, že pracovníci JP mohou využívat při zodpovídání dotazů různé zdroje, aniž by na to explicitně interakčního partnera upozornili.

Když pracovníci JP sdělují tazatelům, že se pokusí dohledat řešení jejich jazykového problému, předcházejí potenciálním interakčním problémům, které by mohla způsobit absence řešení a/nebo absence vizuálních informací – vzhledem k užitému kanálu komunikace totiž nemají její účastníci možnost monitorovat zrakem neverbálně-nevokální projevy komunikačních partnerů (jako je nahlížení do knihy). Pomocí těchto sdělení také udržují pracovníci JP kontakt s volajícími (ovšem mohou ho též na určitou dobu přerušit, požádat tazatele o vyčkání: *vydržte chvilíčku*). Práce se zdroji dále implikuje, že se pracovníci JP snaží doporučovat pouze řešení, jimiž jsou si jisti a která mohou založit na spolehlivých informacích, a tak tato řešení obhájit. Tematizace práce s různými zdroji tedy hraje v jazykověporadenských interakcích velmi důležitou roli.

<sup>1</sup> Pozn.: V abstraktu je užíván výrazně zjednodušený přepis (částí) replik.

## Bibliografie

BENEŠ, M. (2022): No ale, to je fakt správněj tvar, jo? Nesouhlas tazatelů s odpověďmi. In: DUFEK a kol.: *Jazyková poradna, dobrý den. O češtině a jejích uživatelích*. Praha: NLN, s. 173–193.

DVOŘÁKOVÁ, K. (2020): „Bylo by lepší to celé přestylizovat...“ Příspěvek ke stylistické problematice v jazykové poradně. In: JÍLKOVÁ, L., MRÁZKOVÁ, K., ÖZÖRENCIK, H. (eds.): *Jak je důležité mít styl. Pocta Janě Hoffmannové*. Praha: NLN.

DVOŘÁKOVÁ, K. (2024): „Tady mám takovou větu – a já nevím, jestli je prostě správně napsaná česky“: Příspěvek k problematice pokládání dotazů jazykové poradně. *Naše řeč*, 107(1), s. 21–47.

HUTCHBY, I., WOFFITT, R. (2002): *Conversation Analysis: Principles, Practices and Applications*. Cambridge: Polity.

KOPECKÝ, J. (2022): Divergent interests and argumentation in Czech Language Consulting Center Interactions. In: NEKULA, M., SHERMAN, T., ZAWISZOVÁ, H. (eds.): *Interests and Power in Language Management*. Berlin: Peter Lang, s. 73–99.

LIDDICOAT, A. (2007): *An Introduction to Conversation Analysis*. London: Continuum.

POLÍVKOVÁ, A. (1972): Z jazykové poradny. *Naše řeč*, 55(1), s. 27–33.

SACKS, H. (1992): *Lectures on Conversation*. Vol. 1 & 2. Oxford: Blackwell.

SIDNELL, J., STIVERS, T. (eds.) (2013): *The Handbook of Conversation Analysis*. Chichester: Wiley-Blackwell.

SMEJKALOVÁ, K. (2020): Jazyková poradna ÚJČ v době jarního nouzového stavu. *Jazykovědné aktuality*, 57(3–4), s. 129–133.

SVOBODOVÁ, I. (1988): Analýza dotazů z jazykové poradny. *Naše řeč*, 71(3), s. 118–128.

UHLÍŘOVÁ, L. (1999): O co se zajímají ti, kdo posílají dopisy, faxy a e-maily jazykové poradně? *Naše řeč*, 82(5), s. 225–236.

**Smích v sociální interakci:  
Čemu se smějeme v pracovních rozhovorech a čeho smíchem dosahujeme?**

Martin Havlík, Petr Kaderka (Ústav pro jazyk český AV ČR)

Pro etnometodologickou konverzační analýzu je smích sociální aktivita, kterou účastníci interakce metodicky produkují (Jefferson, 1979, 1983, 1984, 1985, 2004; Jefferson – Sacks – Schegloff, 1987; Glenn, 2003; Glenn – Holt, 2013). Je to aktivita, které si účastníci interakce všímají, kterou interpretují a kterou také citlivě manažují. Je to aktivita variabilní a reportovatelná a také ironicky modulovatelná (např. recipient může svůj odmítavý postoj vůči vtípu vyjádřit monotónním, rytmizovaným a dobře artikulovaným vyslovením trojslabičného „ha-ha-ha“). Jeffersonové (1983, 1985) se pomocí detailního transkribování smíchu podařilo dokázat, že smích není nějakou nekontrolovatelnou erupcí, která by na okamžik rozbořila interakční mechanismy, ale že se aktéři při produkci smíchu neustále ohlížejí na chování druhých, svůj smích ovládají a jsou schopni ho nejen dávkovat, ale také přesně umisťovat do proudu interakce (tj. zahajovat a ukončovat smích vzhledem k promluvám a smíchu druhých). Toto zjištění je pozoruhodné, uvážíme-li, že nejsme schopni smích vyprodukovat na povel; takové pokusy vyznívají jako umělé nebo falešné. Jakmile se však začneme po nějakém impulzu upřímně smát, jsme přitom schopni monitorovat chování druhých a svůj smích ovládat (manažovat) a koordinovat s jinými konverzačními aktivitami (Jefferson – Sacks – Schegloff, 1987).

Smích bývá součástí rozvinutých afiliativních sekvencí (expanded affiliative sequence, Jefferson – Sacks – Schegloff, 1987, s. 159), tj. interakčních sekvencí, při kterých dochází k projevům souznění (affiliation) mezi účastníky. Smíchem lidé reagují nejen na vtípy a humorné poznámky, ale také na případy, kdy se někdo z účastníků interakce projeví příliš upřímně, drsně nebo vulgárně. Cílem těchto nepřístojných promluv přitom není druhé šokovat, ale naopak vytvořit prostředí důvěrnosti, resp. vybídnout ostatní účastníky k jeho společnému vytváření (ibid., s. 160). Smích je podle autorů jakousi přípravou k následnému explicitnímu projevu souznění (affiliation): pokud je účastník interakce ochoten nabídku k důvěrnějšímu tónu přijmout, může se nepřístojné poznámce zasmát, následně něco ve stejném duchu poznamenat a tím demonstrovat, že je (či chce být) s autorem nepřístojné poznámky na stejné vlně, tj. že s ním souzní.

V situacích, kdy lidé hovoří o potížích, problémech či nepříjemnostech, má smích jiný význam a zpravidla také jiný interakční průběh. Když se mluví při líčení svých problémů smějí, ukazují tím, že své problémy považují za zvládnutelné a že jsou sami do určité míry vůči těmto problémům odolní (troubles-resistive, Jefferson, 1984, s. 367). Posluchači se v takové situaci k smíchu nepřidávají, ale odpovídají vážně, aby ukázali svou vnímavost vůči problémům druhého (troubles-receptiveness, ibid.). Ke společnému smíchu sice i tak může dojít, ale děje se tak nikoli v sekvencích, ve kterých mluvčí líčí své problémy či potíže, ale zpravidla ve speciálních sekvencích, které slouží k oddechu od vážného tématu (time-out, ibid.). Jen zde, v tematických time-outech, může dojít k souznění (affiliation) mezi mluvčími prostřednictvím společného smíchu.

Ve svém příspěvku budeme analyzovat několik případů smíchu, které pocházejí z našich terénních audiovizuálních nahrávek. V nich jsme zachytili interakce, během nichž jejich účastníci společně usilovali vyřešit nějaký (pracovní) problém a dosáhnout shody o tom, co a jak udělat, připravit, vyrobit apod. Nešlo to tedy až na jednu výjimku o rozhovor pro rozhovor, jehož primárním cílem je udržovat a posilovat sociální vazby. V prezentaci ukážeme několik momentů, v nichž byl smích využit k poukázání na nějaký (technický) problém. Smích v těchto situacích sice sloužil k vyjádření souznění mezi komunikujícími nebo k zmírnění kritiky, především však byl použit jako jeden z prostředků vzájemné domluvy. Smích byl tedy v těchto interakcích užíván primárně nikoli jako prostředek morálního či postojového souznění, nýbrž jako prostředek souznění instrumentálně-kooperativního.



## Bibliografie

Glenn, P. (2003): *Laughter in Interaction*. Cambridge: Cambridge University Press.

Glenn, P. – Holt, E. (eds.) (2013): *Studies of Laughter in Interaction*. London – New Delhi – Sydney: Bloomsbury.

Jefferson, G. (1979): A technique for inviting laughter and its subsequent acceptance declination. In: G. Psathas (ed.), *Everyday Language: Studies in Ethnomethodology*. New York: Irvington Publishers, s. 79–96.

Jefferson, G. (1983): An exercise in the transcription and analysis of laughter. *Tilburg Papers in Language and Literature*, 35, s. 1–15.

Jefferson, G. (1984): On the organization of laughter in talk about troubles. In: J. M. Atkinson – J. Heritage (eds.), *Structures of Social Action: Studies in Conversation Analysis*. Cambridge – London – New York – New Rochelle – Melbourne – Sydney – Paris: Cambridge University Press – Editions de la Maison des Sciences de l'Homme, s. 346–369.

Jefferson, G. (1985): An exercise in the transcription and analysis of laughter. In: T. van Dijk (ed.), *Handbook of Discourse Analysis, 3: Discourse and Dialogue*. London – Orlando – San Diego – New York – Toronto – Montreal – Sydney – Tokyo: Academic Press, s. 25–34.

Jefferson, G. (2004): A note on laughter in 'male–female' interaction. *Discourse Studies*, 6 (1), s. 117–133.

Jefferson, G. – Sacks, H. – Schegloff, E. (1987): Notes on laughter in the pursuit of intimacy. In: G. Button – J. R. E. Lee (eds.), *Talk and Social Organisation*. Clevedon – Philadelphia: Multilingual Matters, s. 152–205.

## Interpretace doprovodného rozhovoru prostřednictvím následného interview

Jana Hoffmannová (Ústav pro jazyk český AV ČR)

Příspěvek založený na přístupech interakční lingvistiky je zaměřen na analýzu interakce při **doprovodném rozhovoru**, který se uskutečňuje při společné pracovní činnosti účastníků a je pro něj příznačná asymetrie rolí (nadřazený – podřazený). (Srov. k tomu už Hoffmannová – Havlík, 2021; Hoffmannová, 2022, 2023.) Pro výzkumné účely byla pořízena rozsáhlá videonahrávka celodenní pracovní komunikace mezi majitelem (vedoucím) houslařské dílny a jeho zaměstnancem; v tomto příspěvku je využita část rozhovoru o rozsahu 48 minut a její transkript. Analýza úryvku naznačí, jak se vedoucí pracovník snaží oslabovat projevy své nadřazenosti, s partnerem co možná vyrovnaně spolupracovat, dávat mu prostor pro vlastní nápady a iniciativy, zmírňovat příkazový ráz komunikace. Užívá proto např. málo imperativů (*udělej to*), volí raději inkluzivní plurál, neosobní 3. os. sg., infinitivy nebo kondicionály (*to uděláme, to se udělá, a co to udělat takhle, udělal bych to tak...*); často po svém podřízeném opakuje, vyjadřuje souhlas s jeho názory. Dosahuje tak hladkého průběhu rozhovoru a vzájemného porozumění. Může tu ovšem vzniknout otázka, zda jsou tyto jeho interakční praktiky intuitivní, podvědomé, nebo uvědomělé; zda jde o záměrně užívanou, osvědčenou strategii, která usnadňuje a optimalizuje společnou práci.

Zodpovězení této otázky umožní analýza **následného interview** (srov. k tomu např. Neustupný, 1999; Sherman – Homoláč, 2014). Dva výzkumníci, kteří pořizovali původní nahrávku doprovodného rozhovoru, uskutečnili po několikátýdenním časovém odstupu rozhovor s dominantním účastníkem komunikace houslařů a tento dialog opět zaznamenali. Během rozhovoru přehrávali svému respondentovi vybrané úseky výchozího rozhovoru a kladli mu otázky týkající se organizace práce v dílně, vznikajících výrobců, používaných materiálů; zajímali se i o některé profesionalismy a slangové výrazy, a také o postoje respondenta ke spolupracovníkům a o komunikační strategie, které vůči nim uplatňuje. Otázky na tyto strategie podnítily neobyčejně rozsáhlou a souvislou odpověď (v rozsahu asi 6–7 minut), v níž vedoucí dílny podrobně vysvětluje své chování ke spolupracovníkovi. Analýza se soustředí na tuto část následného interview a na úryvku z tohoto rozhovoru bude předvedeno, že si vedoucí pracovník uvědomuje své interakční praktiky a hodnotí je jako „manipulaci“. Vždy se snaží vtáhnout svého spolupracovníka/podřízeného do diskuse o pracovních úkonech, nechat ho, aby „si řekl svoje“, aby si myslel, že na vše přišel on sám. Občasným „postrčením“ (*nudge*, „šťouch“; srov. Thaler – Sunstein, 2010) se mu daří dosáhnout toho, že partner domluvený postup „vezme za své“, ztotožní se s ním. Manipulace pro nadřazeného zřejmě znamená, že „obývá jednání toho druhého“ (Goodwin, 2018).

Pojem **manipulace** tu asi není úplně na místě, ostatně oba výzkumníci při komentování doprovodného rozhovoru zdůrazňovali spíše **spolupráci** obou účastníků. A sám vedoucí pracovník chápe manipulaci spíše pozitivně: pomáhá mu překonávat znalostní asymetrii, dosáhnout vzájemné interakční akceptace a zdárného výsledku práce. Jeho interakční praktiky by bylo možno umístit i do přechodného pásma mezi manipulací a **persvazí** (Srpová, 2005).

Příspěvek ukazuje, že následné interview může přinést nejen četné doplňující informace o obsahu a průběhu výchozího rozhovoru, ale může elicitovat i další zajímavé poznatky o interakčních strategiích účastníků a jejich sebereflexi.

## **Bibliografie**

Goodwin, Ch. 2018. *Co-operative Action*. Cambridge: Cambridge University Press.

Hoffmannová, J. 2022. Styl doprovodných dialogů. *Jazykovedný časopis*, roč. 73, č. 1, s. 86–100.

Hoffmannová, J. 2023. Doprovodný rozhovor a následné interview: možnosti analýzy transinterakčního managementu. *Jazykovedný časopis*, roč. 74, č. 2, s. 495–508.

Hoffmannová, J. – Havlík, M. 2021. Opakování jako součást utváření pracovního (doprovodného) rozhovoru. *Časopis pro moderní filologii*, roč. 103, č. 2, s. 210–227.

Neustupný, J. V. 1999. Následné (follow-up) interview. *Slovo a slovesnost*, roč. 60, č. 1, s. 13–18.

Sherman, T. – Homoláč, J. 2014. Management summaries and the follow-up interview in language biography research. *Slovo a slovesnost*, roč. 75, č. 4, s. 294–324.

Srpová, H. 2005. Manipulace a persvaze – janusovské dilema. In: H. Srpová (ed.), *Metody a prostředky přesvědčování v masových médiích*. Ostrava: Ostravská univerzita, s. 200–208.

Thaler, R. H. – Sunstein, C. R. 2010. *Nudge (Šťouch): jak postrčit lidi k lepšímu rozhodování o zdraví, majetku a štěstí*. Zlín: Kniha Zlín.

**Sekvenční struktura diskuse po online přednášce**

Jakub Kopecký (Ústav pro jazyk český AV ČR)

Komunikační online platformy umožňují vstupovat do interakce lidem nesdílejícím společný interakční prostor. Přenos obrazu a zvuku v reálném čase poskytuje účastníkům videohovoru částečnou iluzi bezprostřední interakce tváří v tvář. Ve srovnání s interakcí realizovanou ve fyzické spolupřítomnosti však používané technologie přinášejí také mnohá omezení (srov. Mlynář – González-Martínez – Lalanne, 2018), která mohou mít vliv i na sekvenční strukturu rozhovoru. Videointerakcím byla dosud věnována pozornost především v prostředí výuky a vzdělávání (Gibson, 2014; van Braak et al., 2021; Melander Bowden – Svahn, 2020), zdravotnictví (Seuren et al., 2020; Hansen, 2020) nebo byznysu (Oittinen, 2020). Předložený příspěvek vycházející z etnometodologické konverzační analýzy (Sacks, 1992; Clift, 2016) se zabývá diskusemi po online přednáškách pro veřejnost. Analyzovány jsou popularizační přednášky lingvistů o češtině pořádané Českými centry, jejichž záznamy jsou zveřejňovány na serveru YouTube. Publikum přednášek tvoří zpravidla bohemisté a studenti češtiny z různých zemí i laičtí zájemci o jazyková témata. Analýza se zaměřuje zejména na typy jednání, která účastníci diskusí ve svých replikách realizují, a na sekvenční organizaci těchto interakcí (Schegloff, 2007). Pozornost je přitom věnována tomu, jak se mluvčí v prostředí videokomunikace vztahují k předchozí replice komunikačního partnera, a tak interakčně uskutečňují intersubjektivitu. Z hlediska různých typů jednání byly např. v replikách posluchačů přednášky (během diskuse v roli „tazatelů“) předběžně identifikovány otázky (žádosti o informace), návrhy/podněty, kritické připomínky, vyjádření vlastního stanoviska k diskutovanému problému, poznámky/doplnění, dále různé kombinace těchto typů jednání aj. Výsledky analýzy budou srovnány s popisy jednání v podobných typech interakcí v offline prostředí, např. v diskusích po konferenčních prezentacích (Webber, 2002; Konzett, 2012). Cílem příspěvku je odpovědět na otázku, zda a do jaké míry je sekvenční struktura zkoumaných diskusí ovlivněna videokomunikačním prostředím, tj. nakolik se jejich průběh liší od struktury analogických diskusí realizovaných ve fyzické spolupřítomnosti účastníků.

## Bibliografie

- CLIFT, R. (2016): *Conversation Analysis*. Cambridge: Cambridge University Press.
- GIBSON, W. J. (2014): Sequential order in multimodal discourse: Talk and text in online educational interaction. *Discourse & Communication*, 8, s. 63–83.
- HANSEN, J. P. B. (2020): Invisible participants in a visual ecology: Visual space as a resource for organising video-mediated interpreting in hospital encounters. *Social Interaction: Video-Based Studies of Human Sociality*, 3 (3). <https://doi.org/10.7146/si.v3i3.122609>
- KONZETT, C. (2012): *Any questions? Identity construction in academic conference discussions*. Boston: De Gruyter Mouton.
- MELANDER BOWDEN, H. – SVAHN, J. (2020): Collaborative work on an online platform in video-mediated homework support. *Social Interaction: Video-Based Studies of Human Sociality*, 3 (3). <https://doi.org/10.7146/si.v3i3.122600>
- MLYNÁŘ, J. – GONZÁLEZ-MARTÍNEZ, E. – LALANNE, D. (2018): Situated organization of video-mediated interaction: A review of ethnomethodological and conversation analytic studies. *Interacting with Computers*, 30, s. 73–84.
- OITTINEN, T. (2020): Noticing-prefaced recoveries of the interactional space in a video-mediated business meeting. *Social Interaction: Video-Based Studies of Human Sociality*, 3 (3). <https://doi.org/10.7146/si.v3i3.122781>
- SACKS, H. (1992): *Lectures on Conversation*. Oxford: Blackwell.
- SEUREN, L. M. – WHERTON, J. – GREENHALGH, T. – CAMERON, D. – SHAW, S. E. (2020): Physical examinations via video for patients with heart failure: Qualitative study using conversation analysis. *Journal of Medical Internet Research*, 22, e16694.
- SCHEGLOFF, E. A. (2007): *Sequence Organization in Interaction: A Primer in Conversation Analysis, 1*. Cambridge: Cambridge University Press.
- VAN BRAAK, M. – HUISKES, M. – SCHAEPKENS, S. – VEEN, M. (2021): Shall we all unmute? A conversation analysis of participation in online reflection sessions for general practitioners in training. *Languages*, 6, 72. <https://doi.org/10.3390/languages6020072>
- WEBBER, P. (2002): The paper is now open for discussion. In: E. Ventola – C. Shalom – S. Thompson (eds.), *The language of conferencing*. Frankfurt am Main: Peter Lang, s. 227–254.

### Procedural and substantive clarifications in medical interaction with an AI-based platform

Jakub Mlynář (HES-SO Valais-Wallis University of Applied Sciences Western Switzerland, Sierre)

This paper is grounded in a detailed examination of linguistic interaction with and around a device based on “artificial intelligence” (AI). Scholars studying the relationship between AI, language and society may find themselves in a paradoxical tension (Jaton & Sormani 2023). On the one hand, the notion itself has been subject to a conceptual critique, exploring the dualities of “humans” and “machines”, or the status of “intelligence” and “mind” (e.g., Button et al. 1995; Coulter 2008). On the other hand, AI has already become a social object despite these conceptual issues (Suchman 2023; Collins 2018), and it continues to encroach on our daily lives (Hirsch-Kreinsen 2023; Pflanzner et al. 2023; Mlynář et al. 2024a).

AI-powered tools are also inherent to “radiomics”: the new field of medical imaging analysis based on extracting large-scale quantitative features by machine-learning algorithms (see Guiot et al. 2021). While there have been notable improvements in radiomics over the past decade, little is known about the physicians’ and researchers’ actual conduct while assessing and interpreting radiomic models (Verma et al. 2023). On this background, we focus on semi-experimental trial sessions organized at a university hospital in Switzerland, where novice users (medical technicians and students of medicine) worked with the software platform QuantImage (henceforth QI, see Abler et al. 2023). In 13 one-hour sessions conducted in French, video recordings from multiple angles were obtained to examine interactional details of participants’ collaborative work, following an ethnomethodological and conversation analytic approach (Garfinkel 2002; Sacks 1992; vom Lehn 2018). They include instructions given by the tutor – an expert in radiomics and one of the creators of QI –, in which he explained the main aspects of radiomics’ workflow, the purpose of QI as part of it, and the basic principles of its use (see Figure 1 below; Mlynář et al. 2024b). Responding to the appeal that “AI-facilitated health care requires education of clinicians” (Keane & Topol 2021), the paper is based on an analysis of these explanatory and instructive sequences of interaction. Investigating the organization of “lecturing’s work” (Burns 2012) in this context can offer insights valuable not only for medical education, but also for teaching with AI more generally, and contributes to our understanding of the organization of talk-in-interaction in technologized professional settings (see also Schegloff 1980).

Our findings suggest that, in collaborative work with QI, the presence of the tutor is routinely used as a resource to overcome encountered troubles through sequences of talk that achieve what could be called *substantive* and *procedural* clarification. Substantive clarification addresses more general issues of the domain of expertise, while procedural clarification is related to QI more specifically. These two levels intertwine in the local environment of the trial sessions: QI is a proxy and a synecdoche of radiomics, in the sense that radiomics as a larger domain of medical work is embodied and represented in QI itself. The platform is treated as a representative instance of radiomics, and learning how to use QI is tied to learning about radiomics more generally. These occasions for practical explanations of the operations of the machine are related to current discussions about AI “explainability” in medicine and to questions of machinic agency, as particular actions are ascribed to the platform by the users as well as the tutor. I will argue that the two types of clarification can be seen as context-free and context-sensitive. They are potentially observable in other settings, but they are also specifically fitted to, and constitutive of, the work with QI in the medical setting. In the sessions with the platform, procedural and substantive clarifications are closely tied to the practical task at hand. This task is not primarily accomplished through talking but rather through interacting with the machine in a competent way. What is being clarified in the respective sequences is thus not “the sense of a word”, but “an adequate next action”, and its “meaning” in the larger workflow of radiomics as a specific domain of medical activity. Initiating a proper clarification and responding to it in a locally competent way is an inherent part of the skilled work (even of novices) in a setting that is recognizable as a radiomics setting.



Figure 1. The tutor (in dark t-shirt) instructs two novice users about the use of an AI-based platform for radiomics in one of the trial sessions at a university hospital.

## Bibliography

- Abler, D., Schaer, R., Oreiller, V., Verma, H., Reichenbach, J., Aidonopoulos, O., Evéquo, F., Jreige, M., Prior, J. O., & Depeursinge, A. (2023). QuantImage v2: A comprehensive and integrated physician-centered cloud platform for radiomics and machine learning research. *European Radiology Experimental*, 7(16).
- Burns, S. L. (2012). “Lecturing’s Work”: A collaborative study with Harold Garfinkel. *Human Studies*, 35(2), 175–192.
- Button, G., Coulter, J., Lee, J. R. E., & Sharrock, W. (1995). *Computers, Minds, and Conduct*. Polity.
- Collins, H. (2018). *Artificial Intelligence: Against Humanity’s Surrender to Computers*. Polity.
- Coulter, J. (2008). Twenty-five theses against Cognitivism. *Theory, Culture and Society*, 25(2), 19–32.
- Garfinkel, H. (2002). *Ethnomethodology’s Program: Working Out Durkheim’s Aphorism* (edited by A. W. Rawls). Rowman & Littlefield.
- Guiot, J., Vaidyanathan, A., Deprez, L., Zerka, F., Danthine, D., Frix, A.-N., Lambin, P., Bottari, F., Tsoutzidis, N., Miraglio, B. et al. (2022). A review in radiomics: Making personalized medicine a reality via routine imaging. *Medicinal Research Reviews*, 42(1), 426–440.
- Hirsch-Kreinsen, H. (2023). Artificial intelligence: A “promising technology”. *AI & Society* [online first].
- Jaton, F., & Sormani, P. (2023). Enabling “AI”? The situated production of commensurabilities. *Social Studies of Science*, 53(5), 625–634.
- Keane, P. A., & Topol, E. J. (2021). AI-facilitated health care requires education of clinicians. *The Lancet*, 397(10281): P1254.
- Mlynář, J., de Rijk, L., Liesenfeld, A., Stommel, W., & Albert, S. (2024a). AI in situated action: a scoping review of ethnomethodological and conversation analytic studies. *AI & Society* (online first).
- Mlynář, J., Depeursinge, A., Prior, J. O., Schaer, R., Martroye de Joly, A., & Evéquo, F. (2024b). Making sense of radiomics: insights on human–AI collaboration in medical interaction from an observational user study. *Frontiers in Communication*, 8, 1234987.
- Pflanzer, M., Dubljević, V., Bauer, W.A., Orcutt, D., List, G., & Singh, M. P. (2023). Embedding AI in society: Ethics, policy, governance, and impacts. *AI & Society*, 38, 1267–1271.
- Sacks, H. (1992). *Lectures on Conversation*. Blackwell.
- Schegloff, E. A. (1980). What type of interaction is it to be? In *ACL’80: Proceedings of the 18th annual meeting on Association for Computational Linguistics*, Stroudsburg, pp. 81–82.
- Suchman, L. (2023). The uncontroversial “thingness” of AI. *Big Data & Society*, 10(2).
- Verma, H., Mlynář, J., Schaer, R., Reichenbach, J., Jreige, M., Prior, J. O., Evéquo, F., & Depeursinge, A. (2023). Rethinking the role of AI with physicians in oncology: Revealing perspectives from clinical and research workflows. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI ’23)*, Article 17. ACM.
- vom Lehn, D. (2018). *Ethnomethodologische Interaktionsanalyse: Videodaten analysieren und die Organisation von Handlungen darstellen*. Beltz Juventa.

**Assessing what can be heard and what can be seen:  
The multimodal construction of assessments in smartphone-based showing sequences**

Florence Oloff (Leibniz-Institute for the German Language, Mannheim)

This contribution will look at the multimodal construction of assessments in face-to-face conversations, with a special focus on mundane digital devices and their affordances for joint action. While mobile devices such as smartphones might traditionally be thought of as objects that are designed for individual use and that potentially impede meaningful social interaction (cf. Avgustis 2023; Mantere/Raudaskoski 2017), interactionally inspired research emphasizes their potential for joint social action (DiDomenico et al. 2020; Porcheron et al. 2016; Suderland 2020; Weilenmann/Larsson 2001, *inter alia*).

In smartphone-based showing activities, for instance, participants jointly look at and comment on digital content on screen. Previous research has underlined how mobile devices can provide “epistemic support” for formulating assessments during such showings, as the sharing of the digital content provides direct access to the referents in question (Raclaw et al. 2016). Through providing access to a “social archive” (Keppler 2019), composed of private messages or photo albums, and publicly accessible digital objects on webpages and social media platforms (see Suderland 2020), smartphones can provide access to numerous types of digital *assessables* (e.g., visible and/or audible, dynamic or static, private or public objects). Moreover, smartphones, as physical devices, can themselves be shared in different ways with co-present others (e.g., by reading from the screen, tilting the display toward the recipient, or handing over the device, see Oloff 2021; Porcheron et al. 2016; Raclaw et al. 2016). These various dimensions make smartphone-based showings a mundane yet highly interesting setting for investigating assessments. In conversation analytic literature, assessing turns traditionally refer to storytellings of past events (Goodwin/Goodwin 1987; Pomerantz 1984) rather than to “here-and-now” events or objects (cf. Lindström/Mondada 2009). Although more recently, research on multisensoriality and food tasting activities has introduced a new dimension to the study of assessing objects “in situ” (e.g., Mortensen/Wagner 2019; Mondada 2019, 2021), up to now there is no research on how we multimodally assess digital objects on smartphones, and how their various affordances (Hutchby 2001) come into play when doing so.

This presentation will thus take a closer look at how assessments within smartphone-based showing sequences are multimodally organized. Based on examples from everyday conversations in Czech and German, I will present several excerpts in which assessments play a central role. In addition to video recordings of the overall setting, individual action cameras have been used for capturing the participants’ mobile device screens, which allows to relate speaking turns to precise digital objects on screen. Using a sequential and multimodal approach to social interaction (multimodal interaction analysis, e.g., Deppermann/Streeck 2018; Mondada 2018; Streeck et al. 2011), the analysis will explore how participants select and offer specific digital objects on screen for assessment, and how their co-participants respond to these. A small collection of cases will show that offering assessments of digital objects varies according to their nature: while *audible* assessables (such as music) are jointly oriented to through syntactically co-constructed turns and the establishment of mutual gaze, *visible* assessables will be typically pointed at, the accompanying turns being often minimal and choral in nature (Lerner 2002). This shows that participants adjust the multimodal turn formats to the nature of the digital object in order to produce simultaneous and timely fitted assessments that clearly differ from assessments of past events (Goodwin/Goodwin 1987). More generally, this analysis illustrates how mobile devices can be used as a resource for joint social action, allowing for reflections on how language and the material environment are intertwined.



## Bibliography

- Avgustis, I. (2023). Respecifying Phubbing: Video-Based Analysis of Smartphone Use in Co-Present Interactions. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–15. <https://doi.org/10.1145/3544548.3581052>
- Deppermann, A., & Streeck, J. (Eds.). (2018). *Time in Embodied Interaction. Synchronicity and sequentiality of multimodal resources*. John Benjamins.
- DiDomenico, S. M., Raclaw, J., & Robles, J. S. (2020). Attending to the Mobile Text Summons: Managing Multiple Communicative Activities Across Physically Copresent and Technologically Mediated Interpersonal Interactions. *Communication Research*, 47(5), 669-700.
- Goodwin, C., & Goodwin, M. H. (1987). Concurrent Operations on Talk: Notes on the Interactive Organization of Assessments. *IPrA Papers in Pragmatics*, 1, 1–54.
- Hutchby, I. (2001). Technologies, Texts and Affordances. *Sociology*, 35(2), 441–456.
- Keppler, A. (2019). “Zeig mal”: Smartphones im Gespräch. In K. Marx & A. Schmidt (Eds.), *Interaktion und Medien. Interaktionsanalytische Zugänge zu medienvermittelter Kommunikation* (pp. 177–190). Winter Verlag.
- Lerner, G. H. (2002). Turn-Sharing. The Coral Co-Production of Talk-in-Interaction. In C. E. Ford, B. A. Fox, & S. A. Thompson (Eds.), *The Language of Turn and Sequence* (pp. 225–256). Oxford University Press.
- Lindström, A., & Mondada, L. (2009). Assessments in Social Interaction: Introduction to the Special Issue. *Research on Language and Social Interaction*, 42(4), 299–308.
- Mantere, E., & Raudaskoski, S. (2017). The sticky media device. In A. R. Lahikainen, T. Mälkiä, & K. Repo (Eds.), *Media, Family Interaction and the Digitalization of Childhood* (pp. 135–154). Edward Elgar.
- Mondada, L. (2018). Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality. *Research on Language and Social Interaction*, 51(1), 85–106.
- Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics*, 145, 47–62.
- Mondada, L. (2021). *Sensing in Social Interaction: The Taste for Cheese in Gourmet Shops* (1st ed.). Cambridge University Press.
- Mortensen, K., & Wagner, J. (2019). Inspection sequences – multisensorial inspections of unfamiliar objects. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 20, 399-343. <http://www.gespraechsforschung-online.de/fileadmin/dateien/heft2019/si-mortensen.pdf>
- Oloff, F. (2021). Show imperatives in smartphone-based showing sequences in Czech and German. *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 22, 691-724. <http://www.gespraechsforschung-online.de/fileadmin/dateien/heft2021/si-oloff.pdf>
- Pomerantz, A. (1984). Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In J. M. Atkinson & J. Heritage (Eds.), *Structures of Social Action* (pp. 57–101). Cambridge University Press.
- Porcheron, M., Fischer, J. E., & Sharples, S. (2016). Using Mobile Phones in Pub Talk. *CSWC' 16 - Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1649–1661. <https://doi.org/10.1145/2818048.2820014>
- Raclaw, J., Robles, J. S., & DiDomenico, S. M. (2016). Providing Epistemic Support for Assessments Through Mobile-Supported Sharing Activities. *Research on Language and Social Interaction*, 49(4), 362–379.
- Streeck, J., Goodwin, C., & LeBaron, C. (Eds.). (2011). *Embodied Interaction. Language and Body in the Material World*. Cambridge University Press.
- Suderland, D. (2020). „oh isch FIND\_s nich;“ Eine konversationsanalytische Untersuchung sprachlicher Bezugnahmen auf smartphone-gestützte Suchanfragen in Alltagsgesprächen. *Journal Für Medienlinguistik*, 2(2), 90–122. <https://doi.org/10.21248/jfml.2019.17>
- Weilenmann, A., & Larsson, C. (2001). Local use and sharing of mobile phones. In B. Brown & N. Green (Eds.), *Wireless world: Social and interactional aspects of the mobile age* (pp. 92–107). Springer Verlag.

### Jak studenti při ústních zkouškách z angličtiny zařazují výrazy z tištěných materiálů do probíhající interakce

David Ryška (Masaryk University), František Tůma (Vienna University of Economics and Business)

Standardizované zkoušky z jazyka skládají kandidáti z řad maturantů, vysokoškolských studentů i žadatelů o trvalý pobyt. Ačkoli jde zpravidla o zkoušky vysoké důležitosti a přestože se jazykovým testováním zabývá značné množství odborné literatury (např. Brunfaut, 2023; Fulcher & Harding, 2021), získala problematika interakce během ústní zkoušky výzkumnou pozornost teprve relativně nedávno v souvislosti s rozvíjejícím se zájmem o tzv. cizojazyčnou interakční kompetenci (Pekarek Doehler, 2018). Z pohledu tvorby a administrace testových úloh přitom interakční pohled hraje důležitou roli ve vztahu ke způsobu, jakým zkouška probíhá a jakým jsou hodnoceny jednotlivé výkony, neboť souvisí s nutností zajistit kandidátům podmínky vedoucí ke srovnatelným a zároveň diferencovaným výkonům. Tyto požadavky představují značnou výzvu zejména pro ty části zkoušky, při kterých mají kandidáti interagovat společně, jelikož výpovědní hodnota úkolu se v dané situaci odvíjí od způsobů, jakými kandidáti vyjednávají téma ve vzájemné interakci a (ne)dávají jeden druhému prostor k reakci (srov. Galaczi, 2008; Gan et al., 2009). Náš příspěvek se zaměřuje právě na tuto výzvu tím, že analyzuje interakce kandidátů během diskusního úkolu, při kterém dostanou pracovní list s tematicky souvisejícími položkami a standardizované pokyny.

Data pro náš výzkum tvoří videonahrávky 66 párových a skupinových interakcí během standardizovaných zkoušek z angličtiny, které proběhly na jedné univerzitě v Česku. Nahrávky byly pořízeny dvěma kamerami, přičemž jedna zachycuje kandidáty a druhá se zaměřuje na zkoušejícího. Jazykovou úroveň zkoušky lze označit za vyšší pokročilou, neboť byly použity standardizované materiály pro zkoušku *Cambridge First* (dříve *First Certificate in English*, jedná se o úroveň B2 dle SERRJ, viz Council of Europe, 2020). Naše analýza se zaměřuje na část zkoušky, během které kandidáti dostanou pracovní list s centrálním tématem a několika souvisejícími aspekty daného tématu. V souladu se standardizovanými pokyny mají kandidáti následně dvě minuty o tématu diskutovat. Nahrávky byly podrobně přepsány a analyzovány za použití principů multimodální konverzační analýzy (Goodwin, 2018) s cílem podrobně popsat, jak kandidáti během zkoušky používají pracovní list: jak se k němu vztahují (pohledem, ukazováním) a jak zařazují položky z pracovního listu do rozhovoru.

V příspěvku se konkrétně zaměříme na sekvence, v nichž jeden kandidát položí druhému otázku vztahující se k některé z položek na pracovním listu a druhý kandidát na tuto otázku reaguje. Předběžná analýza ukazuje, že kandidáti se při iniciaci nebo odpovědi k materiálu pozorovatelně vztahují (nahlíží do něj, ukazují) a v některých interakcích organizují diskusi jako sérii otázek a odpovědí, jež se pracovním listem řídí. Položky z něj přitom kandidáti do různé míry gramaticky integrují do svých otázek a odpovědí. Uvedená zjištění odkrývají způsoby, jimiž se položky z pracovního listu odráží v interakcích kandidátů společně s jejich různě rozvinutými interakčními repertoáry (Hall, 2018), nezbytnými pro kladení otázek a odpovídání na ně. Tato zjištění jsou zásadní pro další uvažování o způsobu hodnocení cizojazyčné interakční kompetence při ústních zkouškách: kandidáti dostávají shodný pracovní list a stejné pokyny od zkoušejícího, následně však používají různé jazykové a tělesné prostředky k tomu, aby o zadaném tématu vedli hovor. Studie tak přispívá k poznání v oblasti jazykového testování a cizojazyčné interakční kompetence (Galaczi & Taylor, 2018; Malabarba & Betz, 2023).

## Bibliografie

- Brunfaut, T. (2023). Future challenges and opportunities in language testing and assessment: Basic questions and principles at the forefront. *Language Testing*, 40(1), 15–23.  
<https://doi.org/10.1177/02655322221127896>
- Council of Europe. (2020). *Common european framework of reference for languages: Learning, teaching, assessment – companion volume*. Council of Europe Publishing.
- Fulcher, G., & Harding, L. (Ed.). (2021). *The Routledge handbook of language testing* (2. vyd.). Routledge.
- Galaczi, E. (2008). Peer-peer interaction in a speaking test: The case of the First Certificate in English Examination. *Language Assessment Quarterly*, 5(2), 89–119.  
<https://doi.org/10.1080/15434300801934702>
- Galaczi, E., & Taylor, L. (2018). Interactional competence: Conceptualisations, operationalisations, and outstanding questions. *Language Assessment Quarterly*, 15(3), 219–236.  
<https://doi.org/10.1080/15434303.2018.1453816>
- Gan, Z., Davison, C., & Hamp-Lyons, L. (2009). Topic negotiation in peer group oral assessment situations: A conversation analytic approach. *Applied Linguistics*, 30(3), 315–334.  
<https://doi.org/10.1093/applin/amn035>
- Goodwin, C. (2018). *Co-operative action*. Cambridge University Press.
- Hall, J. K. (2018). From L2 interactional competence to L2 interactional repertoires: Reconceptualising the objects of L2 learning. *Classroom Discourse*, 9(1), 25–39.  
<https://doi.org/10.1080/19463014.2018.1433050>
- Malabarba, T., & Betz, E. (2023). Describing and assessing interactional competence in a second language: Introduction to the special issue. *Applied Pragmatics*, 5(2), 121–141.  
<https://doi.org/10.1075/ap.00020.mal>
- Pekarek Doehler, S. (2018). Elaborations on L2 interactional competence: The development of L2 grammar-for-interaction. *Classroom Discourse*, 9(1), 3–24.  
<https://doi.org/10.1080/19463014.2018.1437759>

## Mediální dialog v podmínkách autoritářského režimu

Marián Sloboda (Univerzita Karlova)

V běžném pojetí je dialog tematicky zaměřenou výměnou replik mezi dvěma nebo více osobami v mluvené komunikaci (srov. Hoffmannová 2016). Na takový typ komunikace se tradičně zaměřuje konverzační analýza – proud sociologie uplatňovaný i v lingvistice, která se zabývá strukturací rozhovorů a interakčními efekty této strukturace. V oblasti mediální komunikace se tedy tato perspektiva uplatňuje obvykle na dialogy typu zpravodajského interview a politické debaty (např. Clayman & Heritage 2002, Hutchby 2006). V nich médium zachycuje střídání replik postav (reportéra a dotazovaného, resp. moderátora a diskutantů) tváří in tvář. Ovšem i repliky pronesené sociálními aktéry v různém čase a na různých místech se dostávají do dialogu, když jsou médiem (např. autorem zprávy, moderátorem apod.) „dány dohromady“. Vícenásobným dialogickým spojením tohoto typu vzniká to, co Leudar a Nekvapil (2022) nazývají „mediální dialogická síť“. Krátký příklad níže ukazuje situační zesíťování více či méně komplexních výpovědí tří aktérů v jedné novinové zprávě (výpovědi, na něž zpráva odkazuje, jsou v úryvku podtrženy a doplněny číslovkou označující jejich chronologické pořadí):

„Stop Sudeťákům“, *Svobodné slovo*, 13. 1. 1993

*PRAHA (fra) – Předseda Československé strany socialistické L. Dvořák [3.] přivítal [2.] prohlášení českého premiéra V. Klause, že vláda nehodlá jednat o dalších restitučních [1.] nárocích sudetských Němců...* (cit. podle Nekvapila & Leudara 2002, s. 484)

Ve zpravodajství televizním je žánrovou normou pořídit „hlavy“ – záběry na obličeje osob, které vyjadřují odlišné názory formulované v různém čase a případně i na různých místech, ale pak v rámci zprávy prezentované dialogicky (Kaderka & Havlík 2010). Leudar a Nekvapil (2004, s. 264) vyslovují názor, že mediální dialogické sítě jsou charakteristické pro společnosti, v nichž jsou veřejná debata a participace důležitou součástí politické kultury, tedy v demokratických zemích.

Jak vypadá prezentace odlišných politických názorů v autoritářském státě, ve kterém vláda kontroluje a cenzuruje média s cílem opoziční názory potlačit? V létě 2005 jsem zaznamenal na hlavních kanálech běloruské televize vyhocenou konfliktní rétoriku, a začal jsem proto zpravodajské relace nahrávat. Postupně tak vznikl materiál, na kterém se zakládá tento příspěvek. Jeho analýza ukázala, že zpravodajské interview reportéra s respondentem je v těchto relacích vzácné a že repliky různých postav, včetně „hlav“, jsou navzájem neprotiřevivé, argumentačně se naopak se podporují. Zaznamenaný konflikt se naopak odehrával v rámci repliky moderátora, případně hlasového doprovodu (voice-overu). Tento jev lépe odpovídá pojetí dialogu užívanému v okruhu M. M. Bachtina (např. Bachtin 1971, Vološinov 1986). Podle něj se nejen mezi výpověďmi dvou osob, ale i přímo v rámci jedné výpovědi střetávají protichůdné představy, odlišná sociální hodnocení, které do výpovědi prolínají z komunikační historie společnosti. Moderátoři a autoři voice-overů, tj. pracovníci televize, v analyzovaných zprávách s pomocí „hlav“ obhajovali obraz Běloruska jako klidné a stabilní země žijící v multietnickém smíru a vymezovali se polemicky proti narušení tohoto obrazu. To kdo, kdy a jak konkrétně tento obraz narušoval, tj. iniciační repliky, na které moderátor či voice-over rozsáhle reagovali, bylo prezentováno neurčitě a někdy i velmi nezřetelně.

Otázkou je, proč se takto potlačené opoziční hlasy přece jen ve státem kontrolovaných médiích objevily. Vysvětlení vidím v tom, že přes různé snahy o státní kontrolu cirkulace informací byly běloruskému obyvatelstvu dostupné i jiné informační zdroje, například vysílání zahraničních televizí, internet nebo telefonáty příbuzných ze zahraničí, z nichž se obyvatelé mohli dovědět, že koncem července 2005 speciální složky státní policie provedly razii v organizaci polské národnostní menšiny Svazu Poláků v Bělorusku, tehdy největší nevládní organizaci v zemi. Tvůrci mediálního obsahu v autoritářském režimu tedy nemohli disentanční hlasy vyloučit zcela, zařadili je, ale snažili se je ovládnout výše uvedenými diskurzivními formami.

## Bibliografie

- Bachtin, M. M. (1971). *Dostojevskij umělec: K poetice prózy*. Československý spisovatel. (Orig. *Проблемы поэтики Достоевского*, 1963.)
- Clayman, S., & Heritage, J. (2002). *The News Interview: Journalists and Public Figures on the Air*. Cambridge University Press.
- Hoffmannová, J. (2016). Dialog. In P. Karlík, M. Nekula & J. Pleskalová (Eds), *Nový encyklopedický slovník češtiny a–m* (s. 335–337). Nakladatelství Lidové noviny.
- Hutchby, I. (2006). *Media Talk: Conversation Analysis and the Study of Broadcasting*. Open University Press.
- Kaderka, P. & Havlík, M. (2010). Vytváření televizních zpráv: pracovní postupy v systému žánrových norem. *Sociologický časopis = Czech Sociological Review* 46(4), 537–567.
- Leudar, I. & Nekvapil, J. (2004). Media dialogical networks and political argumentation. *Journal of Language and Politics* 3(2), 247–266.
- Leudar, I. & Nekvapil, J. (2022). *Dialogical Networks: Using the Past in Contemporary Research*. Routledge.
- Nekvapil, J. & Leudar, I. (2002). Sekvenční struktury v mediálních dialogických sítích. *Sociologický časopis = Czech Sociological Review* 38(4), 483–500.
- Vološinov, V. N. (1986). Marxizmus a filozofia jazyka. In M. M. Bachtin & V. N. Vološinov, *Marxizmus, freudizmus, filozofia jazyka* (s. 171–382). Pravda. (Orig. *Марксизм и философия языка*, 1929.)

**The Effect of Chosen Turn-Taking Strategies on the Perception of the Politician's Public Image**

Alexey Tymbay (Technical University in Liberec)

**Background** The public image of a politician is a multi-level structure incorporating the features of the objective world, such as their age, gender, and accent [3]; elements of cognition and preferred ideology objectified in discourse with the help of conceptual thinking [1]; and features of the speaker's self-expression realized through their speech strategies [2]. Therefore, the development of a speaker's public image can be viewed from the perspective of newly accepted conversational patterns, speech maneuvering, turn-taking models, and methods of achieving a communicative goal. Van Dijk [8] calls such choices the situational "performances of identity" that can significantly modify the speaker's image or contribute to constructing a completely new one.

**Objectives** The research aim of our study is to analyze the strategies of cooperative vs. competitive turn transitions [6] chosen by the participants of televised 2016, 2020, and 2024 US presidential debates and assess their influence on the perception of the public image of American politicians, namely Hillary Clinton, Donald Trump, and Joseph Biden. The rationale for choosing turn-taking for analysis was prompted by the fact that immediately after the 2020 debates, political experts and analysts made multiple claims that excessive interruptions affected the public perception of the politicians [5].

**Contribution** In the present study, we analyze speakers' turn-taking strategies in the context of election debates since this genre of political communication explicitly reflects the times, cultural shifts, and political preferences of the general population of a country [4]. We hypothesize that in order to fit into this paradigm, politicians and their election managers modify certain aspects of their speaking styles and demonstrate relative flexibility in the use of communicative tactics to suit the specific tastes of the electorate [7].

**Method** In our study, we used 538's and Ipsos's opinion poll results to assess the potential effect of debate performances on the politicians' public perception, although realizing that isolating turn-taking from other speech behaviors is quite problematic. We also provide details of specific practices of overlaps and interruptions undertaken by the speakers through which their communicative identities are proposed, adjusted, and managed. By classifying these micro-moments of identity construction, we attempt to pinpoint specific communicative occasions where particular identity issues become salient to the TV viewers of the debates.

## Bibliography

- [1] Antaki, Ch., Condor, S., and Levine M. 1996. Social identities in talk: Speakers' own orientations. *British Journal of Social Psychology* 35(4), 473–492.
- [2] Bucholtz, M., and Hall, K. 2005. Identity and interaction: a socio-cultural linguistic approach. *Discourse Studies* 7(4-5), 585–614.
- [3] Creel, S., and Bregman, M. 2011. How talker identity relates to language processing. *Language and Linguistics Compass* 5, 190–204.
- [4] Ilie, C. 2015. Parliamentary discourse. In Tracy, K. (Ed.) *The International Encyclopedia of Language and Social Interaction*. New York, NY: Jhon Wiley and Sons
- [5] Macaya, M., Rocha, V., Blaine, K., and Estepa J. 2020. First 2020 presidential debate. *CNN Politics*. Available at: <https://edition.cnn.com/politics/live-news/presidential-debate-coverage-fact-check-09-29-20/index.html>
- [6] Oertel, C., Włodarczak, M., Tarasov, A., Campbell, N., and Wagner P. 2012. Context cues for classification of competitive and collaborative overlaps. *Proceedings of the 6th International Conference on Speech Prosody*. Available at: [https://www.isca-speech.org/archive\\_v0/sp2012/papers/sp12\\_721.pdf](https://www.isca-speech.org/archive_v0/sp2012/papers/sp12_721.pdf)
- [7] Tymbay, A. 2023. Effectiveness of interruption as a communicative strategy in the 2020 presidential debates in the USA. *Brno Studies in English*, 48(2), 101-118.
- [8] Van Dijk, T. 2010. Political identities in parliamentary debates. In Ilie, C. (Ed.) *European parliaments under scrutiny. Discourse strategies and interaction practices*. Amsterdam: John Benjamins, 29-56.

9

*Frazeologie češtiny:  
strukturní, kognitivní,  
kulturní a didaktické aspekty*

organizují

Marie Kopřivová

Kateřina Šichová

(CZE)



# Popis panelu

Vývoj bádání v současné evropské frazeologii je ovlivněn rozvojem řady dalších disciplín, které jsou s ní v kontaktu a přináší nové možnosti zkoumání (srov. např. Burger et al. 2007, Mieder 2009 a novější publikační výstupy z konferencí EUROPHRAS). Jedná se především o aktuální dostupnost, dosud nevidaný rozsah dat a technické možnosti pro jejich zpracovávání, replikovatelnost výzkumu, nové teoretické přístupy a další. Na tomto pozadí je cílem tohoto panelu představit aktuální výzkum frazeologie češtiny z různých perspektiv.

Česká frazeologie disponuje (i v kontextu jiných jazyků) nejen velmi dobrým teoretickým popisem (Čermák 2007), ale i kvalitním frazeografickým zpracováním (SČFI 2009). Na ně navazuje ve světovém měřítku ojedinělý projekt anotace frazémů v rozsáhlých korpusových datech (Petkevič a kol. 2020).

Se změnami tradičního pojetí jednojazyčných a překladových slovníků a posunem potřeb jejich uživatelů souvisí i proměny frazeografie a paremiografie. V rámci evropských výzkumů jsou budovány též vícejazyčné databáze, zaměřené např. na přísloví (*SprichWort*, Tölgyesi 2021), nebo tzv. frazeologická a paremiologická optima různých jazyků (Kopřivová 2017).

Na poli didaktiky se lze opřít o propracovanou metodologii pro výuku frazeologie na základních a středních školách (Čechová 2012). K dosud málo zpracovaným oblastem ale patří frazeodidaktika a paremioididaktika češtiny jako cizího jazyka (Šemelík/Šichová 2016, 2017) a češtiny jako druhého a zděděného jazyka, jakož i frazeologická kompetence mluvčích češtiny (jako mateřského i cílového jazyka) obecně a její testování.

Frazémy si nacházejí cestu i do nových mediálních prostředí. Jedná se především o audiovizuální prostředky, které vytváří podmínky pro vznik a šíření tzv. memů. Frazeologický výzkum se tak dále zaměřuje nejen na rozsah znalostí stávajících parémii a jiných frazémů, ale také na to, jak se frazeologie vyvíjí a které nové typy frazémů (Janovec et al. 2013) si mladí mluvčí osvojují pod vlivem angličtiny. Spolu s tím se dostává do popředí také otázka funkce frazému v různých typech textů.

Prostor pro další výzkumy skýtá též oblast kontrastivní frazeologie včetně otázek překladu frazeologických jednotek (Krejčí 2015, Obstová 2022). Řada badatelů analyzuje frazeologický materiál z kognitivního pohledu (srov. Dobrovoľskij/Piirainen 2009) a přináší nové poznatky k sémantickým specifikům frazeologických jednotek (Christou/Táborská 2022). Další teoretickou disciplínou, která přináší do frazeologie nové podněty a možnosti v popisu jednotek stojících na pomezí frazeologie a pravidelného jazyka, je konstrukční gramatika (srov. Ziem 2018, Lehečková 2017). Neustávají konečně ani výzkumy etymologické, diachronní a nářeční (Kloferová 2022). Autorka nově vydaného *Etymologického slovníku české frazeologie* (Stěpanova 2022) přednese úvodní přednášku panelu.

O bohatosti a potenciálu frazeologického bádání chceme diskutovat spolu s kolegy, kteří se mu věnují na českém materiálu. Pro frazeologický panel proto uvítáme příspěvky k tématům z výše naznačených, ale i dalších oblastí. Vybrané příspěvky z konference plánujeme vydat knižně v roce 2025.

## Literatura

- Burger, Harald/Dobrovolskij, Dmitrij/Kühn, Peter/Norrick, Neal R. (eds.) (2007): *Phraseologie. Ein internationales Handbuch der zeitgenössischen Forschung. Phraseology. An International Handbook of Contemporary Research*. Berlin/New York: De Gruyter.
- Čechová, Marie (2012): *Řeč o řeči*. Praha: Academia, 207–212.
- Čermák, František (2007): *Frazeologie a idiomatika česká a obecná. Czech and General Phraseology*. Praha: Karolinum.
- Dobrovolskij, Dmitrij/Piirainen, Elisabeth (2009): *Zur Theorie der Phraseologie. Kognitive und kulturelle Aspekte*. Tübingen: Stauffenburg.
- EUROPHRAS = *Europäische Gesellschaft für Phraseologie*, web s bibliografií: <http://www.euophras.org/en> [29.9.2023].
- Christou, Anna/Táborská, Karolína (2022): The Cultural Concepts of “ZDRAVÍ” (“HEALTH”) and “NEMOC” (“ILLNESS”) in Czech Phraseology. In: *Ethnolinguistics. Issues in Language and Culture*, 34, 303–318.
- Janovec, Ladislav/Adámková, Petra/Hradilová, Darina/Lábus, Václav/Martinková, Michaela/Nagy, Marek/Stěpanova, Ludmila/Svobodová, Diana (2013): *Dynamika českého lexika a lexikologie*. Olomouc: Univerzita Palackého v Olomouci.
- Kloferová, Stanislava (2021): Kdy žába hýbá uchem (k nářeční frazeologii v češtině). In: Janovec, Ladislav (ed.): *Svět v obrazech a ve frazeologii II / World in Pictures and in Phraseology II*. Praha: PedF UK, 479–488.
- Kopřivová, Marie (2017): Contribution Towards a Corpus-Based Phraseology Minimum. In: Mitkov, Ruslan (ed.): *Computational and Corpus-Based Phraseology*. London/Berlin/Heidelberg: Springer Verlag, 220–231.
- Krejčí, Pavel (2015): *Srbská frazeologie v českém a bulharském překladu: Kontrastivní analýza*. Brno: Masarykova univerzita.
- Lehečková, Eva (2017): Využití korpusových metod v popisu češtiny: příklad kolostrukční analýzy. In: *Časopis pro moderní filologii*, 99/2, 167–180.
- Mieder, Wolfgang (2009): *International Bibliography of Paremiology and Phraseology 1., 2.* Berlin: De Gruyter.
- Obstová, Zora (2022): Fraseologismi nelle traduzioni letterarie: analisi delle strategie traduttive basata sui corpora paralleli. In: De Giovanni, Cosimo (ed.): *Fraseologia e paremiologia tra lingua e discorso*. Roma: Aracne, 419–432.
- Petkevič Vladimír/Kopřivová Marie/Hnátková Milena/Jelínek Tomáš/Kopřiva Pavel/Rosen Alexandr/Skoumalová Hana/Vondříčka Pavel (2020): Typologie víceslovných jednotek v češtině a frekvenční zastoupení jejich hlavních vlastností v žánrově vyváženém korpusu. In: *Studie z aplikované lingvistiky*, 11, 37–62.
- SČFI = Čermák, František et al. (2009): *Slovník české frazeologie a idiomatiky 1.–4.* Praha: Leda.)
- SprichWort = *SprichWort-Plattform – Eine Internet-Lernplattform für das Sprachenlernen*, web: <https://www.sprichwort-plattform.org/> [29.9.2023].
- Stěpanova, Ludmila (2022): *Etymologický slovník české frazeologie*. Olomouc: Univerzita Palackého v Olomouci.
- Šemelík, Martin/Šichová, Kateřina (2016, 2017): Was nicht ist, kann noch werden. Zur Parömioididaktik im Tschechisch-als-Fremdsprache-Unterricht (Teil 1: Eine exemplarische Lehrwerk- und Wörterbuchanalyse, Teil 2: Aufgaben und Übungen). In: *Beiträge zur Fremdsprachenvermittlung*, 58, 59, 61–104, 37–76.
- Tölgyesi, Tamás (2021): Přísluví v jazycích střední Evropy na základě údajů z databáze SprichWort. In: Janovec, Ladislav (ed.): *Svět v obrazech a ve frazeologii II / World in Pictures and in Phraseology II*. Praha: PedF UK, 265–281.
- Ziem, Alexander (2018): Construction Grammar meets Phraseology: eine Standortbestimmung. In: *Linguistik Online*, 90/3.

# Obsah

*Koncepce zpracování frazémů v Akademickém slovníku současné češtiny*

Anna Christou (Karla Tvrdá (Ústav pro jazyk český AV ČR)

*Nestandardní adjektivní přirovnání v korpusových datech*

Jaroslav Emmer (Jihočeská univerzita v Českých Budějovicích)

*Víceslovné lexikální entity korpusově*

Marie Kopřivová<sup>1</sup>, Hana Skoumalová<sup>2</sup>, Hana Goláňová<sup>3</sup>, Milena Hnátková<sup>2</sup>, Anna Christou<sup>4</sup>,  
Tomáš Jelínek<sup>2</sup>, Jan Křivan<sup>4</sup>, Vladimír Petkevič<sup>2</sup>, Pavel Vondříčka<sup>2</sup>, Karla Tvrdá<sup>4</sup> (<sup>1</sup>Ústav  
bohemistiky pro cizince a komunikace neslyšících FF UK, <sup>2</sup>Ústav Českého národního korpusu FF  
UK, <sup>3</sup>Katedra českého jazyka a literatury UHK, <sup>4</sup>Ústav pro jazyk český AV ČR)

*„Jak se do lesa volá, tak se z lesa ozývá“: Frazémy jako metakomunikační vyjádření v  
současných psaných mediálních textech*

Kamila Mrázková (Filozofická fakulta Univerzity Karlovy)

*Všechno souvisí se vším, ale některé věci nejdou dohromady: k frazémům s recipročním  
významem*

Sylva Nzimba, Vojtěch Veselý (Ústav pro jazyk český AV ČR)

*Mezi frazeologií a konstrukční gramatikou: frazémy s opakováním verbálního komponentu*

Zora Obstová (Charles University)

*Jak hloupý je Hloupý Honza? Frazémová analýza jazykového prototypu v českém kontextu a  
jeho porovnání s chorvatštinou a makedonštinou*

Slavomira Ribarova (Filozofski fakultet Sveučilišta u Zagrebu)

*K historii vzniku Etymologického slovníku české frazeologie*

Ludmila Stěpanova

*K frazeologické kompetenci druhé generace mluvčích češtiny jako zděděného jazyka na  
příkladu českých přísloví a hlášek*

Kateřina Šichová (Universität Regensburg)

## Koncepce zpracování frazémů v Akademickém slovníku současné češtiny

Anna Christou, Karla Tvrdá (Ústav pro jazyk český AV ČR)

Cílem příspěvku je představit aktuální zpracování frazémů ve vznikajícím *Akademickém slovníku současné češtiny* (srov. Kochová, P. – Opavská, Z. et al. 2016; Šemelík, M. – Lišková, M. – Křivan, J. 2023). Při zpracování všeobecného výkladového slovníku patří oblast frazeologie k důležitým koncepčním tématům. Zatímco ve starším *Slovníku spisovného jazyka českého* a také ve *Slovníku spisovné češtiny* jsou frazémy uváděny především v rámci heslové stati jednoslovného hesla a definice je ztvárněna pouze stručnou glosou, v ASSČ je zpracování frazeologie výrazně rozšířeno. Větší důraz je kladen na zachycení frazémů jako samostatných lexikálních jednotek a jejich definice je podána výkladovou perifrází. Na druhé straně v ASSČ frazeologie není zpracována natolik podrobně jako v *Slovníku české frazeologie a idiomatiky*, jehož hesla zahrnují mj. gramatickou charakteristiku, informaci o intonačním typu, cizojazyčné ekvivalenty a vedle vlastního výkladu také popis komunikační situace.

Frazémem v ASSČ rozumíme ustálené spojení, jehož celkový význam není plně odvoditelný z významu jednotlivých složek. Zachycovány jsou všechny typy frazémů, tj. přirovnání, jmenné frazémy, slovesné frazémy a větné frazémy včetně paremií. Základním zdrojem pro vyhledávání frazémů je ČNK, především korpusy řady SYN. Lemma je zaznamenáváno v reprezentativním tvaru, tj. zpravidla ve spisovné podobě a se slovesy v infinitivu. V zájmu přehlednosti a vstřícnosti k uživatelům zachycujeme pouze nejfrekventovanější varianty frazémů. Součástí frazeologického hesla je rovněž informace o stylové platnosti (srov. Lišková, M. 2013). Počínaje hesly s náslovím *d-* již frazeologická hesla neobsahují exemplifikaci, avšak aktuálně je připravována funkce, která zobrazí doklady užití daného frazému v korpusu SYN. Tato data z ASSČ a ČNK budou propojena s databází víceslovných lexikálních jednotek LEMUR v rámci projektu TA ČR *Víceslovné jednotky pro digitální vzdělávání* (TQ01000177). Na rozdíl od ASSČ poskytuje tato databáze komplexnější popis frazémů a dalších typů víceslovných lexikálních jednotek, které klasifikuje na základě podrobné typologie zachycující jejich vlastnosti na různých jazykových rovinách (Petkevič et al. 2020, s. 38–39).

Příspěvek se zaměří jak na obecné zásady zpracování frazémů v ASSČ, tak na konkrétní problematické jevy.

## Bibliografie

### Literatura

Kochová, P. – Opavská, Z. et al. (eds.) (2016): *Kapitoly z koncepce Akademického slovníku současné češtiny*. Praha: Ústav pro jazyk český AV ČR, v. v. i. <https://bit.ly/3A17i6t>. (4. 3. 2024).

Lišková, M. (2013): Způsob označování stylové příznakovosti ve výkladovém slovníku. *Jazykovědné aktuality*, 50 (3–4), s. 154–158.

Petkevič, V. – Koprivová, M. – Hnátková, M. – Jelínek, T. – Kopriva, P. – Rosen, A. – Skoumalová, H. – Vondříčka, P. (2020): Typologie víceslovných jednotek v češtině a frekvenční zastoupení jejich hlavních vlastností v žánrově vyváženém korpusu. *Studie z aplikované lingvistiky*, 11 (2), s. 37-62.

Šemelík, M. – Lišková, M. – Křivan, J. (2023). K modifikované koncepci Akademického slovníku současné češtiny. *Bohemistika*, 23 (2), s. 275–296.

### Slovníky

Akademický slovník současné češtiny [online] (2017–2024). Praha: Ústav pro jazyk český AV ČR, v. v. i. Cit. 04. 03. 2024. <https://slovníkcestiny.cz>

Čermák, F. et al. (2009): *Slovník české frazeologie a idiomatiky 1.–4.* Praha: Leda.

Červená, V. et al. (2006): *Slovník spisovné češtiny pro školu a veřejnost: s Dodatkem Ministerstva školství, mládeže a tělovýchovy České republiky*. Praha: Academia.

Havránek, B. (ed.) et al. (1989): *Slovník spisovného jazyka českého*. Praha: Academia.

## Nestandardní adjektivní přirovnání v korpusových datech

Jaroslav Emmer (Jihočeská univerzita v Českých Budějovicích)

Adjektivní přirovnání (např. *slepý jako patrona*) patří mezi ustálené frazeologické jednotky, jejichž dvěma hlavními funkcemi jsou atribuce na základě analogie a intenzifikace přisuzované vlastnosti (Moonová, 2008; Norrick, 1986). Z hlediska konstrukční gramatiky (Langacker, 1987; Croft, 2007) se jedná o srovnávací konstrukce objevující se v mnoha jazycích, avšak často s různými lexikálními komponenty pro tutéž vlastnost (srov. angl. *blind as a bat* s čes. *slepý jako patrona*). Lze tedy očekávat určitou míru lexikální interference, zejména z angličtiny jakožto mezinárodního jazyka.

V 21. století se stále častěji setkáváme s digitálními slovníky, které nejsou limitovány počtem stran. Tyto podmínky umožňují neustálé rozšiřování a doplňování nových výrazů, což může vést i k zahrnutí výrazů nestandardních. Problém spočívá v tom, že slovníky nadále působí jako forma institucionalizace výrazu, tedy jakási validace jeho existence v daném jazyce. Hlavní jazykovým zdrojem slovníků jsou dnes korpusy, proto je důležité pečlivě zvažovat, jaké korpusy a jaká data použít pro jejich rozšiřování.

Cílem tohoto příspěvku je ilustrovat rozdíl mezi přeloženými a původními jazykovými daty na příkladu adjektivních přirovnání a poukázat na vliv cizích jazyků (zejména při překladu), který může potenciálně způsobovat nahrazování (domácích) ustálených lexikálních komponentů těmi přejatými. Pro účel této studie byla sesbírána adjektivní přirovnání z paralelního korpusu InterCorp v16 (čeština) – zde byly vyloučeny texty generované v češtině. Jelikož se v tomto korpusu ve většině případů objevují texty překládané, lze očekávat určitý vliv původních jazyků. Vytěžený seznam 239 typů adjektivních přirovnání z InterCorp v16 byl následně konfrontován s daty z korpusu Syn v12, z nějž byly naopak odfiltrovány texty, jejichž zdrojovým jazykem není čeština. Takto vznikly podmínky pro srovnání textů do češtiny přeložených s texty v češtině generovanými. Relativní frekvence (i.p.m.) každého typu přirovnání v Syn v12 slouží jako doklad užívání daného přirovnání v češtině, přičemž významně vyšší frekvence přirovnání v InterCorpu v16 (ve srovnání se Syn v12) naznačuje cizí vliv.

Na základě výsledků zkoumání jsou adjektivní přirovnání rozdělena do čtyř kategorií: (i) ustálené přirovnání (např. *bílý jako stěna*), (ii) řidší lexikální transformace ustáleného přirovnání, často ve formě synonymního slova pro jeden z komponentů (např. *lehký jako facka* vs. prototypické *jednoduchý jako facka*), (iii) přirovnání obsahující komponenty, které se objevují v jiných typech frazému (např. *hladový jako vlk* vs. prototypické *mít hlad jako vlk*), (iv) přirovnání přeložené z cizího jazyka (např. *slepý jako netopýr* vs. standardní *slepý jako patrona*). Ve více než 30 % adjektivních přirovnání z InterCorpu v16 se objevuje pro češtinu nestandardní lexikální komponent. V 11,3 % případů všech vytěžených typů přirovnání jde s velkou pravděpodobností o vliv cizího jazyka, zbylých zhruba 20 % lze interpretovat jako řidší či nestandardní lexikální transformace (Čermák, 2007) ustáleného přirovnání, které nutně nenaznačují cizí vliv. Obecně je možné konstatovat, že v InterCorpu v16 se objevuje mnohem více nestandardních adjektivních přirovnání než v Syn v12, což je s velkou pravděpodobností způsobeno interferencí zdrojových jazyků.

Zjištění předkládaná v této studii ilustrují úskalí spojená se sestavováním slovníkových seznamů na základě korpusových dat a nutnost jejich důsledné analýzy a interpretace. Korpusové frekvenční seznamy nezřídka obsahují adjektivní přirovnání, která nejsou pro češtinu standardní, a zejména překládané texty mohou vyvolávat dojem, že dochází ke zvýšenému výskytu nových variant. Z tohoto pohledu je původní jazyk, ve kterém byl text generován, naprosto zásadní kritérium. Zůstává otázkou, zda vůbec využívat data z překládaných textů při tvorbě slovníků daného jazyka, jelikož je riziko výskytu nestandardních konstrukcí nesporně vyšší.

## Bibliografie

Čermák, F. (2007). *Frazeologie a idiomatika česká a obecná / Czech and General Phraseology*. Praha: Univerzita Karlova v Praze – Nakladatelství Karolinum.

Croft, W. (2007). Construction grammar. In Dirk Geeraerts and Hubert Cuyckens (eds.) *Handbook of Cognitive Linguistics*, 463–508. Oxford: Oxford University Press

Langacker, R.W. (1987). *Foundation of Cognitive Grammar (Vol. 1). Theoretical Prerequisites*. Stanford: Stanford University Press.

Moon, R. (2008). Conventionalized as-similes in English: a problem case. *International Journal of Corpus Linguistics*, 13(1), 3–37.

Norricks, N. R. (1986). Stock Similes. *Journal of Literary Semantics XV/1*, 39–52.

## Korpusy

Křen, M. – Cvrček, V. – Čapka, T. – Hnátková, M. – Jelínek, T. – Koček, J. – Kovářiková, D. – Křivan, J. – Milička, J. – Petkevič, V. – Skoumalová, H. – Šindlerová, J. – Škrabal, M.: *Korpus SYN, verze 12 ze 29. 12. 2023*. Ústav Českého národního korpusu FF UK, Praha 2023. Dostupný z WWW: <https://www.korpus.cz>

Rosen, A. – Vavřín, M. – Zasina, A. J.: *Korpus InterCorp – čeština, verze 16 z 12. 10. 2023*. Ústav Českého národního korpusu FF UK, Praha 2023. Dostupný z WWW <http://www.korpus.cz>

### Víceslovné lexikální entity korpusově

Marie Kopřivová<sup>1</sup>, Hana Skoumalová<sup>2</sup>, Hana Goláňová<sup>3</sup>, Milena Hnátková<sup>2</sup>, Anna Christou<sup>4</sup>, Tomáš Jelínek<sup>2</sup>, Jan Křivan<sup>4</sup>, Vladimír Petkevič<sup>2</sup>, Pavel Vondříčka<sup>2</sup>, Karla Tvrdá<sup>4</sup>  
 (1Ústav bohemistiky pro cizince a komunikace neslyšících FF UK, 2Ústav Českého národního korpusu FF UK, 3Katedra českého jazyka a literatury UHK, 4Ústav pro jazyk český AV ČR)

Cílem příspěvku je představení projektu, který se zaměřuje na rozšíření lexikografické databáze LEMUR, na její softwarové aplikační využití při korpusové anotaci a na uživatelské propojení databáze s jazykovými korpusy a s Akademickým slovníkem současné češtiny (ASSČ) [5]. Projekt vychází ze stávající frazeologické anotace užívané v korpusech řady SYN [4, 3, 2] a z databáze víceslovných lexikálních jednotek (VLJ) LEMUR [8]. Kromě výrazného rozšíření počtu VLJ [1,7] v databázi se zaměřuje na vytvoření nového anotačního programu, který bude schopen zachytit v českých textech i fragmenty víceslovných jednotek a jejich kombinace [6]. Při rozšiřování databáze o nové jednotky dochází také k úpravám a opravám existujících hesel, a to jak v klasifikaci jednotlivých kolokací (dochází např. k revizi přísloví), tak ve změně struktury databáze (přibývají nové kategorie).

V příspěvku uvedeme příklady na rozšíření repertoáru VLJ např. o filmové hlášky nebo regionální VLJ. Nově byla již vytvořena první verze programu, který bude v korpusu vyhledávat a označovat víceslovné jednotky a nahradí tak dosavadní anotační program FRANTA [4]. Rozdíl nebude pouze v použitém programu, ale především v tom, že databáze LEMUR, která nahradí slovník FRANTALEX, popisuje jednotlivé VLJ podle několika ortogonálních os. Tento popis odráží také značka, která bude každé označené VLJ přiřazena a která poskytne uživateli bohatší informaci. Kromě značek popisujících celou lexikální jednotku, budou mít i jednotlivé členy VLJ vlastní značku popisující jejich lexikální vlastnosti. Tak například v kolokacích ‘mrkat na drát’ nebo ‘skočit na špek’ budou všechna slova označena jako stylově neutrální, ale kolokace jako celek bude označena jako expresivní. Kromě detailnějšího popisu nalezených VLJ si od programu slibujeme větší rychlost, což by umožnilo nasazení tohoto programu např. již v průběhu morfologického značkování pro zvýšení jeho úspěšnosti. Zatím se využívá jen malá část slovníku FRANTALEX právě z důvodu zpomalení procesu.

Dále bychom se rádi zaměřili na ukázkou toho, jak propojení slovníku, korpusu a databáze umožní efektivně získat informace využitelné ve výuce (na příkladu přísloví). Již nyní, v pilotní verzi systému, je možné z korpusu prokliknout do databáze a získat další informace o vyhledaném frazému. Plánujeme ale také možnost prokliku z ASSČ do databáze a naopak. Kromě tohoto propojení korpusu, databáze VLJ a slovníku ASSČ vznikne i elektronická cvičebnice frazémů určená pro různé typy škol.



## Bibliografie

- [1] Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In: Fred J. Damerau & Nitin Indurkha (eds.), *Handbook of natural language processing*, 2nd edn., 267–292. Boca Raton, FL: CRC Press, Taylor & Francis Group.
- [2] Čermák, František. 2007. *Czech and general phraseology*. Prague: Karolinum.
- [3] Čermák, František et al. 1983–2009. *Slovník české frazeologie a idiomatiky* (SČFI), vol. 1–4. Prague: Academia/Leda.
- [4] Hnátková, Milena. 2002. Značkování frazémů a idiomů v Českém národním korpusu s pomocí Slovníku české frazeologie a idiomatiky. *Slovo a slovesnost* 63(2). 117–126.  
<http://sas.ujc.cas.cz/archiv.php?art=4064>.
- [5] Kochová, Pavla, Opatková, Zdeňka (eds.). 2016: *Kapitoly z koncepce Akademického slovníku současné češtiny*. ÚJČ AV ČR.
- [6] Skoumalová Hana, Kopřivová Marie, Petkevič Vladimír, Jelínek Tomáš, Rosen Alexandr, Vondříčka Pavel & Hnátková Milena. 2024. A lexicon of Czech multiword expressions. In: Giouli V. & Barbu Mititelu V. (eds.), *Multiword expressions in lexical resources: Linguistic, lexicographic, and computational perspectives, 1–32. (Phraseology and Multiword Expressions)*. Berlin: Language Science Press.
- [7] Villavicencio, Aline, Baldwin Timothy & Waldron, Benjamin. 2004. A Multilingual Database of Idioms. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).
- [8] Vondříčka, Pavel. 2019. Design of a Multiword Expressions Database. *The Prague Bulletin of Mathematical Linguistics* 112. 83–101. <https://ufal.mff.cuni.cz/pbml/112/art-vondricka.pdf>.

**„Jak se do lesa volá, tak se z lesa ozývá“:  
Frazémy jako metakomunikační vyjádření v současných psaných mediálních textech**

Kamila Mrázková (Filozofická fakulta Univerzity Karlovy)

Příspěvek analyzuje užívání frazémů referujících k řečovému jednání (kontextově synonymních s verbi dicendi, Šoltys 1983, Hirschová 2017) i jeho účinku (hodnoceného z pozice referujícího) v současných online publikovaných psaných masových médiích, a to v závislosti na typu či žánru těchto médií, jak jsou kategorizovány v korpusech Českého národního korpusu online2\_now a online2\_archiv (tj. např. hlavní proud, bulvární média, politický bulvár, antisystémové weby, názorové deníky; srov. Cvrček – Jeziorský – Henyš 2022). Jedná se o spojení jako (*padala*) *ostrá slova* či účinek řeči hodnotící *namazat si koho na chleba*. Součástí výzkumu je srovnání frekvence těchto frazémů ve výše jmenovaných korpusech s korpusem staršími a rozsáhlejšími, zejména syn2013pub. Cílem je ověřit hypotézy týkající se stylových rozdílů mezi uvedenými typy médií, resp. toho, zda se tyto rozdíly projevují také v typu a frekvenci metajazykových vyjádření konceptualizujících komunikaci (srov. Lakoff – Johnson 1980, Chrz 1999, Vaňková – Nebeská et al. 2005), o níž se referuje, jako nepřátelský konflikt, nahlížený zejména prostřednictvím emocí komunikačních partnerů. Na této konceptualizaci se podílejí i jednoslovné výrazy ve funkci verba dicendi, zpravidla adherentně expresivní (Zima 1961, Mrázková 2013), jako např. *vpálit (něco někomu)*; předmětem analýzy v tomto příspěvku jsou tehdy, pokud se opakovaně objevují v kolokaci, která je potenciálně frazeologizovatelná, např. *vpálit (něco někomu) do obličeje/tváře/ksichtu*. Zájem o tento typ výrazů souvisí s výzkumným projektem Česká frazeologie a proměny jejího užívání v dobových a žánrových kontextech,<sup>2</sup> jehož je prezentovaný výzkum součástí a mezi jehož cíle patří mj. vyhledávání, identifikace a anotace frazémů a jejich variant a modifikací v nově budovaných korpusech (k variabilitě a modifikaci českých frazémů v úzu srov. Jelínek – Kopřivová et al. 2018).

---

<sup>2</sup> <https://jakobson.korpus.cz/dokuwiki/doku.php?id=frape:start>

## Bibliografie

Cvrček, V. – Jeziorský, T. – Henyš, J. (2022): ONLINE2\_NOW: monitorovací korpus internetové češtiny. Ústav Českého národního korpusu FF UK, Praha 2022 [cit. 2024-03-22]. Dostupný z WWW: <http://www.korpus.cz>

Česká frazeologie a proměny jejího užívání v dobových a žánrových kontextech. GA ČR 24-10254S. URL: <https://jakobson.korpus.cz/dokuwiki/doku.php?id=frape:start>.

Hirschová, M. (2017): VERBUM DICENDI. In: P. Karlík, M. Nekula – J. Pleskalová (eds.), CzechEncy - Nový encyklopedický slovník češtiny.

URL: <https://www.czechency.org/slovník/VERBUM DICENDI> (poslední přístup: 21. 3. 2024).

Chrz, V. (1999): Metafory v politice. Praha: PsÚ AV ČR.

Jelínek, T. – Kopřivová, M. et al. (2018): Variabilita českých frazémů v úzu. Časopis pro moderní filologii 2, 151–175.

Lakoff, G. – Johnson, M. (1980): Metaphors we live by. Chicago: University of Chicago Press.

Mrázková, K. (2013): Emotivní funkce a lexikální expresivita v současné české publicistice. In: K. Klímová – I. Kolářová – J. M. Tušková (eds.), Stylistika v kontextu historie a současnosti. Brno: Masarykova univerzita, s. 199–207.

Šoltys, O. (1983): Verba dicendi a metajazyková informace, 1983.

Vaňková, I. et. al. (2005): Co na srdci, to na jazyku. Praha: Karolinum.

Zima, J. (1961): Expresivita slova v češtině. Praha: Nakl. Československé akademie věd.

**Všechno souvisí se vším, ale některé věci nejdou dohromady:  
k frazémům s recipročním významem**

Sylva Nzimba, Vojtěch Veselý (Ústav pro jazyk český AV ČR)

Reciprocita je vymezována jako symetrická relace mezi participanty vyjadřovaná některými slovesy, resp. jistými syntaktickými konstrukcemi: jsou-li participanty reciproční situace právě dva, pak platí, že A je v určitém vztahu k B a současně B je v druhově stejném vztahu k A (jestliže platí tvrzení *A a B se navzájem respektují*, pak nutně platí i konjunkce *A respektuje B a B respektuje A*). Reciprocita byla kromě sloves popsána u substantiv, adjektiv nebo adverbii (viz např. Panevová 2007, Kettnerová – Lopatková 2018, Kettnerová – Kolářová 2023), frazémům s recipročním významem však bylo dosud věnováno jen málo pozornosti, srov. stručnou zmínku Nedjalkovu (Nedjalkov 2007: 14) o tom, že jedním z typů tzv. lexikálních reciprok jsou verbonominální kolokace typu *shake hands, be on good terms*.

Cílem tohoto příspěvku je postihnout sémantické rysy frazémů s recipročním významem, popsat vztahy mezi frazeologickým a „doslovným“ významem (tím, jak je reciproční situace frazémem konceptualizována) a zjistit, zda existují určité formální prvky typické pro reciproční frazémy.

Naším materiálovým východiskem byly slovesné a větné frazémy zpracované ve *Slovníku české frazeologie a idiomatiky* (Čermák – Hronek 2009) a v nyní vznikajícím *Akademickém slovníku současné češtiny*. Excerpce probíhala manuálně na základě významu lexikálních jednotek, celkem bylo vyexcerpováno cca 200 položek. Tento materiál zahrnoval jak frazémy, které nutně, v každém kontextu vyjadřují reciproční děj nebo stav (*promluvit si s někým jako chlap s chlapem, jeden (je) za osmnáct, druhý bez dvou za dvacet*), tak frazémy, jejichž reciproční význam vzniká teprve v důsledku odpovídajícího syntaktického užití (*držet na sebe* – frazeologickou jednotkou je tu *držet na někoho*).

Významovou strukturu skupiny recipročních frazémů lze znázornit pomocí modelu radiální kategorie (Lakoff 1987): prototypickou významovou skupinou, k níž mají další významové skupiny určitý vztah, jsou frazémy vyjadřující symetrický stav ‚A je stejný / podobný jako B‘ (*být ze stejného / jednoho těsta*). Prototypický význam je asociován např. s frazémy vyjadřujícími navázání či trvání milostného vztahu nebo přátelství (*mít něco s někým / spolu, držet spolu / dohromady*), s frazémy vyjadřujícími dosažení souladu (*plácnout si na to (s někým / spolu)*) či s frazémy označujícími sexuální akt (*hupsnout / hupnout s někým / spolu do postele*) – všechny tyto situace jsou (v různé míře) spojeny s představou souladu povah, smýšlení apod. Významovým opakem prototypu jsou frazémy vyjadřující stav ‚A je odlišné od B‘ (*jeden (táhne / jde) hot a druhý čehý*). Významový rys ‚nesoulad, neshoda‘ lze nalézt u frazémů s významy ‚A a B se rozešli / rozvedli‘ (*být od sebe*), ‚A a B se hádali, nadávali si‘ (*mít s někým výstup*), ‚A a B si vzájemně škodí, dráždí se‘ (*házet si klacky pod nohy*), ‚A a B se perou‘ (*dát si po čuni*) nebo ‚A a B se nemají rádi, nenávidí se‘ (*jsou (na sebe) jako (dva) kohouti (na jednom smetišti)*).

Reciproční děj či stav může být konceptualizován jako děj/stav realizovaný komitativně nebo kooperativně (*sedat s někým za jedním stolem*), jako nereciproční (jednosměrný) děj (*vypadnout z oka někomu*) či nereciproční (nesymetrický) stav (*jeden (je) lepší než druhý*). Multiplexní, tj. rozčlenitelná reciproční situace (srov. Haspelmath 2007, s. 2106n) může být za určitých podmínek konceptualizována jako situace uniplexní, tj. nerozčlenitelná (frazém *stát při sobě* ‚podporovat se‘ denotuje nejméně dva děje) a naopak uniplexní reciproční situace může být konceptualizována jako multiplexní (frazém *dát si sbohem* ‚rozloučit se‘ denotuje pouze jeden děj). Pro frazémy obecně je také charakteristická konceptualizace jako fyzický směrový reciproční děj (*padnout si kolem krku*).

Formální znaky typické pro reciproční frazémy jsme našli zejména u přirovnání (koordináční skupina se spojkou *a* (*něco / to je (jak(o)) nebe a dudy*) nebo komponent *dva* (*vrkat / cukrovat jak(o) (dvě) hrdličky*) v komparátu přirovnání), ale také např. u polypropozičních frazémů (dva členy signalizující opozici dvou rozdílných participantů, např. *já – ty, já – on, jeden – druhý* apod.: *já o voze, ty / on o koze*).

## Bibliografie

*Akademický slovník současné češtiny* [online] (2017–2024). Praha: Ústav pro jazyk český AV ČR. Dostupný online na adrese: <https://slovníkcestiny.cz>.

ČERMÁK, F. – HRONEK, J., et al. (2009): *Slovník české frazeologie a idiomatiky 1–4*. Voznice: LEDA.

HASPELMATH, M. (2007): Further remarks on reciprocal constructions. In: V. Nedjalkov (ed.), *Reciprocal Constructions*. Amsterdam/Philadelphia: John Benjamins, 2087–2115.

KETTNEROVÁ, V. – KOLÁŘOVÁ, V. (2023): K reciprocitě adjektiv v češtině. *Slovo a slovesnost*, 84 (3), 179–200.

KETTNEROVÁ, V. – LOPATKOVÁ, M. (2018): Lexicographic Potential of the Syntactic Properties of Verbs: The Case of Reciprocity in Czech. In: J. Čibej – V. Gorjanc – I. Kosem – S. Krek (eds.), *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana: Ljubljana University Press, 685–698.

LAKOFF, G. (1987): *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

NEDJALKOV, V. et al. (2007): *Reciprocal constructions*. Amsterdam/Philadelphia: John Benjamins.

PANEVOVÁ, J. (2007): Znovu o reciprocitě. *Slovo a slovesnost*, 68 (2), 91–100.

## Mezi frazeologií a konstrukční gramatikou: frazémy s opakováním verbálního komponentu

Zora Obstová (Charles University)

Anglické pojmy *constructional (schematic) idioms* a německé výrazy *Phraseoschablonen* či *Phrasem-Konstruktionen* označují částečně lexikálně obsazené syntaktické struktury obsahující volné pozice, které jsou obsazovány a aktualizovány teprve v kontextu. Navzdory proměnlivým konkrétním realizacím se tyto útvary vyznačují určitou syntaktickou idiomatičností a specifickým významem, který je mnohdy pragmatické či diskurzívní povahy. Pro svou hybridní povahu a postavení na pomezí lexikonu a syntaxe jsou tyto struktury obtížně uchopitelné tradičními lingvistickými přístupy a byly dlouho opomíjeny nejen v syntaktických, ale i ve frazeologických studiích. V mnoha jazycích včetně češtiny jim nebyla věnována téměř žádná pozornost; v jiných jazykových kontextech, zejména v tom ruském a německém, byly v posledních padesáti letech zkoumány jako okrajový typ frazému, jako určitá frazeologická kuriozita. Do popředí zájmu se dostaly teprve díky konstrukční gramatice (KG), pro niž představují ideální předmět zkoumání, mezi jiným i proto, že svědčí ve prospěch teze konstrukčněgramatického kontinua mezi lexikonem a syntaxí. Právě potřeba vytvořit teorii, která by dokázala uspokojivě popsat a vysvětlit i takovéto nesnadno klasifikovatelné útvary, nakonec vyústila ve vznik konstrukční gramatiky: struktury jako "the X–er, the Y–er" (*The harder I try, the less I accomplish*) se objevují už ve slavném článku Ch. J. Fillmora, P. Kaye, and C. O'Connorové (1988), který je považován za jeden z manifestů KG. V posledních dvou desetiletích se však zdá, že zájem konstrukční gramatiky o idiomatičké struktury poněkud klesá a KG a frazeologie se vydávají odlišnými cestami (Ziem 2018). Přesto však vyšlo v posledních letech několik publikací, které poukazují na nesporný přínos propojení frazeologického a konstrukčněgramatického přístupu (např. Mellado Blanco, Mollica & Schafroth (eds.) 2022 či Casadei & Koesters Gensini (eds.) 2024).

Stejný cíl si klade i tento příspěvek. V prezentaci se zaměříme na struktury s reduplikací verbálního komponentu, jako např. *stůj co stůj* či *ať dělá, co dělá*, ukážeme, jakým způsobem jsou zpracovány ve *Slovníku české frazeologie a idiomatiky*, a následně se je pokusíme popsat z konstrukčněgramatického hlediska jako více či méně produktivní konstrukční frazémy. Analýzu doplníme o kontrastivní rozměr: zajímavé srovnání s italštinou, kde nalézáme struktury s obdobnou formou i blízkým významem, poněkud zpochybňuje konstrukčněgramatickou tezi, že „constructions are language-specific“ (Croft 2001: 6) a že v mezijazykovém porovnání lze narazit nanejvýš na sémantické a funkční podobnosti.

Český jazykový materiál čerpáme z referenčních reprezentativních korpusů řady SYN a webového korpusu Araneum Bohemicum; pro italštinu jsme využili referenčního reprezentativního korpusu CORIS (boloňská univerzita) a webového korpusu Araneum Italicum.

## Bibliografie

CASADEI, F. & KOESTERS GENSINI, S. E. (eds.), *I fraseologismi schematici, Questioni descrittive e teoriche*, Roma: Aracne 2024.

CROFT, W., *Radical Construction Grammar. Syntactic theory in typological perspective*, Oxford: Oxford University Press, 2001.

ČERMÁK, F. a kol., *Slovník české frazeologie a idiomatiky (SČFI)*, 5 sv., Praha: Leda 2009-2016.

FILLMORE, Ch. J., KAY, P. & O'CONNOR, C., Regularity and idiomaticity in grammatical constructions. The case of let alone, *Language* 64, 1988, s. 501–538.

MELLADO BLANCO, C., MOLLICA, F. & SCHAFROTH, E., *Konstruktionen zwischen Lexikon und Grammatik: Phrasem-Konstruktionen monolingual, bilingual und multilingual*, Berlin, Boston: De Gruyter, 2022, s. 1-18.

ZIEM, A., Construction Grammar meets Phraseology: eine Standortbestimmung, *Linguistik online* 90 (3), 2018, s. 3–19. Dostupné z: <https://bop.unibe.ch/linguistik-online/article/view/4316/6450>.

## Jak hloupý je Hloupý Honza? Frazémová analýza jazykového prototypu v českém kontextu a jeho porovnání s chorvatštinou a makedonštinou

Slavomira Ribarova (Filozofski fakultet Sveučilišta u Zagrebu)

Příspěvek se zabývá českým frazémem *hloupý Honza* (nebo frazémy obsahujícími komponent *hloupý Honza*), který metaforicky popisuje člověka s nízkou rozumovou schopností, a srovnáním tohoto frazému s podobnými frazémy v makedonštině a chorvatštině.

Podle dosavadních výzkumů (Grenerová, 2011) je nejčetnější podskupinou v rámci českých frazémů obsahující antroponyma, patřících do frazeosémantického pole rozumové schopnosti a způsobilosti člověka, skupina s aspektem *hlouposti*. Prototypem hloupého člověka se v českém jazyce stal právě *hloupý Honza*. I když pohádkový hrdina "hloupý Honza" může ve skutečnosti vykazovat bystrý úsudek, v českém frazeologickém kontextu, tj. ve slovnících (Čermák, 2009) je tento výraz spojován spíše s negativními charakteristikami, jako je hloupost, naivita a prostoduchost.

Cílem příspěvku je s použitím Českého národního korpusu (ČNK) provést analýzu frekvence a kontextuálního užití frazému s komponentem *hloupý Honza* za účelem porozumění jeho významu a kulturním konotacím. Rovněž se zaměřím na možné aktualizace tohoto frazému a změny v jeho používání. Srovnání s chorvatštinou a makedonštinou s cílem identifikovat podobné prototypy v těchto jazycích a zdůraznit podobnosti a rozdíly ve vnímání a užívání frazémů obsahující antroponyma, které vyjadřují koncept *hlouposti* může osvětlit jazykovou a kulturní dynamiku.

Vzhledem k tomu, že frazém odráží specifické aspekty české kultury a vlastní jméno nese internacionální, nacionální a lokální konotace (Taylor 2002) srovnání s chorvatštinou a makedonštinou umožní porozumět rozdílům ve vnímání a užívání podobných prototypů, které jsou ovlivněny kulturními a sociálními faktory každé země. V obou jazycích existuje analogický způsob vytváření frazémů obsahující antroponyma, či význam je podobný - v makedonštině se používá frazém *Имап Пејо*, zatímco v chorvatštině je obdobným frazémem *Mujo iz Doboja*. Nicméně přestože způsob tvorby těchto frazémů v obou jazycích je podobný, jejich užití v kulturních kontextech každého ze zkoumaných jazyků je specifické. Například samotné jméno *Mujo iz Doboja* může nést specifické konotace a asociace pro chorvatskou kulturu („buranství“), které se mohou lišit od konotací jména *Имап Пејо* v makedonštině („vychytralost“) nebo *hloupý Honza* v češtině.

Příspěvek se rovněž dotkne možných strategií překladu frazému do jiných jazyků.



## Bibliografie

- Bezljaj, F. et al. = Безлај Ф. и др. [red.] (1983) *Osnoven sistem i terminologija na slovenskata onomastika / Основен систем и терминологија на словенската ономастика / Скопје: МАНУ.*
- Čermák, F. (2007) *Frazeologie a idiomatika česká a obecná. Czech and General Phraseology*. Prague: Karolinum.
- Čermák, F. et al. (2009) *Slovník české frazeologie a idiomatiky 1.–5*. Praha: Leda.
- Dimitrovski i Širilov = Димитровски, Т. и Ширилов. Т. (2003) *Фразеолошки речник на македонскиот јазик А-Ј том I*. Скопје: Огледало.
- Fink Arsovski, Ž., Kovačević, B. i Hrnjak, A. (2023) *Bibliografija hrvatske frazeologije s frazeobibliografskim rječnikom* Zagreb: KNJIGRA.
- Frančić, A. (2005) Imena u frazeologiji, *Učitelj: godišnjak Visoke učiteljske škole u Čakovcu* (ur. S.Hranjec), 5, Visoka učiteljska škola u Čakovcu, Čakovec, 31–41.
- Grenarová, R. (2011) *Jednoduché ustálené komparace s antroponymy a spojkou jako – jak v češtině a ruštině*. 1. vyd. Brno: Masarykova univerzita.
- Ivanetić, N. i Karlavaris-Bremer, U. (1991) Antroponimske fraze u njemačkom i hrvatskom, U: *Prožimanje kultura i jezika* (ur. M. Andrijašević i Y. Vrhovac), HDPL. Zagreb, 123–126.
- Matešić, J. (1982) *Frazeološki rječnik hrvatskoga ili srpskoga jezika*. Zagreb: Školska knjiga.
- Matešić, J. (1992/1993): Frazemi s komponentom vlastitog imena u hrvatskom jeziku, *Filologija*, 20–21, HAZU, Razred za filološke znanosti, Zagreb, 293–297.
- Menac, A., Fink Arsovski, Ž., Venturin, R. (2014) *Hrvatski frazeološki rječnik*, Zagreb: Naklada Ljevak
- Parizoska, J., Poredbeni frazemi s vlastitim imenima u hrvatskome - korpusni pristup [Similes with proper nouns in Croatian: A corpus-based approach] *Journal: Slavistički studii*, Spisanie na Katedrata za slavistika pri Filološkiot fakultet „Blaže Koneski“, Vol. 18; 139–150.
- Pleskalová, J. (2014) *Vlastní jména osobní v češtině*, Brno: Masarykova universita.
- Raguž, D. (1979) Vlastita imena u frazeologiji, *Onomastica Jugoslavica*, knj. 8, JAZU, Zagreb, str. 17–23.
- Ribarova, S. = Рибарова, С. (2018) Фреквентни машки имиња застапени во фраземите (од македонскиот, чешкиот и хрватскиот јазик), *Славофраз 2017: Имињата и фразеологијата. Имена и фразеологија* (ур. К. Велјановска и Б. Мирчевска-Бошева), Универзитет „Св. Кирил и Методиј“, Филолошки факултет „Блаже Конески“, Скопје, 251–262.
- Širilov = Ширилов, Т. (2008, 2009) *Фразеолошки речник на македонскиот јазик К-II том II., Р-III том III*. Скопје: Огледало.
- Taylor, J. (2002). *Cognitive Grammar*, Oxford: Oxford University Press.
- Veljanovska, K., Mirčevska- Voševa, B. = Велјановска, К., Мирчевска Бошева, Б. (2019) Антропонимите во македонската фразеологија. *Славистички студии*, Скопје, 149–160.
- Veljanovska, K., Mirčevska-Voševa, B. = Велјановска, К., Мирчевска Бошева, Б. (2021) *Имињата во македонската фразеологија*. Скопје: Антолог.

## Korpusové zdroje:

*Croatian Web* (CLASSLA–web.hr) <https://www.clarin.si/ske/#open>

*Czech Web* (2023 csTenTen23) <https://www.clarin.si/ske/#open>

*Macedonian Web* (CLASSLA–web.mk) <https://www.clarin.si/ske/#open>

*Český národní korpus* (SYN (verze 12) <https://www.korpus.cz/>

## K historii vzniku Etymologického slovníku české frazeologie

Ludmila Stěpanova

Diachronní aspekty české frazeologie stále zůstávají jednou z dosud neprobádaných oblastí. Čím intenzivněji se vyvíjí frazeologie jako samostatná lingvistická disciplína, tím zřetelnější je propast mezi její synchronní rozpracovaností a nedostatečným studiem diachronního aspektu. Příčinou může být: nedostatek starých pramenů obsahujících frazémy, absence monografických děl věnovaných výzkumu historie frazeologie, nedostatek historických a etymologických slovníků. Zanedbává se rovněž skutečnost, že historická analýza frazémů se liší od analýzy historie lexémů už proto, že jako etymon tu není slovo, ale slovní spojení, je tedy nesmírně důležité objevit strukturně sémantický model tohoto ustáleného výrazu. K nalezení takového modelu přispívá zachycení co nejvyššího počtu variant daného rčení (včetně nářečních). Je nezbytné též odhalit, zda je daný frazém původní, nebo převzatý. K tomu pomůže stanovení areálu šíření tohoto úsloví, stanovení chronologie vzniku tohoto rčení v různých jazycích aj.

Doufáme, že se impulsem pro historická bádání v oblasti frazeologie stane náš Etymologický slovník české frazeologie (cca 4300 frazémů), vydaný v r. 2022 v Olomouci. Slovník zahrnuje pouze frazémy, tj. ustálené reprodukovatelné expresivní jednotky, které mají celistvou sémantiku. V každém hesle se uvádí frazém, jeho význam, výklad historie a etymologie rčení a seznam pramenů, v nichž je zmíněna etymologie daného frazému. Rozsah informací o původu frazémů se může lišit: jednou to je výklad odpovídajícího místa z Bible nebo z antických mýtů, podruhé vysvětlení zastaralého slova či odborného termínu nebo zastaralé syntaktické konstrukce. Důležitou etapou etymologické analýzy bylo též srovnání určitého frazému s materiálem jiných slovanských a evropských jazyků.

Jsme si vědomi toho, že se žádný etymologický slovník (navíc slovník frazeologický) nemůže vyhnout nedostatkům i chybám, vnímáme to jako nástin možné cesty k historickému popisu bohaté české frazeologie.

## K frazeologické kompetenci druhé generace mluvčích češtiny jako zděděného jazyka na příkladu českých přísloví a hlášek

Kateřina Šichová (Universität Regensburg)

Pod pojmem frazeologická kompetence je třeba chápat dvě roviny: Jedná se jednak o jistý druh vědění o specifičnosti frazeologických jednotek (Čermák 2007:25–42), jednak o určité množství frazémů daného jazyka, které uživatel tohoto jazyka zná a/nebo užívá (srov. také Burger et al. 1982:172, Hallsteinsdóttir 2001:11n.). U každého mluvčího má tato kompetence co do kvality i kvantity jinou podobu. Výzkumy k frazeologické kompetenci jako celku jsou velmi sporé, ať už se týkají mateřského/prvního jazyka mluvčích, nebo jazyka, který se mluvčí učí jako cizí. Pro češtinu byly dosud provedeny pouze sondy, a to převážně ve školním prostředí (např. Čechová 2012:105–142, ke studentům češtiny jako cizího jazyka viz Hasil 2011:64–100) a/nebo primárně s ohledem na znalost určitých typů frazeologických jednotek (např. biblická frazeologie u Šindelářové 2021, přísloví u Bittnerové a Schindlera 2004). Zatím neprozkoumanou skupinou mluvčích jsou tzv. *heritage speakers*, tedy lidé, kteří žijí v zahraničí a různou měrou ovládají určitý jazyk jako jazyk zděděný od (jednoho ze) svých rodičů (obecně k těmto mluvčím např. Bremer/Mehlhorn 2018, sondy k mluvčím češtiny jako zděděného jazyka (čzj) viz Nekula 2021). Frazeologické kompetenci u druhé generace (dospělých) mluvčích češtiny jako zděděného jazyka, jejichž dominantním jazykem je němčina, bude věnována pozornost v předkládaném příspěvku (sociální parametry respondentů a stručná charakteristika jejich jazykového profilu viz Šichová 2024).

První poznatky o znalostech českých frazeologických syntagmat, resp. jejich formy a významu, naznačují, že u mluvčích čzj je množina ovládnutých českých frazémů menší, resp. že míra ovládnutí souvisí s typem frazémů a mírou jejich idiomaticity. Ukazuje se také, že psycholinguvistická stabilita frazémů je u mluvčích čzj slabší než u monolingválních rodilých mluvčích. Zároveň se v testech a v rozhovorech vyjevují nejen strategie, které tito mluvčí při mentálním zpracovávání frazémů užívají, ale i vliv dominantního jazyka na jazyk zděděný (srov. Šichová 2021). Cílem příspěvku je tento výzkum rozšířit a zaměřit se na další dva typy specifických frazeologických jednotek: hlášky a přísloví.

(Filmové) hlášky, v českém kontextu i jako součást internetových memů (k memům z lingvistického pohledu srov. Osterroth 2019) či důležitý zdroj mezitextového navazování (Mrázková/Homoláč 2020) mezi mluvčími velmi oblíbené, pochází především z divácky populárních komedií a filmových pohádek (Kopřivová/Yu Ju 2023). Otázku, zda jsou tato – v jazyce relativně nová – slovní spojení včetně povědomí o zdroji, resp. provázanosti se situacemi, v nichž se (hlavně v mluveném projevu) používají, známá i mladým mluvčím čzj a zda tedy patří k (lingvo)realiím, které znají, zodpovíme pomocí testů a experimentu s obrázky a ukázkami z filmů.

Přísloví naopak patří ke starým, ale velmi vitálním jazykovým jednotkám, a užívají se především v psaném jazyce. Základ pro předkládanou analýzu paremiologických znalostí (práce s textem, cloze-test) tvoří přísloví, která patří v češtině k nejfrekventovanějším a jsou tedy součástí paremiologického optima (Kopřivová/Šichová 2023).

Analyzovaná data (testy, texty, pozorování, interview) také naznačí, jaké mají sledovaní mluvčí čzj povědomí o charakteristických znacích frazémů a zda tyto vědomosti hrají roli při zpracovávání frazémů ve zděděném jazyce.

## Bibliografie

- Bittnerová, Dana / Schindler, Franz (2004): *Česká přísloví. Soudobý stav konce 20. století*. Praha: Karolinum.
- Bremer, Bernhard / Mehlhorn, Grit (2018): *Herkunftssprachen*. Tübingen: Narr.
- Burger, Harald / Buhofer, Annelies/Sialm, Ambros (1982): *Handbuch der Phraseologie*. Berlin: De Gruyter.
- Čechová, Marie (2012): *Řeč o řeči*. Praha: Academia.
- Čermák, František (2007): *Frazeologie a idiomatika česká a obecná*. Praha: Karolinum.
- Hallsteinsdóttir, Erla (2001): *Das Verstehen idiomatischer Phraseologischer in der Fremdsprache Deutsch*. Hamburg: Verlag Dr. Kovač.
- Hasil, Jiří (2011): *Interkulturní aspekty mezikulturní komunikace*. Ústí nad Labem: Univerzita Jana Evangelisty Purkyně.
- Kopřivová, Marie / Šichová, Kateřina (2023): Proverbs In Contemporary Czech. Corpus Probe Into Written Texts. – In: *Jazykovědný časopis SAV*, 74/2023, s. 92–100.
- Kopřivová, Marie / Yu Ju Chen (2023): „Furt králíka!“ a další filmové hlášky. – In: *Jazykovědné aktuality* 1–2/23, s. 76–87.
- Mrázková, Kamila / Homoláč, Jiří (2020): „Ó frétná chrochtobuznosti!“ Mezitextové navazování na twitterových profilech českých novinářů a politiků. – In: Jílková, Lucie et al. (eds.): *Jak je důležité mítí styl. Pocta Janě Hoffmannové*. Praha: Nakladatelství Lidové noviny, s. 337–362.
- Nekula, Marek (2021) (ed.): *Čeština ve světě / Czech in the World. Special issue Naše řeč*, 5/104.
- Osterroth, Andreas (2019): „How to do things with Memes? – Internet-Memes als multimodale Sprechakte“. – In: Bülow, Lars / Johann, Michael (eds.): *Politische Internet-Memes – theoretische Herausforderungen und empirische Befunde*. Berlin: Frank & Timme, s. 41–60.
- Šichová, Kateřina (2024): „Ta gramatika a taki to psání mě tak nejde (jak tady v tém textu vidíte)“: Jazykové kompetence druhé generace mluvčích češtiny jako zděděného jazyka. – In: *Český jazyka a literatura* 2023/24, č. 4, s. 172–177.
- Šichová, Kateřina (2021): Zur phraseologischen Kompetenz bei Herkunftssprechern des Tschechischen. – In: *brücken* 28/2, s. 111–144.
- Šindelářová, Jaromíra (2021): Znalost a chápání biblické frazeologie v České republice podle dosaženého stupně vzdělání jejích obyvatel. – In: Janovec, Ladislav (ed.): *Svět v obrazech a ve frazeologii II*. Praha: Pedagogická fakulta Univerzity Karlovy, s. 70–96.

10

*Manifestations of sound  
symbolism in language:  
phonaesthemes, ideophones  
and onomatopoeia*

organizers

Jan Křivan

Veronika Zikmundová

(ENG)

# Workshop description

Linguistic research oriented towards the language use is increasingly turning its attention to describing, interpreting, and explaining phenomena that violate conventional notions of language as a system of rules based on arbitrary signs (cf. Diessel 2017). While such phenomena have been known for a long time (cf. Jakobson 1957; for instance, knowledge of iconic and indexical signs nowadays belongs to the fundamentals of linguistics), in traditional post-Saussurian structure-oriented accounts they are usually treated rather as exceptions to a grammatical or lexical system. Functionally, typologically and cognitively oriented approaches, on the other hand, aim to describe these phenomena in their entirety, as systems corresponding to principles of a different type. One of such principles is sound symbolism (also known as phono semantics, phonesthesia, phonological iconicity), under whose manifestations one can include, among others, linguistic units referred to as phonaesthemes, ideophones (also mimetics, expressives) or onomatopoeia. While a phonaestheme is a cluster of sounds (of a submorphemic nature) carrying a certain meaning (Firth 1964: 184), i.e. a unit at a lower level, whole classes of words are referred to in languages as ideophones and onomatopoeia. Onomatopoeia are generally known as words that imitate sounds (e.g. Matthews 2014), ideophones (the term coined by Doke 1935) are usually defined as “an open lexical class of marked words that depict sensory imagery” (Dingemanse 2019: 16). In both cases, we are dealing with words bearing a special semantics and a specific morphosyntactic status, not infrequently employing morphologically expressive means such as reduplication or vowel lengthening (Zwicky & Pullum 1987).

The aim of this workshop is to bring together researchers working on sound symbolism in language from different perspectives, not only in the context of Czech linguistics.

Invited speaker: Thomas Van Hoey (University of Leuven)

We welcome contributions addressing comparative and more general methodological questions, as well as original empirical papers devoted to the description of sound symbolism in specific languages, including the treatment of related special and applied topics. Below we give examples of possible topics.

General and comparative questions:

- The question of ideophones as a descriptive category: To what extent does the definition of ideophones as a distinct word class depend on the morphological type of a particular language (cf., e.g., Jendraschek 2001)?
- What is the role of expressive morphology (cf. Dingemanse & Akita 2017; or Oberfalzerová 2009 on Mongolian)? Does expressive morphology occur across language types?
- What semantic categories appear across languages or in particular languages for ideophones and for other sound-symbolic words (cf. Ibarretxe-Antuñano 2019 on motion ideophones)?
- Whether and how are sound-symbolic words subject to language contact?
- What registers (and pragmatic functions) are associated with the use of ideophones and onomatopoeia?

Descriptive contributions:

- Descriptions of phonaesthemes, onomatopoeia or ideophones in specific languages, preferably based on own (e.g. field, experimental or corpus) research and touching upon the abovementioned questions.

Special and applied research questions:

- How should sound-symbolic words be translated across languages?
- How does sound symbolism manifest itself in language acquisition and language disorders?
- The functioning of sound-symbolic words in child directed speech: to what extent does this register bear the features traditionally associated with ideophones in languages?

## References

- Diessel, Holger. 2017. Usage-Based Linguistics [online]. In: Aronoff, M. (ed.), *Oxford Research Encyclopedia of Linguistics*. New York: Oxford University Press.  
<<https://doi.org/10.1093/acrefore/9780199384655.013.363>>.
- Dingemans, Mark. 2019. "Ideophone" as a comparative concept. In: Akita, K. & Pardeshi, P. (Eds.), *Ideophones, Mimetics, and Expressives*. John Benjamins, 13–33.
- Dingemans, Mark, & Akita, Kimi. 2017. An Inverse Relation between Expressiveness and Grammatical Integration: On the Morphosyntactic Typology of Ideophones, with Special Reference to Japanese. *Journal of Linguistics* 53 (3), 501–532.
- Doke, Clement Martyn. 1935. *Bantu Linguistic Terminology*. London: Longmans, Green, and Co.
- Firth, John Rupert. 1964 [1930]. *The Tongues of Men, and Speech*. London: Oxford University Press.
- Ibarretxe-Antuñano, Iraide. 2019. Towards a semantic typological classification of motion ideophones: The motion semantic grid. In: Akita, K. & Pardeshi, P. (Eds.), *Ideophones, Mimetics, and Expressives*. John Benjamins, 137–166.
- Jakobson, Roman. 1957. *Shifters, verbal categories, and the Russian verb*. Massachusetts: Harvard University Press.
- Jendraschek, Gerd. 2001. Semantic and structural properties of Turkish ideophones. *Turkic Languages* 5, 88–103.
- Matthews, Peter. 2014. Onomatopoeia. In: *The Concise Oxford Dictionary of Linguistics*. Oxford University Press.
- Oberfalzerová, Alena. 2009. Onomatopoeia and iconopoeia—as an expressive means in Mongolian. *Mongolo-tibetica Pragensia '09, Ethnolinguistics, Sociolinguistics, Religion, and Culture*, 2(1), 29–60.
- Zwicky, Arnold & Pullum, Geoffrey. 1987. Plain morphology and expressive morphology. *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 330–340.  
<<https://doi.org/10.3765/bls.v13i0.1817>>

# Table of Contents

*From similar place avoidance to sound symbolism in Slavic languages*

Aleš Bičan (Masaryk University)

*The semantic categories of Korean ideophones*

Blanka Ferklová (Palacky University)

*Imitative words and language contact: their borrowing in the history of English*

Maria Flaksman (Ludwig-Maximilian University of Munich)

*Measuring the meaning and structure of ideophones: Lessons from Chinese*

Thomas Van Hoey (KU Leuven)

*Sound of Fire, Water and Wind – Contrastive study of Czech and Japanese*

Petra Kanasugi (ÚAS, FFUK)

*What can a data-driven approach reveal about ideophones?*

Tatiana Korol (independent researcher), Tatiana Nikitina (INALCO, CNRS)

*A general perspective on ideophones – Where does Abui fit in?*

František Kratochvíl, Georges Saad (Palacký University)

*Large language models' ability to generate iconic pseudowords*

Anna Marklová<sup>1</sup>, Jiří Milička<sup>1</sup>, Leonid Ryvkin<sup>2</sup>, Ludmila Bennet<sup>3</sup>, Libuše Kormaníková<sup>3</sup> (<sup>1</sup>Charles University, <sup>2</sup>Claude Bernard University Lyon 1, <sup>3</sup>Palacký University Olomouc)

*Morphosyntactic properties of ideophones of Central and Northeast Asia*

Vít Ulman (Palacký University Olomouc)

*The effect of sound symbolism on the processing of Czech words in persons with aphasia*

Eva Pospíšilová, Michal Láznička (Charles University)

*The importance of Sakha 'ideophones' for the general theory of ideophones*

Jan Židek, Jonáš Vlasák, Aldana Vlasáková (Charles University)

*Mongolian ideophones: Semantics and change*

Veronika Zikmundová, Jan Křivan, Vít Ulman, Veronika Kapišovská, Alena Oberfalzerová (Charles University)



## From similar place avoidance to sound symbolism in Slavic languages

Aleš Bičan (Masaryk University)

It has been proposed that sound-symbolic lexemes display phonological properties markedly different from those of neutral arbitrary vocabulary (Hinton et al. 1994). At the same time, phonologists have described many deviations from cross-linguistical phonological preferences such as sonority reversals, velar consonant + front vowel combinations, or syllabic obstruents. Since phonologists are traditionally interested in conventional vocabulary, little is known about the extent sound-symbolic items are driven by the same preferences or whether they are marked by their own phonological peculiarities.

Another example of a strong cross-linguistic preference is the principle known as Similar Place Avoidance (SPA; Pozdniakov & Segerer 2007), also known as Obligatory Contour Principle (McCarthy 1986), according to which repetitions of consonants of the same place of articulation are avoided in CVC sequences. The principle has been mostly studied for non-Indo-European languages, but our lexical data for 18 Slavic and 42 non-Slavic languages confirm the effect for the Indo-European family. Despite its near-universal status, the SPA principle is not a categorial restriction. Languages are known to contain lexical items violating it, but the existence of these items is just passingly acknowledged without further analysis (a notable exception is Pozdniakov's 2007 discussion of Russian). Sound-symbolic vocabulary and expressive words are suggested as potential exceptions to the principle. This calls for further consideration.

To determine the nature of SPA violations in Slavic languages, we have excerpted non-borrowed roots containing a CVC sequence with same-place consonants from etymological dictionaries of Old Church Slavonic, Slovenian, and Czech. As such dictionaries usually account for conventional vocabulary, we have supplemented this material with items from our own hitherto undescribed database of 4,700+ basic verbal roots (plus 4,000+ of their variants) collected mainly from Czech dialect dictionaries. Dialects are known for their vividness, expressivity, and playfulness and are thus expected to contain more sound-symbolic vocabulary than literary-language dictionaries record.

The data reveal that the homorganic CVC sequences within native roots are indeed endowed with a sound-symbolic function. However, this origin prevails only among labial and velar sequences but not among alveolar and palatal sequences. Coronals thus seem less suitable for expressing sound symbolism, which may be explained by their unmarked status, less restricted occurrence, and greater combinatory potential in contrast to non-coronal places (Paradis & Prunet 1991).

As Table 1 shows, the Slavic roots with SPA-violating CVC sequences having a sound-symbolic function can be divided into three broad classes: (A) onomatopoeic and interjectional roots denoting sounds emitted by humans, animals, or (moving) objects, (B) child-language formations, and (C) phonaesthetic patterns. Phonaesthemes have been shown to exist in Slavic languages (e.g., Mompean et al. 2020, Sláma 2023). We will argue that they cannot be limited to particular phoneme sequences such as (C1) but should also include more abstract phonotactic patterns such as (C2). The latter roots are realized with different vocalism and consonantism that cannot be explained by regular sound changes or dialectal variation. Yet, these phonotactic patterns recur both in SPA-violating and SPA-compliant roots and can be associated with a particular semantic concept.

In short, we will demonstrate that sound-symbolic words are indeed phonologically unique in contrast to the conventional vocabulary. By setting off from the well-known phonotactic constraint on the co-occurrence of consonants, we will offer an analysis of potential sound-symbolic vocabulary in Slavic languages. Since our extensive database of verbal roots offers many examples of sound symbolism, the analysis will not be restricted to the items violating the mentioned restriction. Our contribution will be primarily descriptive. Although sound symbolism has been studied for Slavic languages (Fidler 2013, Koleva-Zlateva 2020), it needs to be substantiated with more evidence before any theories and explanations can be proposed. The latter, however, will be attempted in the contribution too.

Table 1: Sound-symbolic patterns in SPA-violating Slavic roots (and other examples in parentheses)

Type	Pattern	Meaning/function	Examples (selective)
A	various	sounds of humans, animals, and (moving) objects	<i>brblat</i> “to mumble”, <i>bafat</i> “to smoke noisily”, <i>pifnout</i> “to shoot”, <i>huhňat</i> “to twang”, <i>gágat</i> “to gaggle”
B	Stop <sub>i</sub> -V-Stop <sub>i</sub>	related to the immediate world of a child	<i>papat</i> “to eat”, <i>máma</i> “mother”, <i>kakat</i> “to poo”, <i>bubat</i> “to drink”, <i>dadat</i> “to sleep”
C1	/šk/	rattling noise (of tearing or rubbing)	<i>škohrtat</i> “to emit rattling noise”, <i>rozškrkovat se</i> “to tear apart” (cf. also <i>škrtat</i> “to scratch”, <i>škablat</i> “to scramble”, <i>škrpat</i> “to scuff one’s feet”, <i>škýrat</i> “to scribble”)
	/af/ /ap/ (a more common variant)	“to grasp, to take, to touch; to ingest”	<i>bafat</i> “to eat” (cf. also <i>blafat</i> “to eat greedily”, <i>t’afnout</i> “to catch”; <i>čapat</i> “to catch”, <i>zjapat</i> ditto, <i>lapat</i> ditto, <i>chňapat</i> “to catch (in one’s mouth)”, <i>rapnout</i> “to grasp”, <i>drapnout</i> “to grasp quickly”, <i>chapnout</i> “to catch, to take”, <i>d’abat</i> “to touch”, <i>gabat</i> ditto, <i>chlapat</i> “to drink”)
	/pVp/, /bVb/, /bVmb/	“to be round or to become round (voluminous)”	<i>pupek</i> “belly”, <i>pupen</i> “bud”, <i>bublina</i> “bubble”, <i>boubelatý</i> “chubby”, <i>bobule</i> “berry”, <i>bob</i> “bean”, <i>bobtnat</i> “to swell”, <i>bubřet</i> ditto, <i>bambule</i> “pompom”
C2	Nasal-Stop	“repeated motion or sound”	<i>bambat</i> “to wiggle”, <i>honkat se</i> ditto (cf. also <i>kondat</i> “to wiggle”; <i>cindat se</i> , <i>kambat se</i> , <i>selungat se</i> , <i>pšundat</i> all “to ramble around”)
	C <sub>i</sub> VC <sub>i</sub> C <sub>i</sub> řC <sub>i</sub>	“(repeated) motion”	<i>babonit</i> “to walk around and seek sth.”, <i>drdat</i> “to walk with small steps”, <i>cackovat</i> “to run, lollop”, <i>sisat</i> “to waggle”, <i>d’oudat se</i> “to walk slowly”, <i>kyklat</i> “to sway”
	C <sub>i</sub> VC <sub>i</sub> -Son.	“to toy with sth.; to do something wrongly or uncleanly”	<i>prplat se</i> “to niggle”, <i>bablat se</i> ditto, <i>fufňat se</i> ditto, <i>kykmat se</i> “to mess around in”, <i>gigrat</i> “to pick at one’s food”, <i>uchuchlat</i> “to prepare food badly”, <i>umrmlat</i> “to do sth. unskillfully”

### Bibliography

- Fidler, M. 2013. *Onomatopoeia in Czech: A window into grammar and discourse*. Slavica.
- Hinton, L., J. Nichols & J. J. Ohala. 1994. Introduction: sound-symbolic processes. In L. Hinton, J. Nichols & J. J. Ohala (eds.), *Sound symbolism*, 1–12. Cambridge University Press.
- Koleva-Zlateva, Z. 2020. Phonosemantics (phonosymbolism, sound symbolism). In Marc L. Greenberg & Lenore A. Grenoble (eds.), *Encyclopedia of Slavic Languages and Linguistics Online*. [http://dx.doi.org/10.1163/2589-6229\\_ESLO\\_COM\\_032009](http://dx.doi.org/10.1163/2589-6229_ESLO_COM_032009) (accessed 18/3/2024).
- McCarthy, J. 1986. OCP effect: Gemination and antigemination. *Linguistic inquiry* 17(2). 207–263.
- Mompean, J. A., A. Fregier & J. Valenzuela. 2020. Iconicity and systematicity in phonaestemes: A cross-linguistic study. *Cognitive linguistics* 31(3). 515–548.
- Paradis, C. & J.-F. Prunet (eds.). 1991. *Phonetics and phonology 2: The special status of coronals*. Academic Press.
- Pozdniakov, K. 2007. O ‘plohih’ i ‘horoshih’ zakah v russkom jazyke. In N. Vakhtin & G. Levinton (eds.), *Hommage Albert Baiburin*, 78–119. Université Européenne. [http://pozdniakov.free.fr/publications/2007\\_Baiburin.pdf](http://pozdniakov.free.fr/publications/2007_Baiburin.pdf) (accessed 18/3/2024).
- Pozdniakov, K. & G. Segerer. 2007. Similar place avoidance: A statistical universal. *Linguistic typology* 11. 307–348.
- Sláma, J. 2023. Inicialní konsonantické fonestémy v češtině. *Bohemistyka* 23(1). 39–54.

## The semantic categories of Korean ideophones

Blanka Ferklová (Palacky University)

The Korean language is covered by several large digital dictionaries and corpora produced by the National Institute for Korean Language and by some university departments (Seoul National University's Kkokkoma project, Yonsei University's Newspaper corpora, etc.). The research on ideophones, however, is largely limited by the fact that ideophones are not classified as a separate word class in Korean. On the contrary, interjections in Korean constitute a separate word class, although the distinction and relation to Onomatopoeia in Korean have also first had to be clearly defined (Song 1988, CHae 2006). Ideophones in Korean are classified mainly based on syntactic criteria as a subcategory of adverbs (Sohn 1999, Pucek 2005). Therefore, they are also tagged as such in digital sources, and searching for them in corpora is still quite laborious, even in specialized corpora. In Czech, the problem is similar, as onomatopoeias are classified as part of interjections and tagged with the same tag, yet corpora have been successfully used in their research (Fidler 2006, 2014). Despite these difficulties, more and more research on Korean ideophones is making use of these resources, and this paper also plans to present results obtained both from the study of the relevant literature and specialized dictionary (Lee 2007) and from available Korean corpora.

As Van Hoey states in his survey study of the semantic categories of ideophones (2022, 2023), Korean ideophones cover all the semantic categories he postulates (sound, sight, somatosensation, and inner sense), including the proposed subcategories. However, Korean ideophones are unevenly distributed across categories, and some specific subcategories show both higher frequency of use and higher number of members (such as in subcategory movement, e.g. manner of walking like *kōntül-kōntül* 'swishly', *pitül-pitül* 'sweaty', or the sound-image category of smiling and laughing eg. *pangkūs* 'beamily', *pangsil* 'with a wide smile', *silsil* 'gigglingly'). In this presentation, these categories will be presented in the context of their features such as sound symbolism and reduplication (Park 2008). These phenomena cause difficulties in the statistical exploration of Korean ideophones because, as one of the earliest Western grammars of Korean notes, "Koreans are particularly fond of constructing onomatopoeic expressions of various kinds, almost ad libitum and at the choice of the speaker himself". (Ramsted 1938, 159). Another point that will be mentioned are the so-called *üisōngüitchaeö*, words that imitate sound and appearance at the same time (Chae 2003), such as *sogon-sogon* 'to speak in a whisper, secretly' or *ttallang-ttallang* 'to tinkle', representing both auditory or visual perception, and such as cover more than one semantic category and, in this regard, has to be considered with special attention.

## Bibliography

CHae, Wan, *Hanguğöüi üisöngö üitçaeö [Korean onomatopoeias and mimetics]*. Seoul: Söultaehakkyo chulpchanbu, 2003.

CHae, Wan, "Üisöngöwa üitçaeö ch'ai [The difference between onomatopoeia and ideophones] ". *Sae kugö saenghwal* 16/ 4 (2006): 125-131.

Fidler, Masako. *Onomatopoeia in Czech : a conceptualization of sound and its connections to grammar and discourse*. Bloomington, Indiana : Slavica, 2014.

Fidler, Masako. "Conceptualization of Onomatopoeia: Observations from Czech Data" *Zeitschrift für Slawistik* 51, no. 4 (2006): 375-385. <https://doi.org/10.1524/slwa.2006.51.4.375>

I Kiwön (Lee Kai Won). *Korean Onomaotpoiea and Mimesis*. Seoul: Hangungmunhwasa, 2007.

Park, Dong-geun. *Han'gugüi hyungnaemarüi ihä. [Understanding Korean imitation words]*. Seoul: Youkrack, 2008.

Pucek, Vladimír. *Gramatika korejského jazyka*. Prague: Karolinum, 2005.

Song Mun-jun. "Sori hyungnaemarüi ssikarüme taehayö [On word class classification of onomatopoeia] ". *Hangeul* 200/6: 139-164, 1988.

Van Hoey, Thomas. "A semantic map for ideophones." (2022). (to be printed and accesed from <https://osf.io/preprints/osf/muhpd>

Van Hoey, Thomas, Arthur L. Thompson, Youngah Do, and Mark Dingemans. "Iconicity in ideophones: Guessing, memorizing, and reassessing." *Cognitive Science* 47, no. 4 (2023): e13268.

## Imitative words and language contact: their borrowing in the history of English

Maria Flaksman (Ludwig-Maximilian University of Munich)

Imitative words (also termed “iconic”, and variously classified as “onomatopoeic”, “sound symbolic” words, and “ideophones” by different authors) are found in languages of different families [6], [13] as well as in ancient and reconstructed languages [2], [8], [9]. These are words with *iconic* (non-arbitrary) correlation between form and meaning. I differentiate imitative words according to *means* of imitation – acoustic means (onomatopoeia) and articulatory imitation (mimesis, imitation of movements of speech apparatus (SAM) for non-speech production purposes – coughing, sneezing).

There is an objective relation of *similarity* between the acoustic form of imitative words and the natural sounds / SAM which they denote. For example, low-pitched tones (blowing of wind, crying of wolves, etc.) are rendered by words containing low-pitched vowels (*hoot* [hu:t], *boom* [bu:m], *zoom* [zu:m], etc.), whereas the high-pitched ones (cries of small birds or mice) – by words containing high-pitched vowels (*peep* [pi:p], *cheep* [tʃi:p], *yip* [jɪp]). Objective form-meaning similarity in imitative words creates a basis for their universally recognized cross-linguistic similarity (cf. E. *meow*, Est. *näü*, Vietn. *meo*, and Chin. *miao* [7]). Persistent iconic coinage and apparent cross-linguistic similarity of imitative words seem to diminish the need for borrowing in this stratum of vocabulary: why borrow something which can be coined so easily? However, my previous research [4] has shown that English imitative lexicon (by which I understand the sum of *all* words imitative by origin from different time strata) is comprised of loanwords from other languages, mostly Latin and French, up to 20%.

The investigated *material* was comprised of 1244 lexemes selected by continuous sampling from etymological dictionaries of the English language [1], [10], [11], [12]. Words marked “onomatopoeic”, “echoic”, etc. in these dictionaries were included into the investigated data corpus. These included non-lexicalized ideophone-like imitative interjections like *zzz!*, *brr!*, *splish-splash*, as well words derived from imitative roots (e.g., *ping-pong* ‘game of table tennis’, *gaga* ‘a madman’). The *methods* of research included: (1) etymological analysis, (2) componential (semantic) analysis, (3) phonosemantic analysis (after [14], [15]), and (4) the method of diachronic evaluation of the imitative lexicon [4], [5]. The latter is a research tool applied for the description of imitative words in diachrony. I proposed measuring the degree of iconicity (form-meaning similarity) loss in imitative words by distinguishing four stages of *de-iconization* (SDs). SD-1 words are non-lexicalized onomatopoeic interjections (*shh!*, *zzz!*); SD-2 words are content imitative words without form or meaning changes (*plop*, *hiss*, *swish*); SD-3a words are words which have lost their original form after regular sound changes (*laugh* /la:f/ < Old English *hlehhan*); SD-3b words are words which have lost their original meaning due to semantic shifts (*ping-pong* ‘sound → game’); SD-4 imitative words have lost both their original form and meaning (*crazy* ‘mad’ < (?) Scandinavian, cf. Norw. *kras* ‘crashing sound’ [10]). The present study has *resulted* in revealing the following facts and tendencies:

- time and sources of borrowing of imitative words generally coincide with those for English vocabulary in general (as compared against [3])
- the majority of the borrowed imitative words come from French and Latin, followed by Scandinavian languages – e.g., *murmur* (Latin), *cliché* (French), *bang* (Scandinavian)
- the share of borrowed imitative words has been declining from 47% in the 13<sup>th</sup> century to 3% in the 20<sup>th</sup> century, whereas the number of source languages has been increasing
- phonetic adaptation and by borrowing accelerates de-iconization (triggers form changes and the loss of the original meaning)
- there is an overall negative correlation between an imitative words’ de-iconization stage and the probability of its being borrowed.

Thus, it can be *concluded* that language contact results in borrowing of imitative words just as it results in borrowing of the non-imitative ones. However, the more iconic the imitative word is, the less likely it is to be borrowed.

## Bibliography

- [1] BARN – Barnhart, R. K. 2006. *Chambers Dictionary of Etymology*. New York: Chambers.
- [2] Carling, G. 2017. Etymology and Iconicity in Onomatopoeia and Sound Symbolism. A Germanic Case Study. In B.A. Olsen (ed.), *Etymology and the European Lexicon*, Wiesbaden: Ludwig Reichert. P. 93-103.
- [3] Durkin, Ph. 2014. *Borrowed Words: A History of Loanwords in English*. Oxford: OUP.
- [4] Flaksman, M. 2015. Diachronic Development of English Imitative Vocabulary. PhD Thesis. University of St. Petersburg [in Russian].
- [5] Flaksman, M. 2017. Iconic Treadmill Hypothesis – The Reasons behind Continuous Onomatopoeic Coinage. In Bauer, M., Zirker, A., Fischer, O., and Ch. Ljungberg (eds.). *Dimensions of Iconicity*. Amsterdam: Benjamins. P. 15-38. <https://doi.org/10.1075/ill.15.02fla>
- [6] Hinton, L., J. Nichols and J.J. Ohala (eds.). 1994. *Sound Symbolism*. Cambridge: CUP.
- [7] *IA – Iconicity Atlas Project: A Multilingual Online Dictionary of Iconic Words and Ideophones*, ed. by M. Flaksman and N. Noland. URL: <http://www.iconicity-atlas.com/index.htm>. (accessed 29.01.2024).
- [8] Koleva-Zlateva, Zh. 2008. Slavianskaya leksika zvukosimvolocheskogo proiskhozhdeniya [Slavonic Words of Phono-Iconic Origin]. In *Tractata Slavica Universitatis Debreceniensis*, 1. Debrecen.
- [9] Kozlova, T. 2013. Iconically Expressible Meanings in Proto-Indo-European Roots and Their Reflexes in Daughter Branches. In L. Ellenstrom, O. Fischer and C. Ljungberg (eds). *Iconic Investigations*. Amsterdam: Benjamins. P. 311-330.
- [10] OED — *The Oxford English Dictionary* (3d edition). URL: <http://www.oed.com> (accessed 29.01.2024).
- [11] PARTR — Partridge, E. 2009. *An Etymological Dictionary of Modern English*. London-New York: Routledge.
- [12] SKEAT – Skeat, W.W. 2006. *An Etymological Dictionary of the English Language*. Oxford: Clarendon Press.
- [13] Voeltz, E.F.K., and Ch. Kilian-Hatz (eds.). 2001. *Ideophones*. Amsterdam: Benjamins.
- [14] Voronin, S. V. 2005. *Iconicity. Glottogenesis. Semiosis: Sundry Papers*. St. Petersburg: St. Petersburg University Press.
- [15] Voronin, S.V. 2006 [1982]. *Osnovy Phonosemantiki* [The Fundamentals of Phonosemantics]. Moscow: Lenand.

**Measuring the meaning and structure of ideophones: Lessons from Chinese**

Thomas Van Hoey (KU Leuven)

As researchers of lexical iconicity (e.g., onomatopoeia, ideophones) we encounter several challenges with regard to the identification and study of these words. (1) While some languages have ideophonic systems that are relatively well-studied, e.g., Japanese (Akita 2009), Siwu (Dingemanse 2011), Quechua (Nuckolls 1996), Basque (Ibarretxe-Antuñano 2017), the scholarly treatment of ideophonic systems in many other languages is still in its infancy or has not been recognized traditionally. (2) Although we have cross-linguistic terms like ‘ideophone’ (Dingemanse 2019), ‘mimetic’, ‘expressive’, and ‘onomatopoeia’, applying these terms to different languages with similar phenomena is not always an easy task. (3) And because iconic words are often relatively low in frequency (compared to words that are typically studied), we might not have immediate access to novel techniques like word embeddings to study their semantics.

This presentation uses Mandarin Chinese as a case study to address these three issues. First, iconic words in Chinese can be identified balancing the canonical typology features advanced by Dingemanse (2019) with previously identified linguistic phenomena. This approach helps establish a prototypical notion of the ideophone class in Chinese. Second, I present two ways of studying the semantic evolution of ideophones. The first takes a traditional corpus linguistic approach, focusing on the dynamic interplay of semasiological developments over time as well as onomasiological categorization and choice between different variants (see Geeraerts 1997). The second revisits the same data with the more recent methodology of count-based vector space modeling to measure meaning changes and variant choice (Sahlgren 2006; Silge 2017). It will be shown that both approaches offer benefits and pitfalls, leading to the conclusion that a multipronged approach is better, but also doable for languages that are typologically different from the better-known ideophone systems.

## Bibliography

- Akita, Kimi. 2009. *A grammar of sound-symbolic words in Japanese: theoretical approaches to iconic and lexical properties of mimetics* (日本語音象徴語文法：擬音・擬態語の類像的・語彙的特性への理論的アプローチ). Kobe: Kobe University PhD dissertation.
- Dingemanse, Mark. 2011. *The meaning and use of ideophones in Siwu*. Nijmegen: Radboud University Nijmegen dissertation.
- Dingemanse, Mark. 2019. “Ideophone” as a comparative concept. In Kimi Akita & Prashant Pardeshi (eds.), *Ideophones, mimetics and expressives* (Iconicity in Language and Literature, ILL 16), 13–33. Amsterdam: John Benjamins. DOI:10.1075/ill.16.02din.
- Geeraerts, Dirk. 1997. *Diachronic prototype semantics: a contribution to historical lexicology* (Oxford Studies in Lexicography and Lexicology). Oxford; New York: Clarendon Press; Oxford University Press.
- Ibarretxe-Antuñano, Iraide. 2017. Basque ideophones from a typological perspective. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 62(02). 196–220. DOI:10.1017/cnj.2017.8.
- Nuckolls, Janis B. 1996. *Sounds like life: sound-symbolic grammar, performance, and cognition in Pastaza Quechua* (Oxford Studies in Anthropological Linguistics 2). New York: Oxford University Press.
- Sahlgren, Magnus. 2006. *The word-space model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. Stockholm: Stockholm University PhD dissertation.
- Silge, Julia. 2017. *Tidy word vectors, take 2!* [2017-11-27]. <https://juliasilge.com/blog/word-vectors-take-two/>. (4 November, 2018).



## Sound of Fire, Water and Wind – Contrastive study of Czech and Japanese

Petra Kanasugi (ÚAS, FFUK)

It is a well-known fact that Japanese has a special lexical stratum of ideophones – signs with iconic nature, i.e. the form of the sign is directly or indirectly motivated by its meaning. This mimetic lexical stratum can be defined in terms of form, function and meaning. The group is relatively broad, comprising over 2000 types. Moreover, its meaning is not limited only to the auditory domain; its individual lexical items express meanings pertaining to visual and haptic modalities, as well as physical or psychological states. (Akita, Tsujimura: 2016) It is precisely because of its formal and semantic specificity that Japanese mimetics have been and continue to be the subject of research in Japan and beyond. (Akita, Pardeshi: 2019)

In recent years, however, there has been a growing interest and substantial research devoted to highlighting the iconic features of vocabulary even in languages that do not have a special layer of ideophones, or that have a very narrow one that concerns practically only the auditory domain. Recent research points to iconicity in the domains of size (Winter, Perlman: 2017), shape (Cwiek: 2022), and roughness. (Winter et al.: 2022) It is becoming increasingly clear from the totality of this research that iconicity is a widespread phenomenon that is complicated both by the fact that the phonological form of words is formed by the phonological repertoire of each given language, and by the fact that the meaning of particular phonemes is highly context-dependent (Akita, Imai: 2023). However, it is equally clear that it is possible to find parallels even in very different languages that point to the universal nature and power of iconicity.

For the present paper, I have chosen three basic forces of nature: *fire*, *water* and *wind*, which seem to be basic enough for there to be only a limited possibility of language contact effect. I will retrieve the analyzed lexical items from corpora of written language (for Japanese BCCWJ; for Czech syn 2020). I will focus on verbs (e.g. *kolotat*, *lokotat*, *zurčet*, *šumět*, *hučet*, *bublat*, *řvát*, etc.) in case of Czech since Czech as a satellite-framed language tends to express manner on the verb; and on the phonomimetics in case of Japanese (e.g. *sarasara*, *shitoshito*, *taratara*, *chorochoro*, *dosha*, *bachabacha* etc.). Apart from the two mentioned corpora I will also make use of Intercorp 16, parallel corpus of Czech and Japanese, to confirm equivalence of meanings.

The goal is first to determine how finely are the given semantic fields divided and along what lines; and define phonological parallels appearing in the lexical items carrying the same meaning (e.g. repeated occurrence of /o/ in /klokotat/ vs. /torotoro/ etc.) and link them to particular meanings.

The results should be of interest not only theoretically because we are dealing with two languages in which iconicity asserts itself very differently - in *conventional vocabulary* in Czech language and in a special lexical stratum in Japanese but also since we may earn some useful hints for translation.

## Bibliography

AKITA, Kimi; PARDESHI, Prashant (ed.). Ideophones, mimetics and expressives. John Benjamins Publishing Company, 2019.

AKITA, Kimi; TSUJIMURA, Natsuko. 4. Mimetics. Handbook of Japanese lexicon and word formation, 2016, 133-160.

AKITA, Kimi. Cognitive Science of Onomatopoe. Tankobon, 2022 (秋田喜美「オノマトペの認知科学」新曜社 2022)

ĆWIEK, Aleksandra, et al. The bouba/kiki effect is robust across cultures and writing systems. Philosophical Transactions of the Royal Society B, 2022, 377.1841: 20200390. DINGEMANSE, Mark, et al. Arbitrariness, iconicity, and systematicity in language. Trends in cognitive sciences, 2015, 19.10: 603-615.

WINTER, Bodo, et al. Trilled/r/is associated with roughness, linking sound and touch across spoken languages. Scientific Reports, 2022, 12.1: 1035.

WINTER, Bodo, et al. Which words are most iconic? Iconicity in English sensory words. Interaction

**What can a data-driven approach reveal about ideophones?**

Tatiana Korol (independent researcher), Tatiana Nikitina (INALCO, CNRS)

This study explores the properties of ideophones in Wan (ISO 639-3: wan) a Southeastern Mande language spoken in Côte d'Ivoire. The research is based on data collected from a corpus of traditional narratives. Building on earlier descriptive studies (Ameka 2001; Newman 1968; Childs 1988; Akita and Dingemans 2019) we seek to develop a new approach to identifying ideophones their properties and the role of expressive morphology. We focus on the meaning of ideophones syntactic restrictions on their use and their compatibility with other lexical items.

We start by analysing the morphosyntactic behavior of ideophones in Wan and contrasting it to that of other lexical classes. We arrive at a language-specific definition of Wan ideophones as well as a general methodology for defining the class of ideophones that can be applied across languages. We believe that such a methodology is beneficial in particular to non-native speakers engaged in the challenging task of describing ideophones of the language “in its own terms”. Such approach is important because we are dealing with words bearing a special semantics and a specific morphosyntactic status.

The study's primary aim is to develop a data-driven approach that avoids assuming pre-existing definitions or ad hoc classifications. As our starting point we rely on a database of 320 naturally occurring examples extracted from a corpus featuring 140 ideophone types. Each token was annotated its position in the clause morphophonological properties such as reduplication and co-occurrence with specific types of semantically related verbs and nouns. The examples were cross-checked with a native speaker consultant to ensure accuracy of our interpretation of the particular uses. Based on these annotations we arrive at natural classes of ideophones that turn out to only partially replicate the existing semantic classifications intended to capture ideophones' cross-linguistic diversity (Dhoorre and Tosco 1998; Hoey 2024; Watson 2001) including the implicational hierarchy proposed by Dingemans (Dingemans 2012).

We further explore correlations between the meaning of ideophones and their phonetic form revealing in particular associations between the ideophone's form and the aspectual properties of the corresponding event.

The results of the study illustrate the advantages as well as the challenges of a data-driven approach and highlight the importance of developing cross-linguistically applicable methodology for the study of ideophones.

## Bibliography

- Akita, Kimi, and Mark Dingemanse. 2019. 'Ideophones (Mimetics, Expressives)'. In *Oxford Research Encyclopedia of Linguistics*, by Kimi Akita and Mark Dingemanse. Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.477>.
- Ameka, Felix K. 2001. 'Ideophones and the Nature of the Adjective Word Class in Ewe'. In *Ideophones*, edited by F. K. Erhard Voeltz and Christa Kilian-Hatz, 25–48. Typological Studies in Language. John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.44.04ame>.
- Childs, G. Tucker. 1988. 'The Phonology of Kisi Ideophones' 10 (2): 165–90. <https://doi.org/10.1515/jall.1988.10.2.165>.
- Dhoorre, Cabdulqaadir Salaad, and Mauro Tosco. 1998. 'Somali Ideophones'. *Journal of African Cultural Studies* 11 (2): 125–56. <https://doi.org/10.1080/13696819808717831>.
- Hoey, Thomas Van. 2024. 'A Semantic Map for Ideophones', January. <https://doi.org/10.31219/osf.io/muhpd>.
- Newman, Paul. 1968. 'Ideophones from a Syntactic Point of View'. *J.W.A.L.*, 1968, v. 2 edition.
- Watson, Richard L. 2001. 'A Comparison of Some Southeast Asian Ideophones with Some African Ideophones'. In *Ideophones*, edited by F. K. Erhard Voeltz and Christa Kilian-Hatz, 385–405. Typological Studies in Language. John Benjamins Publishing Company. <https://doi.org/10.1075/tsl.44.29wat>.

**A general perspective on ideophones - Where does Abui fit in?**

František Kratochvíl, Georges Saad (Palacký University Olomouc)

Ideophones are a family of words that depict sensory perceptions in forms that evoke that perception in an iconic or mimetic fashion (Akita and Dingemanse 2019; Dingemanse 2019). Sometimes they are studied in broader context which includes other expressives (i.e. names or bad words) (cf. Klamer 2002) and cannot be separated from the larger context of human imitative behaviors which include facial expressions, gestures, etc.

Various aspects of ideophones have been studied in great detail. In addition to compiling detailed inventories, studies have focused on the question of iconicity and systematicity (cf. Dingemanse et al. 2015; Winter et al. 2017; Thompson and Do 2019; Van Hoey et al. 2023).

Another aspect of ideophones is the mapping of sensory domains and possible implicational hierarchies which predict which sensory domains are expressed universally (cf. Dingemanse 2012; McLean 2021). The attempts to establish ideophone universals have stimulated detailed mapping of ideophones cross-linguistically and ideophones are now established as a comparative concept (cf. Dingemanse 2019). This has led some to call for a mapping approach to come to a greater understanding of the conceptual space of lexicalised perception cross-linguistically (Van Hoey, to appear). This is a turn to the ‘unsayable’, and ‘untranslatable’, which suspects that languages may be more specialized and ‘exuberant’ in describing one type of experience of reality over another (cf. Ortega y Gasset 1957, through Nick Evans’ work).

The standard comparison framework for ideophones relies on the classification of sensory domains in Winter et al. (2017): (i) auditory (sound) > onomatopoeia; (ii) physical (movement); (iii) visual (shape & color); (iv) tactile (touch and texture); (v) olfactory (smell); or (vi) gustatory (taste).

This classification may need to be elaborated with other domains, such as (vii) somatic (words encoding inner body appraisals, often with emotions, sometimes called psychomimetics).

In addition to sensory domain mapping (following Van Hoey’s approach), we are interested in mapping the ideophones spanning multiple sensory domains as well as the possible transfer from one domain into another and the cultural and communicative practices in which the transfer takes place (humor, stance, emotion).

In our paper we take the Abui videophone system as a test case and try to fit it in the more general typology of ideophones that emerges from the works above.

**Large language models' ability to generate iconic pseudowords**

Anna Marklová<sup>1</sup>, Jiří Milička<sup>1</sup>, Leonid Ryvkin<sup>2</sup>, Ludmila Bennet<sup>3</sup>, Libuše Kormaníková<sup>3</sup>

(<sup>1</sup>Charles University, <sup>2</sup>Claude Bernard University Lyon 1, <sup>3</sup>Palacký University Olomouc)

While the principle of arbitrariness is crucial when connecting signifier and signified (Saussure, 1959), there are principles causing that the connection between form and meaning is not completely arbitrary. Namely, there are some systematic regularities (structural resemblances within word classes or regularities in length, e.g., Farmer et al., 2006; de Varda & Strapparava, 2022), and phenomena usually referred to as iconicity or sound symbolism. While phonological iconicity is the most researched and pronounced in ideophones, the extent of iconicity goes beyond this particular class, as shown in various experimental studies (see Perniss et al. 2010 for an overview).

In the present study, we tested the abilities of large language models (LLMs) to generate iconic pseudowords. Natural languages contain features not easily replicable by artificial means. In the present study, we prompted GPT4 to invent an iconic language and provide translations of Czech and German words into that language. The language of prompting was Czech and German, resulting in a list of Czech pseudowords, and a list of German pseudowords.

We used the task of pairing two Czech words with two words from an unknown language. Our previous studies on natural languages unknown to Czech speakers uncovered that in such a task, the success rate is around 60%, therefore higher than a random guess (Diatka & Milička, 2017). On the other hand, pairing the words with random pseudowords does not yield success higher than a random guess. We had four conditions with 100 participants per condition: (1) German speakers were presented with German pseudowords; (2) German speakers were presented with Czech pseudowords; (3) Czech speakers were presented with German pseudowords, and (4) Czech speakers were presented with Czech pseudowords. It was discovered that overall, the success rate was around 70%, therefore, higher than in the same task with natural languages (see Figure 1). This rate was comparable in all four conditions.

Our study shows that (1) LLMs demonstrate a capacity for generating pseudowords with a degree of iconicity, and (2) the generated pseudowords exhibit consistent levels of iconicity regardless of the language of the prompt. The second finding has several interpretations: firstly, there might be some underlying phonetic qualities of ideophonically sounding words that are not language-specific (we need to point out that German and Czech are both Indo-European languages from similar cultural-geographical backgrounds, therefore, we cannot easily generalize). Secondly, the reason for similar results may be caused by the fact that while the interface of GPT4 was in Czech (resp. German), the process of generating pseudowords might have been through English in both cases since it is an internal pivot language of LLMs (Wendler et al., 2024). That means that the LLM might not have taken specific languages into account. Lastly, the results might have been influenced by the particularities of the prompt. Our study provides initial insights into the capabilities of LLMs in replicating the iconicity found in natural languages.

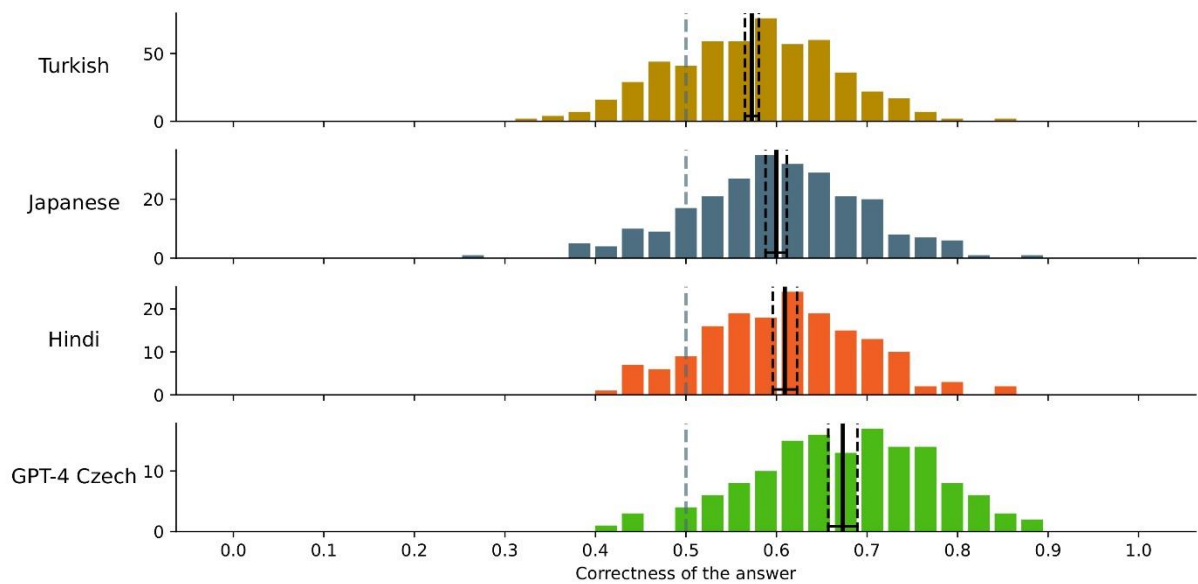


Figure 4: Proportion of correct guesses of Czech speakers assessing meanings of Czech GPT-4 generated pseudowords, compared to guesses of Turkish, Japanese, and Hindi words.

## Bibliography

Diatka, V., & Milička, J. (2017). The effect of iconicity flash blindness. *Dimensions of Iconicity*, 15, 3.

Farmer, T. A., Christiansen, M. H., and Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12203–12208. doi: 10.1073/pnas.0602173103.

Perniss, P., Thompson, R. L., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psychol.* 1:227. doi: 10.3389/fpsyg.2010.00227.

Saussure, F. de. (1959). *Course in General Linguistics*. New York, NY: Philosophical Library. (Original work published 1922).

de Varda, A. G., & Strapparava, C. (2022). A cross-modal and cross-lingual study of iconicity in language: Insights from deep learning. *Cognitive Science*, 46 (6), e13147. <https://doi.org/10.1111/cogs.13147>

Wendler, C., Veselovsky, V., Monea, G., & West, R. (2024). Do Llamas Work in English? On the Latent Language of Multilingual Transformers. Retrieved from arXiv preprint server (eprint 2402.10588)

**The effect of sound symbolism on the processing of Czech words in persons with aphasia**

Eva Pospíšilová, Michal Láznička (Charles University)

**Background:** During the last two decades, there has been a growing interest in the study of sound symbolism and iconicity across various disciplines including typology, evolutionary linguistics, and psycholinguistics. Both sound symbolism and iconicity are hypothesized to have played a pivotal role in the emergence of language (Perlman 2017; Ćwiek et al. 2021) and have been observed to facilitate language acquisition and processing (Imai et al. 2008; Lockwood, Dingemanse & Hagoort 2016; Sidhu, Vigliocco & Pexman 2020). While there has been a discussion about a processing advantage for both language acquisition and language impairments (Ramat 1995), the majority of research has primarily focused on the former, with a notable exception being the study conducted by Meteyard et al. (2015). Through employing four distinct tasks (repetition, reading aloud, auditory lexical decision and visual lexical decision), they have demonstrated that iconicity aids spoken word processing for individuals with aphasia. Building upon the findings of Meteyard et al., our study aims to investigate whether iconic words are processed easier than non-iconic words among Czech speakers with aphasia.

**Methodology:** Firstly, we have selected 36 Czech iconic words and 36 matched control words. Most of these words were translations of the stimuli used by Meteyard et al., the rest of the stimuli were words that were marked as onomatopoeia in the Czech Etymological Dictionary (Rejzek 2015), all the selected iconic words were onomatopoeic in nature. The control words have been matched in terms of word class, corpus frequency, and length. Subsequently, the set of selected words have been rated for sound iconicity by Czech speakers without aphasia. The selected and rated words were then used in lexical decision task and a repetition task with a group of participants with aphasia and a group of matched neurotypical speakers. Data analysis (in progress) encompassed both error rates and reaction times. We expect to find similar advantages as Meteyard et al. for iconic words, i.e. the iconic stimuli will be faster and have fewer errors.

**Implications:** The outcomes of this experiment are expected to enrich our understanding of sound symbolism and its manifestation and function in language impairments. Research of sound symbolism in aphasia could enhance the interest of studying iconicity in the context of other language disorders, such as specific language impairment or dyslexia. The results also have the potential to be used in therapeutic practice (e.g. in the creation of diagnostic tools and new therapeutic procedures).



## Bibliography

Ćwiek, Aleksandra, Susanne Fuchs, Christoph Draxler, Eva Liina Asu, Dan Dediu, Katri Hiovain, Shigeto Kawahara, et al. 2021. Novel vocalizations are understood across cultures. *Scientific Reports*. Nature Publishing Group 11(1). 10108. <https://doi.org/10/gjz5x3>.

Imai, Mutsumi, Sotaro Kita, Miho Nagumo & Hiroyuki Okada. 2008. Sound symbolism facilitates early verb learning. *Cognition* 109(1). 54–65. <https://doi.org/10.1016/j.cognition.2008.07.015>.

Lockwood, Gwilym, Mark Dingemans & Peter Hagoort. 2016. Sound-symbolism boosts novel word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42(8). 1274–1281. <https://doi.org/10.1037/xlm0000235>.

Meteyard, Lotte, Emily Stoppard, Dee Snudden, Stefano F. Cappa & Gabriella Vigliocco. 2015. When semantics aids phonology: A processing advantage for iconic word forms in aphasia. *Neuropsychologia* 76. 264–275. <https://doi.org/10.1016/j.neuropsychologia.2015.01.042>.

Perlman, Marcus. 2017. Debunking two myths against vocal origins of language: Language is iconic and multimodal to the core. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems* 18(3). 376–401. <https://doi.org/10.1075/is.18.3.05per>.

Ramat, Anna Giacalone. 1995. Iconicity in the Grammaticalization Processes. In Raffaele Simone (ed.), *Iconicity in Language*, 119–139. John Benjamins.

Rejzek, Jiří. 2015. *Český etymologický slovník*. Leda.

Sidhu, David M., Gabriella Vigliocco & Penny M. Pexman. 2020. Effects of iconicity in lexical decision. *Language and Cognition*. Cambridge University Press 12(1). 164–181. <https://doi.org/10/ggnkmv>.

## Morphosyntactic properties of ideophones of Central and Northeast Asia

Vít Ulman (Palacký University Olomouc)

This paper presents an overview of morphosyntactic properties of ideophones in multiple languages of Central and Northeast Asia, such as Mongolian, Sibe, Sakha, Udihe, Korea or Japanese, and differences between them. The main aim of this paper is to show how the ideophones of the individual languages can be clustered based on their morphosyntax. First, important features shared among the ideophones in this area will be listed. Then features that distinguish the ideophones of individual languages will be discussed in detail. Finally, a proposal for grouping of languages according to these features will be presented.

While it can be said that ideophones in the aforementioned languages share many basic features that are very common in ideophones across the whole world, such as is their expressiveness (Dingemans 2012), morphological markedness (in this case represented by uncommon sequences of phonemes on the level of phonotactics, or for this paper more importantly by being frequently reduplicated in languages that otherwise do not readily employ reduplication), certain level of iconicity (specific to the language, Dingemans 2019), etc., there are clear differences between otherwise typologically very similar languages that have been in contact for many centuries. A good example is Khalkha Mongolian (Mongolic) and Sibe (Tungusic). While they both are left-branching agglutinative SOV languages with layers of shared vocabulary, their ideophones clearly differ from each other.

Mongolian ideophones have complex morphology, they contain ideophonic roots that cannot function on their own and must therefore be followed by various suffixes. They also readily form not only analytic forms (as would be expected from other languages of the area, such as Japanese), but also deideophonic verbs, as in (1). On the other hand, in Sibe the morphology employed in ideophone formation is rather limited. Sibe ideophones are mostly quite morphologically simple, either reduplicated or non-reduplicated, and verbs are usually not derived from them (Zikmundová 2002), see (2).

- (1) *saxal čī-n'*      *arz-ai-čixa-ŋ*,      *xus!*  
 Beard 2SG-GEN      **rough-VBLZ-INTS-PST** shave.IMP  
 'Your beard is sticking out, shave (it)!' (Oberfalzerová 2010: 20)
- (2) *erk eim-maq*      *palang palang soxt-xei*  
 alcohol drink-CVB.PRF **drunk**      get.drunk-PST  
 'Having drunk alcohol, he got totally wasted.' (Zikmundová, personal communication: 2023)

This is not to say that there are no reduplicated Mongolian ideophones (interestingly in Mongolian even the reduplicated forms contain derivational morphology), or that there are no verbs in Sibe that were derived from ideophones, but the tendencies in these languages are rather clear.

The aforementioned contrasting features make ideophones of Sibe, and other Tungusic languages (Kilen, Udihe, Evenki) that will be discussed in the paper much more akin to the ideophones of Korean (Ferklová 2023) or Japanese (Akita 2016), whereas other languages of the area, such as Sakha (Turkic) (Zhirikova 2020) exhibit features much more similar to Mongolian.

## Bibliography

- Akita, Kimi. 2016. Grammatical and functional properties of mimetics in Japanese. In: Iwasaki, Noriko; Sells, Peter & Akita, Kimi (eds.) *The Grammar of Japanese Mimetics*. London & New York: Routledge. 20-34.
- Dingemanse, Mark. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics compass* 6/10: 654-672.
- Dingemanse, Mark. 2019. "Ideophone" as a comparative concept. In Akita, Kimi & Pardeshi, Prashant (eds.) *Ideophones, Mimetics, Expressives*. Iconicity in Language and Literature 16. Amsterdam: John Benjamins. 13–33.
- Ferklová, Blanka. 2023. Textual distribution of Korean onomatopoeias and ideophones and its specific features. *SKASE Journal of Theoretical Linguistics* 20/1.
- Oberlfalzerová, Alena. 2010. The use of iconopoeic words in spoken Mongolian. *Mongolo-Tibetica Pragensia* 3/1: 7–34.
- Ulman, Vít. 2024. Morphosyntactic Properties of Ideophones in the Languages of Central and Northeast Asia. *International Journal of Eurasian Linguistics*, 6/1: 145-171.
- Zikmundová, Veronika. 2002. 'Iconopoeia' in Mongolian and Manchu. *Mongolica Pragensia 02, Ethnolinguistics and Sociolinguistics in Synchrony and Diachrony*: 129-149.
- Zhirkova, Eugenia E. 2020. Structural features of figurative and onomatopoeic vocabulary of the Yakut and Japanese language. *Teoreticheskaya i prikladnaya lingvistika* 6/4: 51-60.

## The importance of Sakha ‘ideophones’ for the general theory of ideophones

Jan Židek, Jonáš Vlasák, Aldana Vlasáková (Charles University)

We argue that ideophones across languages share family resemblances (“*Familienähnlichkeiten*”) rather than being a monolithic class (i.e. having self-evident defining features the researcher can always approach with the same set of tools getting uniform results; or, less exactly, having a prototype), for which Sakha ideophones serve as prime example.

Sakha iconic words tend to behave neatly in relation to POS, there are onomatopoeia, ideophonic verbs, and a small closed class of prototypical ideophones.

Sakha ideophonic **verb** are traditionally considered an onomatopoeia-like class (звучоизобразительные глаголы) in the Sakha philological tradition (see e.g. [1]). Nevertheless, they do share various properties with ideophones rather than onomatopoeia, e.g. (1). They are unusual in that they constitute a subclass of verbs rather than being a cross-POS class (see e.g. [4], [5]). We intend to discuss their properties, and in doing so show that they are indeed best studied in the ideophones reference frame. We further our argument by contrasting similar roots in contact languages that do have “prototypical” ideophones (1b).

(1)

a. Sakha	b. Mongolian	c. Czech	d. Sakha
<i>sar-dyrghaa-</i>	<i>sar sar hiih</i>	<i>šust-it</i>	<i>bax gyn-na</i>
<b>IDEO-VCL</b>	IDEO IDEO do	ONOM-INF	<b>ONOM do-PST</b>
approx. ‘to scuff/scrape, to shatter’	approx. ‘to scuff/scrape, to abruptly rustle’	‘to rustle’	‘to thud’

Please note that in (1a) and (1b), the semantics are to be interpreted as ‘to move in some manner associated with a sensory input’, not as ‘to make a sound associated with a certain movement’, as in (1c) and (1d).

If we look at “ideophonic verbs” across languages, we usually get the Mongolian pattern (1b), like in Japanese (for their discussion see e.g. [2],[3]) or Korean. Cross-linguistically, ideophonic roots are usually promiscuous in relation to POS, however, in Mongolian, the *sar* ideophone (1b) seems to be restricted to verbal use cases, in which it is similar to Sakha. Also note the similar form, which may be the result of contact. Some other parallels exist with other languages in restricted contexts. We compare the behaviour in those contexts with respect to categoriality, morphosyntax, semantics and pragmatics (as applicable). The scope of the paper does not allow to discuss all aspects of iconicity present in the Sakha language, like reduplication.

## Bibliography

- [1] Afanasiev, L. A. 1992. *Fonosemantika obraznyh slov jakutskogo jazyka: avtoreferat* [Phonosemantics of the depicting words in the Yakut language: thesis account]. Yakutsk. YaGUMKA
- [2] Akita, K. 2017. The linguistic integration of Japanese ideophones and its typological implications. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 62(2), 314-334.
- [3] Akita, K. 2017. The “dynamic” nature of stative mimetic verbs in Japanese. *Journal of Cognitive Linguistics*, 2, 39-56.
- [4] Vlasák, J. 2024. Sakha. in *Onomatopoeia in the World's Languages: A Comparative Handbook*. Eds. Körtvélyessy, L. & Štekauer, P. Berlin, Boston. De Gruyter Mouton. pp. 627-636.
- [5] Vlasák, J. & Vlasáková, A. & Židek, J. In print. From intimate to insulting: Ideophonic representations of "chubby" in Sakha. in *Mongolica Pragensia '20*. Eds. Zikmundová, V., Kapišovská, V. Prague. Charles University Press.

## List of abbreviations

IDEO ideophone  
ONOM onomatopoeion  
VCL verbal class

## Mongolian ideophones: Semantics and change

Veronika Zikmundová, Jan Křivan, Vít Ulman, Veronika Kapišovská, Alena Oberfalzerová  
(Charles University)

In this paper we present the preliminary results of a project focusing on Mongolian ideophones (“an open lexical class of marked words that depict sensory imagery“, Dingemanse 2019: 16). First, we introduce the peculiarities of the semantics and morphosyntax of Mongolian ideophones in a comparative perspective (following Oberfalzerová 2010), and second, we focus on changes in the ideophone usage in the context of the ongoing shift of the speakers from rural to urban environment. We draw upon an emerging database of ideophones containing examples from both spoken and written registers. It is based on fieldwork by A. Oberfalzerová (between 2000-2010) and on a multi-sited fieldwork (between 2022-23) in rural and urban areas (30+ speakers of different background and age interviewed for a set of ca. 100 ideophones).

Areally, Mongolian clusters in general with other languages of Central and Northeast Asia. However, unlike most of the languages of this area, Mongolian has a rather complex morphological apparatus used specifically with ideophones. Mongolian ideophonic words consist of an ideophonic root with distinct expressive semantics, and suffixes which derive (i) ideophones *per se* (forming an analytic construction with quotative verbs), (ii) ideophonic nouns, adjectives, verbs. Stems of type (ii) take regular derivational and inflectional suffixes. All aforementioned morphological means participate in the resulting expressive-descriptive meaning (1).

- (1) *mord-ood*                      *cüinde-gn-uul-chih-lee*  
Get.on.a.horse-CVB      round.shape(of.a.belly)-REPEATED.MOVEMENT-CAUS-INTENSITY-PST  
‘[She] got on a horse and rode, featuring her round-shaped belly.’  
(literally: started letting a round belly shape being seen as repeatedly moving)

Semantics of Mongolian ideophonic words is thus compositional, showcasing a notable distinction between Mongolian and other regional languages like Sibe/Manchu, in which ideophones are mostly reduplicated adverbial expressions which cannot be semantically decomposed (2).

- (2) *bumber~bamber*  
‘Walking of an old and weak person.’

We also show that Mongolian ideophonic semantics is particularly elaborate in the following domains (using classification of sensory domains in Winter et al. 2017, cf. Dingemanse & Majid 2012; McLean 2021): (a) visual perception of objects (e.g. quantity, array) on the horizon, (b) visual perception of objects in the landscape (vegetation, stone fields, mountains, water bodies), (c) tactile perception of object surfaces (e.g. roughness, smoothness, undulation), (d) visual perception of objects in motion (riders, herds, birds, wild animals).

Concerning current usage, ideophones occur in several speech domains/registers, which differ between the rural and urban environment. In the former, ideophones are mainly a property of everyday language. In the latter, several developments have been revealed: (i) Use of shifted and metaphorical meanings in new domains with concurrent loss of use in the primary meanings. (ii) Usage in scientific terminology accompanied by loss of expressivity. (iii) Changes caused by recent sociolinguistic developments, such as reinforcement of certain ideophones through instruction at schools (particularly ideophones occurring in folklore and literature) and their spread via media.

We can conclude that the repertoire of Mongolian ideophones reflects to a considerable extent the environment speakers live in, and it is being modified in accordance with recent social changes.

## Bibliography

- Dingemanse, Mark & Majid, Asifa. (2012). The semantic structure of sensory vocabulary in an African language. In: Miyake, N. (ed.), *34th Annual Meeting of the Cognitive Science Society (CogSci 2012)*, 300-305.
- Dingemanse, M. (2019). Ideophone as a comparative concept. In: Akita, K. & Pardeshi, P., *Iconicity in Language and Literature*, vol. 16, 13–33. Amsterdam: John Benjamins.
- Oberfalzerová, A. (2010). The use of iconopoeic words in spoken Mongolian. *Mongolo-Tibetica Pragensia* 3(1), 7–34.
- McLean, Bonnie. 2021. Revising an implicational hierarchy for the meanings of ideophones, with special reference to Japonic. *Linguistic Typology* 25(3): 507-549.
- Winter, Bodo, Perlman, M., Perry, L. K. & Lupyan, Gary. (2017). Which words are most iconic? Interaction and Iconicity in the Evolution of Language. *Interaction Studies* 18(3), 443–464. DOI: 10.1075/is.18.3.07win



11

*Nové empiricky podložené  
pohledy na mnohojazyčnost*

*New empirically based  
perspectives  
on multilingualism*

organizují / organizers

Barbara Mertins

Kateřina Šormová

(CZE, ENG)

Bienále české lingvistiky 2024  
Biennial of Czech Linguistics 2024

## Popis panelu / Workshop description

**(CZE)** Z globální perspektivy jsou dvojjazyčnost a vícejazyčnost spíše pravidlem než výjimkou; z evropského a českého pohledu ale není vícejazyčnost nutně vnímána jako běžné jazykové prostředí a často je poměřována s existující jednojazyčnou normou. (De Houwer, 2009; Delucchi Danhier & Mertins, 2018; Bayram, Miller, Rothman, & Serratrice, 2018; Grosjean, 2021). Nicméně pro mnoho mluvčích je být dvoj-/vícejazyčným mluvčím spíše přirozený stav než záměrné/vědomé rozhodnutí (De Houwer, 2009). Panel zdůrazňuje význam empirických dat tím, že zahrnuje různé metodologické přístupy (rozsáhlé průzkumy, reakční časy, eye-tracking, preference otáčení hlavy u kojenců). Výzkumná témata zahrnují jevy na různých jazykových úrovních (fonologické, gramatické, konceptuální i pragmatické), které jsou zkoumány z různých teoretických hledisek (osvojování, jazyková relativita, usage-based přístupy, jazyková atrice, didaktika jazyka/aplikovaná lingvistika). Prezentovaná lingvistická a psycholingvistická data se zaměřují na češtinu, ale zahrnují také angličtinu, perštinu, němčinu, norštinu, polštinu, ruštinu a turečtinu v různých dvojjazyčných párech a prostředích. Celkový cíl panelu je dvojitý: Za prvé poskytnout empiricky podložené poznatky o vícejazyčném zpracování na různých kognitivních a lingvistických úrovních na základě dat z méně studovaných jazyků, za druhé přispět k současnému diskurzu o mnohojazyčnosti ve vědeckém, didaktickém i aplikovaném kontextu a poskytnout tak nové podněty a inspiraci při studiu mnohojazyčnosti.

**(ENG)** From a global perspective, bilingualism and multilingualism constitute a rule rather than an exception; from a European and Czech perspective, multilingualism is not necessarily seen as a common linguistic setting and is often measured against an existing monolingual norm. (De Houwer, 2009; Delucchi Danhier & Mertins, 2018; Bayram, Miller, Rothman, & Serratrice, 2018; Grosjean, 2021). Nevertheless, for many speakers, being bilingual/multilingual is a natural state rather than an intentional/conscious decision (De Houwer, 2009). The panel highlights the relevance of empirical data by including different methodological approaches (large-scale surveys, reaction time data, eye tracking data, head turn preference paradigm). The research topics comprise phenomena at different linguistic levels (phonological, grammatical, conceptual as well as pragmatic) that are examined from various theoretical viewpoints (acquisition, linguistic relativity, usage-based approaches, language attrition, didactics/applied linguistics). The linguistic and psycholinguistic data presented focus on Czech, but include also English, Farsi, German, Norwegian, Polish, Russian, and Turkish in different bilingual pairs and settings. The overall goal of the panel is twofold: First, to provide empirically based insights into multilingual processing at different cognitive and linguistic levels based on data from less-studied languages. Second, to contribute to the current discourse on multilingualism in scientific, didactic as well as applied contexts and, thus, to give new impulses and inspiration in the study of multilingualism.

## References

Bayram, F., Miller, D., Rothman, J., & Serratrice, L. (2018). Studies in bilingualism: 25 years in the making. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Hrsg.), *Studies in Bilingualism* (Bd. 54, S. 1–12). Amsterdam: John Benjamins Publishing Company.

De Houwer, A. (2009). *Bilingual First Language Acquisition*. Multilingual Matters.

Delucchi Danhier, R., & Mertins, B. (2018). Psycholinguistische Grundlagen der Inklusion: Schwerpunkt Bilingualismus. In S. Hußmann & B. Welzel (Hrsg.), *DoProfiL – Das Dortmunder Profil für inklusionsorientierte Lehrerinnen- und Lehrerbildung* (S. 161–178). Münster: Waxmann.

Grosjean, F. (2021). *Life as a bilingual: Knowing and using two or more languages*. Cambridge: Cambridge University Press.

## Obsah / Table of Contents

*Word segmentation by multilingual infants: Challenges and new perspectives*

Joël Alipaß (TU Dortmund University)

*Looking under the surface. The investigation of bilingual children's conceptualization with the use of linguistic, eye-tracking, and memory data*

Hanna Andresen (MultiLing, University of Oslo)

*On the project Educating Heritage Language Learners*

Marie Boccou Kestránková (Charles University)

*Language specific-patterns in a multilingual context: The case of grammatical gender*

Maryam Fatemi (TU Dortmund University)

*Podpora vícejazyčnosti v Ústeckém kraji. Pilotní studie ve školách preprimárního a primárního vzdělávání*

Martin Lachout (Filozofická fakulta Univerzity J. E. Purkyně, Ústí nad Labem)

*The notion of 'being bilingual': A large-scale survey*

Barbara Mertins (TU Dortmund University)

*Advantages of bilingualism due to an earlier developed metalinguistic awareness*

Katrin Odermann, Barbara Mertins (TU Dortmund University)

*Plurilingual study-abroad students' perspective on language interaction in their linguistic repertoires*

Silvie Převrátilová (Charles University)

*Language and cognition in dementia. A call for a more detailed assessment of bilingualism*

Aldona Rzitki (Institut für Slavische Sprachen und Literaturen, Universität Bern)

*Psaná čeština romských žáků – etnolektní projevy*

Kateřina Šormová (FF UK)

## Word Segmentation by Multilingual Infants: Challenges and New Perspectives

Joël Alipaß (TU Dortmund University)

What should a sound empirical basis of infant multilingualism encompass? One well-attested skill of preverbal infants is certainly word segmentation: the ability to segment lexical units from running speech, despite the absence of clear perceptible boundaries between lexical units. Besides using statistical probabilities of syllable transitions (Saffran et al., 1996), one major strategy for the segmentation of speech has been found to lie in prosody, starting with the seminal studies by Jusczyk and Aslin (1995) and Jusczyk, Houston and Newsome (1999), who showed that first syllable (trochaic) lexical stress in English nouns helps 7,5-month-old infants to extract syllables from speech. Prosodic segmentation has been replicated for rhythmically similar (Dutch: Houston, Jusczyk, Kuijpers, Coolen, Cutler, 2000; German: Bartels, Darcy, Höhle, 2009, *inter alia*) and different languages (e.g. French, Polka & Sundara, 2012; Portuguese, Butler & Frota, 2018). While this type of segmentation ability seems to be connected to later language development (e.g., Newman, Ratner, Jusczyk, Jusczyk & Dow, 2006, Marimon, Höhle & Langus, 2022), the number of languages studied so far remains small. In addition to this, research with bilingual infants has only recently been initiated (since Bosch, Figueras, Teixidó & Ramon-Casas, 2013). Bilingual infants, it has been shown, have a greater ability to deal with the statistics of two different artificial languages (Antovich & Graf Estes, 2018), successfully use strategies from two prosodically different languages (English & French, Orena & Polka, 2019) and even exploit prosodic variability in one language to segment rare prosodic patterns in their other language earlier than monolinguals (English & Spanish, Mateu & Sundara, 2022).

At the same time, including into the much needed language diversity the acquisition of Czech may shed new light on the question of prosodic segmentation strategies and the importance of syllable “prominence”: Despite the consensus about Czech as a trochaic language with fixed stress on the first syllable of a lexical item (Dogil, Gvozdanović & Kodzasov, 1999), the typical correlates signaling syllable prominence (length, loudness, pitch height, vowel quality) appear to be hardly prominent with the initial syllable in Czech (Skarnitzl & Eriksson, 2017).

This talk discusses the challenges and outlooks of this complexity. First, it critically reviews past research designs on bilingual word segmentation (Bosch et al., 2013; Polka et al., 2017; Orena & Polka, 2019; Mateu & Sundara, 2022) and its consequences for the interpretation of results. This is complemented by preliminary data from a Turkish-German bilingual experiment, testing familiarized words using the *head turn preference* procedure, by which methodological implications are discussed. In a second step, it will argue for the relevance of studying Czech infant word segmentation and bilingual constellations with Czech, based on the specifics mentioned above.

Finally, two further perspectives will be laid out: (1) Locating word segmentation studies with bilingual infants into the larger discussion about familiarity vs. novelty preferences (are familiarized stimuli looked at for a longer time or vice versa?). This will be done with reference to Singh’s (2021) observation that there might be a bilingualism-general orientation towards novel information and maturity by diversified linguistic input (Braun et al., 2021). (2) The potential of EEG-based studies on bilingual word segmentation in infancy will be highlighted – the rapid response in time and the breadth of neural activity in the ERP signal, which has been related input-related language dominance in bilinguals (Unger, 2023). The potential of EEG will be discussed with regard to Czech mono- and bilingual language acquisition, pointing at the possibility to check ERP correlations with the particular segmental (e.g. sound contrasting vowel length) and supra-segmental (e.g., post-stress  $f_0$  rise) features occurring in natural speech.

## Bibliography

- Antovich, D. M., & Graf Estes, K. (2018). Learning across languages: Bilingual experience supports dual language statistical word segmentation. *Developmental Science*, 21(2), e12548.
- Bartels, S., Darcy, I., & Höhle, B. (2009). Schwa syllables facilitate word segmentation for 9-month-old German-learning infants. In J. Chandless, M. Franchini, S. Lord, & G.-M. Rheiner (Ed.), *Proceedings of the 33rd Annual Boston University Conference on Language Development* (S. 73–84). Cascadilla Press.
- Bosch, L., Figueras, M., Teixidó, M., & Ramon-Casas, M. (2013). Rapid gains in segmenting fluent speech when words match the rhythmic unit: Evidence from infants acquiring syllable-timed languages. *Frontiers in Psychology*, 4.
- Braun, B., Czeke, N., Rimpler, J., Zinn, C., Probst, J., Goldlücke, B., Kretschmer, J., & Zahner-Ritter, K. (2021). Remote Testing of the Familiar Word Effect With Non-dialectal and Dialectal German-Learning 1–2-Year-Olds. *Frontiers in Psychology*, 12, 714363.
- Butler, J., & Frota, S. (2018). Emerging word segmentation abilities in European Portuguese-learning infants: New evidence for the rhythmic unit and the edge factor. *Journal of Child Language*, 45(6), 1294–1308.
- Dogil, G., Gvozdanović, J., & Kodzasov, S. (1999). Slavic languages. In H. V. D. Hulst (Ed.), *Eurotyp* (1999, p. 813–876). Mouton de Gruyter.
- Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A. (2000). Cross-language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review*, 7(3), 504–509.
- Jusczyk, P. W., & Aslin, R. N. (1995). Infants' Detection of the Sound Patterns of Words in Fluent Speech. *Cognitive Psychology*, 29(1), 1–23.
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology*, 39(3–4), 159–207.
- Marimon, M., Höhle, B., & Langus, A. (2022). Pupillary entrainment reveals individual differences in cue weighting in 9-month-old German-learning infants. *Cognition*, 224, 105054.
- Mateu, V., & Sundara, M. (2022). Spanish input accelerates bilingual infants' segmentation of English words. *Cognition*, 218, 104936.
- Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*, 42(4), 643–655.
- Orena, A. J., & Polka, L. (2019). Monolingual and bilingual infants' word segmentation abilities in an inter-mixed dual-language task. *Infancy*, 24(5), 718–737.
- Polka, L., Orena, A. J., Sundara, M., & Worrall, J. (2017). Segmenting words from fluent speech during infancy—Challenges and opportunities in a bilingual context. *Developmental Science*, 20(1), e12419.
- Polka, L., & Sundara, M. (2012). Word Segmentation in Monolingual Infants Acquiring Canadian English and Canadian French: Native Language, Cross-Dialect, and Cross-Language Comparisons. *Infancy*, 17(2), 198–232.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, 274(5294), 1926–1928.
- Singh, L. (2021). Evidence for an Early Novelty Orientation in Bilingual Learners. *Child Development Perspectives*, 15(2), 110–116.
- Skarnitzl, R., & Eriksson, A. (2017). The Acoustics of Word Stress in Czech as a Function of Speaking Style. In Francisco Lacerda, David House, Mattias Heldner, Joakim Gustafson, Sofia Strömbergsson, & Marcin Włodarczak (Ed.), *Proc. Interspeech 2017* (p. 3221–3225).
- Unger, A. (2023). *Frühe Wortsegmentierung im monolingualen und bilingualen Spracherwerb: Behaviorale und neurophysiologische Untersuchungen mit 9 Monate alten Säuglingen* [Universität Potsdam]. <https://publishup.uni-potsdam.de/58297>

**Looking under the surface. The investigation of bilingual children's conceptualization with the use of linguistic, eye-tracking, and memory data**

Hanna Andresen (MultiLing, University of Oslo)

In this talk, I present a study on how simultaneous bilingual Norwegian-English children conceptualize goal-oriented motion events in their two languages, which have different, and partly conflicting, language-specific conceptualization patterns. While speakers of the aspect-language English often prefer a so-called phasal perspective on events, with a focus on the ongoing action, speakers of Norwegian (a non-aspect-language) prefer a so-called holistic, endpoint-oriented perspective, with more attention to potential endpoints (Andresen, 2023; von Stutterheim & Nüse, 2003). But what about young, bilingual speakers who grow up acquiring both perspectives simultaneously?

A group of bilingual Norwegian-English children (*Mage* 8) participated in an experiment, which combined three different methodologies to measure attention to endpoints: linguistic elicitation, eye-tracking, and memory tests, and the children were tested twice, on a separate day for each language. The results showed an effect of the language of operation on conceptualization, independently of language dominance, but depending on the situation. In their first session, the bilingual children showed a strong awareness of the prototypical conceptualization pattern in the language they were speaking. In their second session, the children's conceptualization was influenced by their first. Hence, we can talk about a flexible conceptual dominance linked to language, the situation, and to previous experience (cf., Andresen, Danhier Delucchi, & Mertins, 2024).

In this presentation, I discuss what impact factors such as language, memory, linguistic proficiency, as well as linguistic confidence and/or anxiety have on the level of conceptual awareness and/or dominance, suggesting that interdisciplinary research at the intersection of socio- and psycholinguistics is an important venue for future research within the field of bi- and multilingualism. Furthermore, I show the significance of using methodologies such as eye-tracking in addition to the linguistic data – to be able to “look under the surface” – when investigating bilingual children with different language combinations (for the future, perhaps including Czech?). The combination of different methodologies leads to new insights in our understanding of early bilingualism at a conceptual level, with implications for the field, but also for parents, care givers, and stakeholders, who at times need to be reminded about children's unique capacity for language learning.

## Bibliography

- Andresen, H. (2023). With Norwegian eyes. The verbalization of goal-oriented motion events in Norwegian. In B. Mertins & R. Delucchi (Eds.), *Diversity in cognition*. Cognition: Kognitions- und neurowissenschaftliche Beiträge zur natürlichen Sprachverarbeitung (pp. 147–181). Peter Lang Verlag.
- Andresen, H., Danhier Delucchi, R., and Mertins, B. (2024.) Dominant-while-speaking: How bilingual Norwegian–English children conceptualize goal-oriented motion events. *International Journal of Bilingualism*, 0(0). <https://doi.org/10.1177/13670069241251781>
- von Stutterheim, C., & Nüse, R. (2003). Processes of conceptualization in language production: Language-specific perspectives and event construal. *Linguistics* 41(5), pp. 831–881.



## On the project Educating Heritage Language Learners

Marie Boccou Kestřánková (Charles University)

The paper presents the 4EU+ mini-grants of 2022 and 2023, two small interconnected projects, referred to as a project in the abstract. The project is led by experts from Charles University and is aimed at streamlining and systematising of teaching of heritage languages. It also involves experts from top universities in France, Poland, Italy, Germany, Denmark, Switzerland and Iceland.

The thematic area of the project is Educating Heritage Language Learners in Europe (on Heritage language education see e.g. Valdés, G. 2005, Montrul 2015), which raises several issues in contemporary education in plurilingual Europe. The main theme explored in the project is the paradox facing European policy, namely, the guarantee of free and compulsory primary education, which should be accessible to all and should be implemented without discrimination (The Universal Declaration of Human Rights 1948, Charter of Fundamental Rights of the Council of Europe 2000, Beacco, Byram 2007). On the other hand, there is a real commitment of European countries in the field of migrants' languages, in the education of migrants and in the effective development of their plurilingual profile. This paradox is present even though "Education shall be directed to the full development of the human personality and the strengthening of respect for human rights" (The Universal Declaration of Human Rights 1948) and plurilingualism and linguistic diversity are the goals of the EU (Charter of Fundamental Rights of the Council of Europe 2000).

Many other issues are closely related to the above. For example, our team points to the measurement of heritage language learners' discourse with monolingually educated speakers, which is often the basis for making reference in heritage language learners' education (De Houwer 2009, Schader 2015, Grosjean 2021). Concerning this issue, our project researchers have been conducting pilot studies to demonstrate the need for a specific approach to teaching heritage language learners and to show that teachers are not receiving necessary methodological support. Last but not least, the project highlights the lack of interdisciplinary research that a) provides the basis for a thoughtful systematic improvement of the platform for the education of heritage language users and b) is the basis for functional inclusive education (Delucchi Danhier, Mertins 2018). This is closely linked to the basic principle of our team, which is international and includes professionals with different expertise related to our topic. One of the mottos of our project is therefore: „International and interdisciplinary cooperation is essential to promote the inclusion of HL education in European educational structures.“

The work is dependent on obtaining international grant support, as in this way it is possible to set international goals and secure interdisciplinary experts from different countries (e.g. experts in law, pedagogy, linguistics). Therefore, a long-term strategy for the development of team activities was planned, of which two projects have been implemented so far and a follow-up project has now been applied for. Interdisciplinarity and international cooperation are essential for all steps of the strategy.

In the presentation, we will first introduce the general objective of the project and explain where the idea of the project came from. Then we will present the objectives of the project, which coincide with selected EU objectives. In this context, we will point out the facts that prevent the full and functional fulfilment of the mentioned European objectives in practice. One of the aims is to find answers to the below questions. The first question concerns school systems in Europe: How can current educational systems develop heritage language acquired outside schools? As the group of heritage language users is small compared to the users of the language of instruction, individual European countries consider this issue to be of lesser importance. For this reason, the team's goal is to find answers to the second question: Are European education systems causing an undesirable consequence: to "forget" heritage language and re-learn it as a foreign language? The presentation briefly describes the steps taken by the team, and also the difficulties the team faces in meeting the goals and in setting subsequent goals. The final phase of the contribution will be devoted to the description of the planned steps of the

team looking for solutions to make plurilingualism and linguistic diversity in contemporary Europe more effective.

## **Bibliography**

Beacco J.-C., Byram M. (2007). From linguistic diversity to plurilingual education: Guide for the development of language education policies in Europe. DOI <https://www.coe.int/en/web/language-policy/from-linguistic-diversity-to-plurilingual-education-guide-for-the-development-of-language-education-policies-in-europe>

(2000). Charter of Fundamental Rights of the Council of Europe. DOI [https://www.europarl.europa.eu/charter/pdf/text\\_en.pdf](https://www.europarl.europa.eu/charter/pdf/text_en.pdf)

De Houwer, A. (2009). Bilingual First Language Acquisition. *Multilingual Matters*.

Garcia, V., Pineault, C., Bryfonski, L. (2023). An evaluation of a multidimensional identity measurement instrument: The heritage language speaker identity tool (HLS-IT). *The Modern Language Journal*, 107(1), p. 353–372.

Grosjean, F. (2021). *Life as a bilingual: Knowing and using two or more languages*. Cambridge: Cambridge University Press.

Delucchi Danhier, R., & Mertins, B. (2018). Psycholinguistische Grundlagen der Inklusion: Schwerpunkt Bilingualismus. In S. Hußmann & B. Welzel (Hrsg.), *DoProfiL – Das Dortmunder Profil für inklusionsorientierte Lehrerinnen- und Lehrerbildung* (S. 161–178). Münster: Waxmann.

Montrul, S. (2015). *The Acquisition of Heritage Languages*. Cambridge. Cambridge University Press. 384 p. DOI <https://doi.org/10.1017/CBO9781139030502>

(1948). The Universal Declaration of Human Rights. DOI <https://www.ohchr.org/en/human-rights/universal-declaration/translations/english>

Schader, B. (2015). Principes et contextes. Matériel pour l'enseignement de la langue et de la culture d'origine. Centre IPE (International Projects in Education) de la Haute école pédagogique de Zurich. University of Zürich. p. 184.

Valdés, G. (2005): Bilingualism, heritage language learners, and SLA research: Opportunities lost or seized? *The Modern Language Journal*, 89(3), p. 410–426.

## Language Specific-Patterns in a Multilingual Context: The Case of Grammatical Gender

Maryam Fatemi (TU Dortmund University)

In communication between native speakers and L2 speakers, the topic of non-nativeness often emerges, even among advanced L2 speakers who produce grammatically error-free sentences (Carroll & Stutterheim 2003). While some scholars have emphasized the presence of non-nativeness primarily in the production of complex linguistic constructions (Schmiedtova, Stutterheim, & Carroll 2011), this study posits that this phenomenon may also manifest in non-complex linguistic structures. This observation stems from the influence of language-specific patterns on the conceptualization process of L2 speakers, resulting in different encoding of the same phenomena by speakers of different languages. This idea aligns with the linguistic relativity hypothesis, also known as the Sapir-Whorf hypothesis (Lucy 2016, Whorf 1956). This phenomenon arises from the absence of certain grammatical features in L2 speakers' native language, or their presence in a different form, leading L2 speakers to formulate sentences in their L2 influenced by specific patterns from their dominant native language. While this process is natural for L2 speakers, it often results in non-native-like linguistic formulations.

Grammatical gender represents one of the linguistic features unique to specific languages. This linguistic feature can manifest in various forms across languages, with some languages possessing two or more grammatical genders. While this categorization often aligns with real-world distinctions of sex, such alignment is frequently not observed (Corbett 1991). This obligatory grammatical feature appears in different configurations; for instance, in German, it is represented by articles “der/masculine”, “die/feminine”, and “das/neutral”, while in Polish, it takes the form of post-nominal suffixes. In Czech, which also features three grammatical genders, this attribute is further indicated by post-verbal suffixes -y or -i, particularly marking the past tense of verbs in the written form. In English, grammatical gender is only evident in pronominal form. Contrary to the mentioned languages, Farsi is another language that lacks grammatical gender completely, neither in article form nor in pronominal nor post-nominal form, or post-verbal form (Fatemi 2023).

Accepting bilingualism and multilingualism as the norm rather than the exception (Delucchi Danhier & Mertins, 2018; Grosjean, 2021) and acknowledging the prevalence of multilingual contexts in many countries including the Czech Republic, the theoretical question arises: How do bilingual speakers of different languages with distinct grammatical gender systems interact with linguistic gender complexities? To answer this question, the study presents the results of a behavioral study on monolinguals, confirming the presence of the effect of grammatical gender on cognition. This study involved 74 German and 60 Farsi speakers. Participants engaged in a property judgment task, describing objects with the first three adjectives that came to mind. Subsequently, another group of German (n=102) and Farsi (n=60) speakers evaluated these adjectives on a Likert scale ranging from 1 (very feminine), 4 (gender neutral), and 7 (very masculine). Applying a linear mixed model (Fahrmeir et al., 2013), the study uncovered distinct influences of grammatical gender on German and Farsi speakers. These findings suggest that individuals with diverse grammatical systems perceive the same entity differently. Furthermore, the study underscores the importance of examining languages with unique grammatical gender systems, emphasizing the role of such structures in eliciting non-native feelings among L2 speakers. It is especially crucial to consider the complexity of grammatical gender when learning a second language, such as English or Farsi as an L2, which possesses less complexity in this regard compared to the speaker's native language.

	<i>b</i>	<i>SE</i>	<i>df</i>	<i>t</i> -value	<i>p</i> -value
Intercept	4.085	0.141	43.114	28.886	< 0.001
GG = feminine	-0.434	0.181	29.279	-2.405	0.023
GG = neutral	-0.243	0.181	29.384	-1.346	0.189
Adjective order = second	-0.135	0.045	4718.795	-2.983	0.003
Adjective order = third	-0.114	0.046	4721.493	-2.446	0.014
Participant gender = female	-0.024	0.031	68.637	-0.761	0.449
Age (in years)	0.002	0.002	78.375	1.277	0.205
GG = feminine × Adjective order = second	0.240	0.064	4720.435	3.739	< 0.001
GG = neutral × Adjective order = second	0.162	0.065	4719.305	2.477	0.013
GG = feminine × Adjective order = third	0.145	0.065	4720.276	2.214	0.027
GG = neutral × Adjective order = third	0.136	0.067	4722.555	2.041	0.041

The linear mixed model is based on a total of 4808 observations from 74 participants. Both the participant's ID and the 30 objects to be described with an adjective were considered random intercepts. Abbreviations: *b* = estimated regression coefficient, *SE* = standard error, *df* = degrees of freedom, GG = grammatical gender

	<i>b</i>	<i>SE</i>	<i>df</i>	<i>t</i> -value	<i>p</i> -value
(Intercept)	3.797	0.166	131.850	22.867	1.014×10 <sup>-47</sup>
Adjective Order = second	-0.062	0.057	1616.307	-1.081	0.280
Adjective Order = third	-0.024	0.058	1616.989	-0.413	0.680
Participants' Gender = female	0.040	0.048	1616.399	0.834	0.404
Age	-5.624×10 <sup>-4</sup>	0.003	1616.481	-0.175	0.861

The linear mixed model is based on a total of 1648 observations from 65 Participants for 30 Objects

## Bibliography

Carroll, M. and Stutterheim, Ch. v. (2003) Typology and information organization: Perspective taking and language-specific effects in the construction of events. In A. Ramat (ed.) *Typology and Second Language Acquisition* (pp. 365–402). Berlin: Mouton de Gruyter.

Corbett, G. G. (1991). *Gender*. Cambridge: Cambridge University Press.

Delucchi Danhier, R., & Mertins, B. (2018). Psycholinguistische Grundlagen der Inklusion: Schwerpunkt Bilingualismus. In S. Hußmann & B. Welzel (Hrsg.), *DoProfil – Das Dortmunder Profil für inklusionsorientierte Lehrerinnen- und Lehrerbildung* (S. 161–178). Münster: Waxmann.

Fahrmeir, L., Kneib, T., Lang, S., Marx, B., Fahrmeir, L., Kneib, T., ... & Marx, B. (2013). *Regression models* (pp. 21-72). Springer Berlin Heidelberg.

Fatemi, M. (2023). Synchronic Overview of Grammatical Gender Marking in Indo-European Languages. In Barbara Mertins & Renate Delucchi Danhier (Eds.), *Diversity in Cognition* (pp. 13-60). Berlin: Deutschland.

Lucy, John. A. 2016. Recent Advances in the Study of Linguistic Relativity in Historical Context: A Critical Assessment. *Language Learning*, 66(3), 487-515.

Grosjean, F. (2021). *Life as a bilingual: Knowing and using two or more languages*. Cambridge: Cambridge University Press. Schmiedtová, B., Von Stutterheim, C., & Carroll, M. (2011). *Language-specific patterns in event construal of advanced second language speakers. Thinking and speaking in two languages*, 66-107. Whorf, Benjamin. 1956. *Language Thought and Reality*. MIT Press.

**Podpora vícejazyčnosti v Ústeckém kraji.  
Pilotní studie ve školách preprimárního a primárního vzdělávání**

Martin Lachout (Filozofická fakulta Univerzity J. E. Purkyně, Ústí nad Labem)

V tomto příspěvku autor představuje výzkum realizovaný na katedře germanistiky FF UJEP, v němž se snažil získat odpověď na otázku, do jaké míry je německý jazyk používán v předškolním vzdělávání (mateřských školách) a jak je vyučován na základních školách (1. až 9. třída základních škol a nižší ročníky gymnázií) ve školských zařízeních v Ústeckém kraji. Zajímala nás také otázka, zda existuje vzájemná provázanost mezi výukou německého jazyka mezi těmito vzdělávacími stupni.

Ústecký kraj se nachází v severozápadní části České republiky a přímo sousedí se Spolkovou republikou Německo, konkrétně se Svobodným státem Sasko. Vzhledem k této geografické poloze lze předpokládat, že se mnoho žáků v tomto kraji bude učit německy a že školy (případně i mateřské školy) využijí možnosti, aby se jejich žáci zdokonalovali v německém jazyce v rámci přeshraniční spolupráce s německými školskými institucemi. Formulovali jsme proto hypotézu, že bude výuka němčiny v Ústeckém kraji podporována zejména tím, že dochází k vzájemnému propojení výuky mezi mateřskými a základními školami, že žáci a děti v tomto regionu budou úzce spolupracovat se žáky a dětmi z Německa, čímž se zvýší jejich motivace pro učení se němčině, což se odrazí i na jejich známkách (v případě základních škol).

Součástí příspěvku tak bude i prezentace výsledků vzešlých z provedené studie.

### **Bibliografie**

Bhatia T. K.; Ritchie W.C. (Eds.) (2004): *The handbook of bilingualism*. Oxford: Blackwell Publishing.

Harding-Esch, E.; Riley, P. (2008): *Bilingvní rodina*. Praha: Portál.

Lachout, M. (2017): *Bilingvismus a bilingvní výchova na příkladu bilingvismu česko-německého*. Praha: Togga

Statistická ročenka školství – výkonové ukazatele:  
<https://statis.msmt.cz/rocenka/rocenka.asp/rocenka/rocenka.asp>

**The notion of ‘being bilingual’: A large-scale survey**

Barbara Mertins (TU Dortmund University)

From a global perspective, bilingualism and multilingualism constitute a rule rather than an exception. In the context of the Czech Republic or, the Europe context in general, bilingualism/multilingualism is not necessarily seen as a common, wide-spread linguistic setting. Nevertheless, for many speakers, being bilingual/multilingual is a natural state, in which they are born in, rather than a conscious choice (De Houwer, 2009). One of the challenges in the research of multilingualism is how to define/determine how is a bilingual person. In a number of studies, *being bilingual* is seen in terms of the age of acquisition (AoA). According to this approach, only speakers who have been raised with more than one language before/from birth on are considered bilingual (cf. Houwer, 2009; Armon-Lotem, Rose & Altman, 2021). At the same time, other studies have provided evidence that language proficiency, independent of the AoA, achieved in the languages at stake is the most relevant factor for passing as bilingual (Grosjean 1996; 2021; Bongaerts, 1997; Perani et al., 1998; Consoni et al, 2013;). In the majority of papers found in (psycho)linguistics and other relevant disciplines, however, the definition of bilingualism/multilingualism is either vague or not specified at all (Delucchi Danhier & Mertins, 2018). The challenge seems to be how to assess of the degree of bilingualism/ multilingualism and the methods used to this end that reach from standardized grammatical tests to highly controversial self-assessment tools (cf. Mertins, 2023). In this paper, I will present the results and the underlying statistical procedure of a large-scale survey aiming at a more objective assessment of the degree of bilingual/multilingual in bilingual children (N=100) and adults (N=100), including a large group of bilingual speakers with Czech as one of their languages. The findings show on the one hand the great heterogeneity of the bilingual community, on the other they highlight the importance of language input in the acquisition of two or more languages and the relative unimportance of the AoA. In conclusion, the results will be discussed within the general study of multilingualism and also in terms of the so-called bilingual advantages (De Houwer, 2009; Delucchi Danhier & Mertins, 2018; Bayram et al., 2018; Grosjean, 2021).

## Bibliography

- Armon-Lotem, S., Rose, K., & Altman, C. (2021). The development of English as a heritage language: The role of chronological age and age of onset of bilingualism. *First Language*, 41(1), 67-89.
- Bayram, F., Miller, D., Rothman, J., & Serratrice, L. (2018). Studies in bilingualism: 25 years in the making. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Hrsg.), *Studies in Bilingualism* (Bd. 54, S. 1–12). Amsterdam: John Benjamins Publishing Company.
- Bongaerts, T., van Summeren, C., Planken, B. & Schils, E. (1997). Age and Ultimate Attainment in the Pronunciation of a Foreign Language. *Studies in Second Language Acquisition*, 19 (4), 447–465.
- Consonni, M., Cafiero, R., Marin, D., Tettamanti, M., Iadanza, A., Fabbro, F. & Perani, D. (2013). Neural convergence for language comprehension and grammatical class production in highly proficient bilinguals is independent of age of acquisition. *Cortex*, 49 (5), 1252– 1258.
- Delucchi Danhier, R., & Mertins, B. (2018). Psycholinguistische Grundlagen der Inklusion: Schwerpunkt Bilingualismus. In S. Hußmann & B. Welzel (Hrsg.), *DoProfiL – Das Dortmunder Profil für inklusionsorientierte Lehrerinnen- und Lehrerbildung* (S. 161–178). Münster: Waxmann.
- Grosjean, F. (2021). *Life as a bilingual: Knowing and using two or more languages*. Cambridge: Cambridge University Press.
- Grosjean, F. (1996). Bilingualismus und Bikulturalismus: Versuch einer Definition. In Schneider, H. & Hollenweger, J. (Hrsg.), *Mehrsprachigkeit und Fremdsprachigkeit: Arbeit für die Sonderpädagogik?* Lucerne: Edition SZH.
- De Houwer, A. (2009). *Bilingual First Language Acquisition*. Multilingual Matters.
- Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S. F., Fazio, F. & Mehler J. (1998). The bilingual brain. Proficiency and age of acquisition of the second language. *Brain*, 121 (10), 1841–1852.

**Advantages of bilingualism due to an earlier developed metalinguistic awareness**

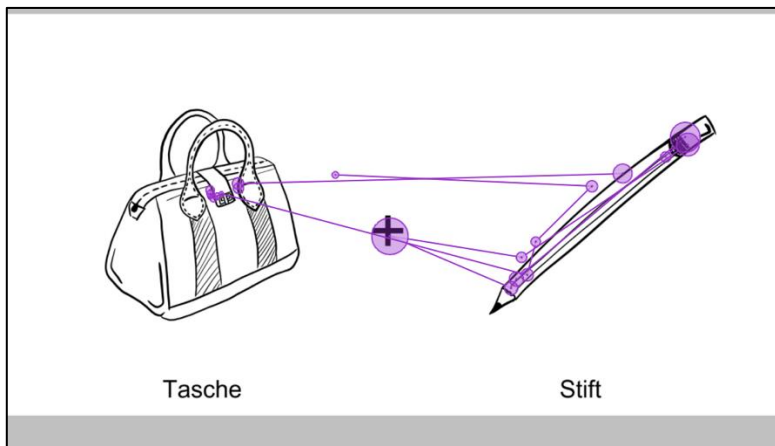
Katrin Odermann, Barbara Mertins (TU Dortmund University)

The ability to think about language at the meta-level is considered as an important prerequisite for an optimal development of written language (Schneider 2017). Literacy represents a key competence for school success. Children who are in possession of metalinguistic awareness before starting school show significant advantages in the acquisition of written language (Schneider 2017). In addition, previous research (Bialystok 1992; Reder et al. 2013) has provided evidence for an early development of metalinguistic awareness in multilingual children.

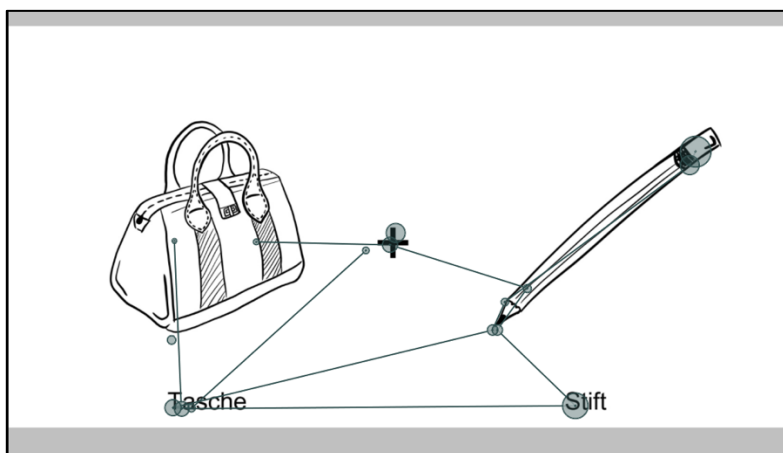
This talk presents the results of an eye-tracking study on the development of metalinguistic awareness at the phonological level measured by understanding the arbitrariness of signs. Bilingual and monolingual preschool children (N=60) aged four to six years were in the first step tested for their word comprehension in German using the PPVT test to guarantee their vocabulary in the German language. In the second step their phonological awareness was examined using an eye-tracking experiment. It was investigated whether the children are able to understand the arbitrariness of the signs by fixating the written words in addition to the pictured objects and in this way answering the questions correctly. The bilingual group was divided into three subgroups according to the degree of bilingualism (other language dominant, balanced bilingual, German dominant). For this subgroup classification, an extensive questionnaire consisting of 46 items assessing the socio-economic status of the family as well as the input quality and quantity in relation to various linguistic settings was developed and used in this study. The working hypotheses were: 1) Balanced bilingual children develop metalinguistic awareness earlier than monolingual children. 2. children who are dominant in the German language develop metalinguistic awareness like monolingual children. 3. the socio-economic status of the parents influences the development of metalinguistic awareness in all three groups.

The results verify all three hypotheses. In addition, the findings show that the influence of parental socioeconomic status on the development of metalinguistic awareness is higher in monolingual children than in bilingual children. Bilingual children benefit from the management of their two languages and can therefore weakly compensate a low socio-economic status. In line with previous research (Bialystok 1992; Delucchi Danhier & Mertins 2018), we argue that the cognitive underpinning of bilingualism (co-activation, code-switching, inhibition/activation) in combination with the parental socio-economic status, which in turn affects the quality and quantity of the linguistic input provided to the children, substantially influences the development of the metalinguistic awareness and thus literacy in young children. These important findings on bilingualism research seems especially relevant for the language pair Czech-Vietnamese – a frequent bilingual setting found in the Czech Republic.





Example: Monolingual child



Example: bilingual child

## Bibliography

Bialystok, E. (1992). Attentional control in children's metalinguistic performance and measures of field independence. *Developmental Psychology*, 28(4), 654–664. <https://doi.org/10.1037/0012-1649.28.4.654>.

Delucchi Danhier, R., & Mertins, B. (2018). Psycholinguistische Grundlagen der Inklusion. Schwerpunkt Bilingualismus. DoProfiL–Das Dortmunder Profil für inklusionsorientierte Lehrerinnen- und Lehrerbildung. Münster: Waxmann, 161-178.

Reder, F., Marec-Breton, N., Gombert, J.-E. & Demont, E. (2013). Second-language learners' advantage in metalinguistic awareness: a question of languages' characteristics. *The British journal of educational psychology*, 83(Pt 4), 686–702. <https://doi.org/10.1111/bjep.12003>.

Schneider, W. (2017). Lesen und Schreiben lernen: Wie erobern Kinder die Schriftsprache? (1. Aufl.). Kritisch hinterfragt. Springer Berlin Heidelberg.

**Plurilingual study-abroad students' perspective on language interaction  
in their linguistic repertoires**

Silvie Převrátlová (Charles University)

Plurilingualism, the integrated command of multiple languages within an individual's linguistic repertoire, plays a pivotal role in shaping language learning and use (Council of Europe, 2001; North et al., 2022). This concept extends beyond mere multilingual ability, highlighting the dynamic interconnections among languages that facilitate positive cross-linguistic influences. Such influences are known to enhance motivation and foster favorable attitudes towards learning new languages (Henry, 2011, 2017). This study ventures into the plurilingual landscapes of international students, particularly those participating in the Erasmus+ program, at Czech universities to explore the nature and domains of language interactions within their plurilingual repertoires.

Focusing on three core questions, this research investigates the perceived interactions among languages studied by the students, the domains in which these interactions occur, and how the Czech language integrates within their plurilingual frameworks. Data were gathered from 106 international students via an open-ended online questionnaire, providing insights into their linguistic experiences and perceptions.

Analysis reveals that students frequently recognize positive interactions among the languages they have learned, especially those within the same language family. Remarkably, even languages that are typologically distinct, such as Czech and German, demonstrate significant interrelations. These findings challenge the traditional monolingual bias in language education (Cook, 2016), suggesting that plurilingual individuals often benefit from the synergies among their languages when acquiring new ones.

The study underscores the didactic potential of leveraging positive language interactions in the pedagogy of language teaching and learning. It argues for a paradigm shift towards a plurilingual approach in language instruction, which acknowledges and utilizes the interconnected nature of languages. By doing so, it opens new avenues for research into plurilingualism's role in enhancing language acquisition and instruction, proposing a reevaluation of current teaching methodologies to incorporate these insights.

## Bibliography

- Cook, V. (2016). Premises of multi-competence. In V. Cook & L. Wei (Eds.), *The Cambridge handbook of linguistic multicompetence* (pp. 1–25). Cambridge University Press.
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Henry, A. (2011). Examining the impact of L2 English on L3 selves: A case study. *International Journal of Multilingualism*, 8(3), 235–255. <https://doi.org/10.1080/14790718.2011.554983>
- Henry, A. (2017). L2 motivation and multilingual identities. *Modern Language Journal*, 101(3). <https://doi.org/10.1111/modl.12412>
- North, B., Piccardo, E., Goodier, T., Fasoglio, D., Margonis-Pasinetti, R., & Rüschoff, B. (2022). *Enriching 21st century language education - The CEFR Companion volume in practice*. Council of Europe.
- Thompson, A. (2016). How do multilinguals conceptualize interactions among languages studied? Operationalizing perceived positive language interaction (PPLI). In M. Uno, H. I. Park, A. Tyler, & L. Ortega (Eds.). *The usage-based study of language learning and multilingualism* (pp. 91-111). Georgetown University Press.

**Language and cognition in dementia. A call for a more detailed assessment of bilingualism**

Aldona Rzitki (Institut für Slavische Sprachen und Literaturen, Universität Bern)

The talk is planned to be part of the panel “New Empirically Based Perspectives on Multilingualism.” It aims to shed light on the intersection of age, cognition, dementia, and multilingualism by illustrating the challenges that can be encountered in the design, implementation, and evaluation of an empirical study with multilingual participants affected by dementia. The talk will focus on the theoretical and practical construction of this key variable in psycholinguistic and sociolinguistic studies and offer a "best practice" guideline by presenting a study with monolingual and bilingual participants affected by dementia (n=20) which wants to examine cognitive benefits in the German-Polish bilinguals.

The life expectancy of the population is continuously increasing. However, as people get older, the prevalence of age-related diseases such as dementia also increases. There are significant differences between those who experience a natural aging process and those who are affected by pathological limitations due to neurodegenerative diseases. There is an interdisciplinary interest in finding which individual factors in a person's life promote a natural, healthy aging process. Multilingualism research is also interested in language(s) as an influencing factor for cognition. Studies conducted by Bialystok et al. have significantly boosted the research on cognitive benefits. They have shown that bilinguals affected by dementia experience symptoms up to 4 years later than comparable monolingual individuals (Bialystok et al., 2007). A closer look at the research landscape on cognitive benefits in old age reveals that the results of the studies are wildly divergent (Van den Noort et al., 2019). Some studies confirm the benefits, while others find no influence of multilingualism. Bialystok et al. comment on this controversial research discourse with varying results, stating that a closer look at the constructs is needed. Other researchers also call for a more transparent and reflective approach to the constructs and methods (de Bruin, 2019).

This is essential in the research discourse on cognitive advantages in old age in order to gain valid insights, as three significant variables are involved, which require a more precise definition. Firstly, there is the construct of age, which is understood differently across disciplines as well as within linguistic fields. Secondly, dementia is another varying key term that encompasses various brain diseases whose symptoms can vary and be different depending on the degree of the disease. Furthermore, the essential variable of multilingualism is understood very heterogeneously in theoretical terms and reflects the methodology, so the research results may vary significantly.

The workshop will explain all variables and focus on the last one, contextualizing it within the lifelong approach. This emphasizes that multilingualism cannot simply be seen as a static state but has to be conceptualized as a dynamic construct. The lifelong accumulation of experiences, which particularly become significant for older participants, plays an essential role in the outcomes of empirical studies (de Bot & Schrauf, 2009).

This workshop talk aims to enable a closer look at one specific psycholinguistic study design, which compares mono and bilingual participants affected by dementia (n=20) and aims to find cognitive benefits in the German-Polish bilingual group. This study includes an assessment of language proficiency, cognitive abilities, executive functions, and a questionnaire with (linguistic) biographical information.

The goal of this talk is to sensitize the audience to theoretical and methodological considerations in multilingualism research and enable a reflected dialogue about multilingualism and the potential cognitive benefits in old age.

### **Bibliography**

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, *45*(2), 459–464.

<https://doi.org/10.1016/j.neuropsychologia.2006.10.009>

De Bot, K., & Schrauf, R. W. (2009). *Language development over the lifespan*. Routledge New York.

De Bruin, A. (2019). Not All Bilinguals Are the Same: A Call for More Detailed Assessments and Descriptions of Bilingual Experiences. *Behavioral Sciences*, *9*(3).

<https://doi.org/10.3390/bs9030033>

Van den Noort, M., Vermeire, K., Bosch, P., Staudte, H., Krajenbrink, T., Jaswetz, L., Struys, E., Yeo, S., Barisch, P., & Perriard, B. (2019). A systematic review on the possible relationship between bilingualism, cognitive decline, and the onset of dementia. *Behavioral Sciences*, *9*(7), 1–37.

## Psaná čeština romských žáků – etnolektní projevy

Kateřina Šormová (FF UK)

Příspěvek se vztahuje k problematice funkční gramotnosti romských žáků, konkrétně ke schopnosti písemné produkce, která je se čtením, tedy recepcí, bezprostředně spojená.

Jak uvádí Šormová (Šormová 2015, s. 9), význam gramotnosti pro dnešního jedince je zcela klíčový; výsledky výzkumů naznačují, že lidé s rozvinutými gramotnostními dovednostmi mají zvýšenou pravděpodobnost mít vyšší sebevědomí, lepší zdravotní stav, uspokojivější zaměstnání a vyšší mzdy než lidé s nízkými gramotnostními dovednostmi. Rizikem snížené úrovně gramotnosti jsou ohroženy nejčastěji osoby z odlišného kulturního prostředí, nezaměstnané a nekvalifikované osoby, osoby s velmi nízkým vzděláním a lidé ze sociálně slabých vrstev. Romští žáci patří do skupiny osob ohrožených tímto rizikem, a je tedy třeba nastavit opatření pro eliminaci těchto rizik. Význam rozvoje jazykové kompetence romských žáků je deklarován také ve strategických dokumentech, např. Strategie rovnosti, začlenění a participace Romů 2021–2030 uvádí nutnost *Zajistit podpůrné mechanismy v rámci povinné školní docházky dětem a žákům s odlišným mateřským jazykem či romským etnolektem češtiny, a to v oblastech jejich řečového vývoje, jazykové a komunikační kompetence,*

V některých oblastech romské komunity je romština vystřídána etnolektem ve funkci prvního jazyka (Bořkovicová, 2006, s. 46), romské děti tedy nepřicházejí do škol bez kompetence v češtině, nicméně je nutno brát v úvahu, že jde o substandardní varietu češtiny ovlivněnou mnoha prvky z romštiny. Při výuce českého jazyka je na místě zaměřit se zejména na jevy, které mohou nabývat etnolektní povahy a které sami žáci nedokážou rozklíčovat.

Příspěvek představí písemnou produkci romských žáků základních škol. Presentovaná data pocházejí z kvantitativního výzkumu čtenářské gramotnosti, který proběhl ve školním roce 2013/2014 na základních školách běžného typu a na tehdejších základních školách praktických. Data pocházejí z otevřených otázek testu čtenářské gramotnosti, jejichž cílem bylo ověření porozumění přečteného textu, schopnost verbální formulace odpovědi a její zápis.

Z analyzovaných dat je patrné, že psaný projev výběrového souboru romských žáků je nejen zatížen vysokou chybovostí, ale vykazuje také výrazné obtíže v oblasti porozumění slov (významová záměna slov) a v oblasti stylizace vlastní odpovědi (nepřesné a nečesky formulované odpovědi). V mnoha případech žák nerozumí položené otázce a odpovídá na něco jiného, případně v odpovědi pouze opakuje informace již obsažené v otázce. Odpovědi žáků nesou rysy mluveného jazyka, v některých případech jsou spíše záznamem proudu myšlenek připomínajícím asociativní metodu než myšlenkově ucelenou odpovědí. V otázkách, ve kterých bylo úkolem interpretovat informace získané z textu nebo formulovat závěr dochází často k tomu, že žáci vyberou nějakou konkrétní informaci z textu, a použijí ji jako odpověď. Nejméně chyb z hlediska porozumění dělají žáci v otázkách zaměřených na vyhledání konkrétní informace z textu, ačkoliv mohou pracovat přímo s textem, jsou i tyto chyby zatíženy vysokým výskytem chyb způsobených nepozorným přepisem.

Výsledky naznačují, že romští žáci často používají etnolektní prvky, které mohou být odlišné od standardní češtiny, a že jejich omezená schopnost vyjadřování se psanou češtinou může být důsledkem nedostatečné jazykové podpory a kulturních bariér. Tato znevýhodnění mohou vést k nedostatečnému rozvoji jazykových dovedností a k omezenému kontaktu s normativní češtinou, zejména psanou formou.

Závěry této studie mají důležité implikace pro vzdělávací politiku a praxi, zejména v oblasti podpory jazykového rozvoje a inkluзивity pro romské žáky ve školním prostředí.

## **Bibliografie**

BOŘKOVCOVÁ, M. (2006) Romský etnolekt češtiny: případová studie. Praha: Signeta.

ŠORMOVÁ, K. (2016). Jak čtou Romové. Praha, FF UK.

Strategie rovnosti, začlenění a participace Romů 2021–2030. Strategie 20-30 +

12

*Lingvistické aspekty překladu:  
nové impulsy  
i návrat ke kořenům*

organizují

David Mraček

Vanda Obdržálková

(CZE)



# Popis panelu

Translatologický panel je otevřen příspěvkům zpracovávajícím témata zaměřená na překlad či tlumočení z lingvistické perspektivy. Očekáváme a vítáme tematickou i metodologickou diverzitu a předpokládáme, že mezi přednášejícími budou jak translatologové se zájmem o lingvistická hlediska překladu a tlumočení, tak filologové, u nichž překladatelská témata tvoří pouze část jejich badatelského zájmu.

Příspěvky mohou být zaměřeny na tradiční translatologická témata, jako jsou stopy jazyka originálu v překladových a tlumočnických výstupech (interference, překladatelština, srov. Snell-Hornby et al. 1999), specifika převodu textů různých funkčních stylů a oborů, včetně odborné terminologie a problematiky terminologických zdrojů. Díky stále propracovanějším metodám i rozšiřujícímu se množství dostupných dat zůstává i po třiceti letech čilou a přínosnou oblastí výzkumu korpusová translatologie a její snaha vysledovat rysy typické pro překladové texty, případně zachytit individuální styl jednotlivých překladatelů (srov. Bernardini, Kenny 2020).

Poslední léta přinesla zvýšený zájem o celou řadu nových výzkumných témat, především v reakci na proměňující se překladatelskou praxi. Příspěvky se tak mohou týkat například široké oblasti audiovizuálního překladu s důrazem na přístup komparativní, psycholingvistický či sociolingvistický (srov. Pérez-González 2020). Vzhledem k nevyhnutelnosti překladu do nemateršského jazyka v každodenní praxi moderního překladatele je žádoucí, aby součástí panelu byly rovněž příspěvky zaměřené na kvalitativní i procesuální hlediska tzv. direkcionality (srov. Duběda, Mraček, Obdržálková 2018).

Problematika strojového překladu je předmětem lingvistického zájmu už sedm desítek let, avšak razantní nástup neuronových překladačů a umělé inteligence přináší nové výzkumné otázky zaměřené na kvalitu výstupů strojového překladu a jeho posteditace a případné důsledky těchto technologií pro překladatelský proces. Vítáme tedy i příspěvky s touto tematikou, jejíž zkoumání zároveň vyžaduje nové metodologické přístupy.

S ohledem na zavedené paradigma výzkumu překladu a tlumočení lze výše naznačená (a pochopitelně i mnohá další) témata uchopit z hlediska teoretického, deskriptivního a aplikačního, včetně didaktiky překladu a tlumočení.

Translatologický panel si klade za cíl otevřít následující konkrétní otázky.

(1) S ohledem na komplexnost překládání jako procesu a s tím i překladu jako jeho produktu není možné se v korpusové translatologii spoléhat pouze na čistě deskriptivní, lingvistický přístup k samotným korpusovým datům. Translatologicky validní závěry si vyžadují kontextualizaci dat, což předpokládá zapojení analytických nástrojů z jiných oblastí. Které metody se v tomto ohledu obzvláště dobře osvědčily?

(2) Jakým způsobem a s jakými výsledky lze využívat strojový překlad (MT), ChatGPT a podobné velké jazykové modely (LLM) v překladatelské praxi? Dílčí výzkumné otázky se mohou týkat mj. hodnocení kvality výstupů strojového (posteditovaného) překladu, vlivu MT a LLM na překladatelský proces, posteditace v překladu do nemateršského jazyka, didaktiky posteditace, role těchto technologií v překladu audiovizuálním, uměleckém či odborném, včetně jejich využitelnosti jako zdrojů terminologie.

(3) Jak dnes zkoumat lingvistické aspekty literárního, audiovizuálního či odborného překladu?

(4) S ohledem na stále rozšířenější a rozmanitější využití nových technologií v tlumočnické praxi se mj. nabízí otázka, jak hodnotit kvalitu výstupů strojového tlumočení ve srovnání s lidským tlumočením, případně jaká je efektivita ChatGPT při tvorbě terminologických glosářů.

## Literatura

Bernardini, S. – Kenny, D. (2020): [Corpora](#). In M. Baker & G. Saldanha (eds.) *Routledge Encyclopedia of Translation Studies*. London: Routledge, 110–115.

Duběda, T. – Mraček, D. – Obdržálková V. (2018): *Překlad do nemateřského jazyka. Fakta, otázky, perspektivy*. Praha: Karolinum.

Pérez-González, L. (2020): *Audiovisual translation*. In M. Baker & G. Saldanha (eds.) *Routledge Encyclopedia of Translation Studies*. London: Routledge, 30–34.

Snell-Hornby, M. – Hönl, H. G. – Kussmaul, P. – Schmitt, P. A. (1999): *Handbuch Translation*. Zweite, verbesserte Auflage. Tübingen: Stauffenburg.

# Obsah

*Jazykový obraz českého práva ve francouzských překladech*

Tomáš Duběda (Univerzita Karlova)

*Teorie relevance ve výuce překladu dnes – cost-benefit analysis*

Renata Kamenická (Masarykova univerzita)

*Asymetrie v právním jazyce jako překladatelský problém*

Ondřej Klabal (Filozofická fakulta Univerzity Palackého v Olomouci)

*Rozvíjanie prekladateľských zručností na báze paralelných korpusov – koncepcia cvičebnice*

Jana Levická, Miroslav Zumrík (Jazykovedný ústav Ľ. Štúra SAV)

*Tvorba optimálních referenčních překladů: aplikace Levého konceptů překlad jako rozhodovací proces a maximinimová strategie*

David Mraček (Univerzita Karlova, Filozofická fakulta)

*Slovosled ve strojovém a lidském překladu právního textu*

Vanda Obdržálková (Ústav translatologie, FF UK)

*Koreferenční vztahy v odborných textech a strojový překlad (na příkladu španělštiny a češtiny)*

Jana Pešková (Jihočeská univerzita v Českých Budějovicích)

*Slovtvorba v básnickém překladu aneb jak dosáhnout prostřednictvím nových kompozit funkční ekvivalence (na příkladu českých překladů šibeniční poezie Christiana Morgensterna)*

Jana Pokojová (Univerzita Karlova)

*Komparace české a ruské jazykové verze počítačové hry „Craftomation 101: Programming & Craft“ vyhotovené OpenAI GPT-4 v porovnání s lidským překladatelem*

Mgr. Tereza Poledníková (Ústav slavistiky FF MU)

**Jazykový obraz českého práva ve francouzských překladech**

Tomáš Duběda (Univerzita Karlova)

Potřeba mezinárodní právní komunikace činí z právního překladu významné odvětví překladatelství: podle studie ELIS (2023) je právo a administrativa nejčastější specializací evropských překladatelů a překladatelek a v mnoha evropských zemích je soudní překlad jediným regulovaným odvětvím překladu. Skrze překlad se cizí právní systémy otvírají českému publiku a české právo se naopak otvírá mluvčím dalších jazyků. Právní systém a jazyk, jehož prostřednictvím se tento systém materializuje, k sobě přitom pojí velmi těsný vztah (Cornu 2005, s. 4; Chromá 2014, s. 23).

Soustředíme-li se na překlad českých právních textů do cizích jazyků, zjišťujeme, že probíhá převážně na domácí překladatelské scéně a mezi jeho aktéry dominují čeští překladatelé. Tato situace je typická pro kultury s málo rozšířenými jazyky (Pavlovič 2007; IAPTI 2015) a z hlediska sociologie překladu má další koreláty: obecně sdílené povědomí o tom, že právní komunikace směrem k cizím kulturám leží na bedrech místních překladatelů (Prunč 2000), značnou toleranci vůči nerodilému překladu, nízkou „viditelnost“ problematických překladů, a tudíž obtížnou samoregulaci kvality v tomto překladatelském segmentu (Duběda 2018), ale také např. skutečnost, že právníkové slovníky s češtinou mají téměř výhradně české autory (Duběda, v tisku).

Toto sociologické pozadí má zásadní vliv na překladatelské strategie a kvalitu překladů do cizích jazyků. Veřejně přístupná část přeložených textů vytváří „stínový“ obraz českého práva v jednotlivých cizích jazycích, stává se vzorem pro další překlady a potenciálně i materiálem pro webové vyhledavače ekvivalentů (např. Linguee) a strojové překladače. Specificky české právní koncepty jsou předmětem „terminologického vyjednávání“, do nějž vstupují autoritativní texty včetně právnických slovníků, online zdroje, překladatelova vlastní zkušenost, ale také tradice a v neposlední řadě i náhoda. Pro právní texty přeložené z češtiny je příznačná terminologická roztržičnost, která může dosahovat značných rozměrů, jak ukazuje nedávný průzkum provedený u soudních překladatelů (Duběda 2020, 2021). Terminologická standardizace cizojazyčné terminologie, jak ji známe z vícejazyčných zemí či mezinárodních institucí, je v českém kontextu neproveditelná.

Empirický vhled do francouzských překladů českých právních textů nabízí specializovaný paralelní korpus budovaný v Ústavu translologie FF UK od roku 2023 (v současnosti cca 550 tis. slov v české části – původní a přeložené verze české legislativy, smluv, obchodních podmínek firem atd.). U textů, které vykazují rysy nerodilého překladu, potvrzuje analýza vysokou nejednotnost francouzských terminologických ekvivalentů, tendenci k doslovným řešením a omezenou textovou kompetenci v cílovém jazyce (v tomto ohledu plní korpus paradoxně funkci žakovského korpusu), a v důsledku toho celkově nevyrovnanou kvalitu překladů (včetně institučně zaštitěných, jako je např. občanský zákoník). Překlady, které nejeví rysy nerodilého překladu (a byly tedy vyhotoveny nebo alespoň revidovány rodilým mluvčím), se mnohem více blíží textovým konvencím právní francouzštiny a z hlediska terminologického je v nich významně větší podíl funkčních ekvivalentů. Celkově se tedy ukazuje, že na ose dokumentárnost – instrumentálnost (Nord 2005) jsou čeští překladatelé blíže prvnímu pólu a frankofonní překladatelé blíže druhému pólu.

První poznatky získané z této textové databáze potvrzují význam korpusového výzkumu právního jazyka (Pontrandolfo 2019), který je u nás zatím okrajový. Umožňují činit závěry o jazykovém obrazu českého práva ve francouzštině, který je určující pro komunikaci směrem k dalším kulturám.

## Bibliografie

- Biel, L.** (2008). Legal terminology in translation practice: Dictionaries, googling or discussion forums? *SKASE Journal of Translation and Interpretation* 3(1), pp. 22–38.
- Cornu, G.** (2005). *Linguistique juridique*. Montchrestien.
- Duběda, T.** (2018). La traduction vers une langue étrangère et son rôle dans la formation des futurs traducteurs. *Synergies Europe* 13, pp. 161–168.
- Duběda, T.** (2020). Soudní překladatelé a terminologické nástrahy. Průzkum k překladu právních termínů z češtiny do francouzštiny (1. část). *Soudní tlumočnick 2/2020*, pp. 25–28.
- Duběda, T.** (2021). Soudní překladatelé a terminologické nástrahy. Průzkum k překladu právních termínů z češtiny do francouzštiny (2. část). *Soudní tlumočnick 1/2021*, pp. 30–32.
- Duběda, T.** (v tisku). Překladové právníkové slovníky: budou brzy minulostí? *Časopis pro moderní filologii*.
- ELIS (European Language Industry Survey) 2023*. <<https://elis-survey.org/repository>>
- Chromá, M.** (2014): *Právní překlad v teorii a praxi. Nový občanský zákoník*. Karolinum.
- IAPTI** (2015). *Translation into a non-native language*. <[https://www.iapti.org/files/surveys/2/IAPTI\\_non-native\\_report.pdf](https://www.iapti.org/files/surveys/2/IAPTI_non-native_report.pdf)>
- Linguee*. <<https://www.linguee.com>>
- Nord, C.** (2005) *Text Analysis in Translation*. Rodopi.
- Pavlović, N.** (2007) Directionality in translation and interpreting practice. Report on a questionnaire survey in Croatia. *Forum* 5(2), pp. 79–99.
- Pontrandolfo, G.** (2019) Corpus methods in legal translation studies. In: *Research Methods in Legal Translation and Interpreting*. Routledge. pp. 13–28.
- Prunč, E.** (2000). Translation in die Nicht-Muttersprache und Translationskultur. In: *Translation into Non-Mother Tongues: In Professional Practice and Training*. Stauffenburg. pp. 5–20.

## Teorie relevance ve výuce překladu dnes – cost-benefit analysis

Renata Kamenická (Masarykova univerzita)

Příspěvek reaguje na třetí podotázku translatologického panelu (Jak dnes zkoumat lingvistické aspekty literárního, audiovizuálního či odborného překladu?) se zacílením na oblast didaktiky překladu. Pokouší se zodpovědět konkrétní otázku, zda se jeví jako smysluplné zařazení základů teorie relevance coby moderní v pragmatice zakořeněné lingvistické teorie do výuky překladu na magisterské úrovni studia, zejména s ohledem na její a) poměrně vysokou explanatorní hodnotu a souvztažnosti s jinými aktuálně využívanými teoriemi překladu a b) možnosti jejího praktického využití při rozhodování v procesu překladu u studentů překladu. Položit si tuto otázku se nám jeví jako vhodné zejména proto, že podrobnosti teorie relevance nepatří k teoriím standardně zařazovaným do překladatelských kurikul, částečně jistě proto, že uvažování v tomto teoretickém rámci obnáší nezanedbatelnou vstupní investici. Pokládáme tedy za vhodné přezkoumat v měnícím se kontextu uvažování o překladu (mj. vlivem současných technologických změn), co může teorie relevance v dialogu mezi vyučujícími a studenty překladu a v interakci s didakticky zacílenými překladovými situacemi nabídnout a zda se bilance mezi touto investicí a tím, co při výuce překladu získáme, jeví jako pozitivní.

Otázkou, kterou si klademe, je, zda může být vystavení budoucích překladatelů základům teorie relevance v kontextu magisterských překladatelských programů časově efektivní, pokud má „překlad“ této teorie pro účely výuky zůstat dostatečně „věrný“. K uvažovaným přínosům teorie relevance v kontextu dalších teorií patří: návaznost na rámec moderní pragmatiky (pragmatický výklad strukturače diskurzu, zásadní role kontextu), souvislost interpretativního užití jazyka (tzv. interpretive use, Gutt 1991) s aktuálními přístupy k ekvivalenci (Nidova dynamická ekvivalence, Pymova přímá a direkční ekvivalence, Chestermanova relevantní podobnost), analogie interpretačně zaměřeného módu komunikace (Alves 2003, 2007, 2010) s překladem jako metakomunikací (Popovič) a moderním širším sémiotickým pojetím překladu (Kobus Marais a další) a v neposlední řadě také kognitivní orientace v teorii relevance, zahrnující stranu komunikátora i adresáta komunikace/překladu. Příspěvek dále ukazuje využití opory v teorii relevance na konkrétních příkladech z jazykového páru angličtina/čeština zahrnujících studentské překlady a práci s výstupy NMT, kde se konkrétně soustředí např. na restrikcí nežádoucích implikatur překladových verzí nebo na tzv. lingvistický *foreground* a *background* (Unger 2006) v syntaktické výstavbě překladového diskurzu.

## Bibliografie

Alves, Fabio, and Gonçalves, José Luiz (2003) “A relevance theory approach to the investigation of inferential processes in translation.” In: *Triangulating Translation: perspectives in process oriented research*, Alves, Fabio (ed.), Amsterdam/Philadelphia: John Benjamins, 11–34.

Alves, Fabio, and Gonçalves, José Luiz (2007) “Modelling translator’s competence: relevance and expertise under scrutiny.” In: *Translation studies: doubts and directions*. Yves Gambier, Miriam Shlesinger, Radegundis Stolze (eds.), Amsterdam/Philadelphia: John Benjamins, 41–55.

Alves, Fabio, and Goncalves, José Luis (2010) “Relevance and Translation”, in: Gambier, Yves, and van Doorslaer, Luc (eds.) *Handbook of Translation Studies*, Volume 1, Amsterdam/Philadelphia: John Benjamins Publishing Company, 279–284.

Clark, Billy (2013) *Relevance Theory*. Cambridge: Cambridge University Press.

Gutt, Ernst-August (1991) *Translation and relevance: cognition and context*. Oxford: Blackwell.

Unger, Christoph (2006) *Genre, Relevance and Global Coherence*. Palgrave Macmillan.

**Asymetrie v právním jazyce jako překladatelský problém**

Ondřej Klabal (Filozofická fakulta Univerzity Palackého v Olomouci)

Právní překlad se oproti jiným oblastem vyznačuje celou řadou specifíků, jedním z nichž je i asymetrie právních systémů a právních jazyků (cf. Prieto Ramos 2021), jejíž součástí je i tzv. směrová asymetrie (cf. Duběda 2021). Obecně lze pod pojmem asymetrie rozumět situace, kdy nelze nalézt ekvivalent jedné ku druhé, což může mít různé příčiny např. synonymii, polysémii nebo systémové rozdíly. Jednotlivé případy si pak vyžadují různá řešení. Zatímco některé konkrétní případy asymetrie v právním jazyce jsou zejména pro jazykový pár čeština-angličtina podrobně popsány (např. Chromá 2014), podrobnější systematický popis strategií pro jejich překonání chybí, případně se omezuje na jiné jazykové páry (cf. Duběda (2021) pro jazykový pár čeština-francouzština). Tento příspěvek si tak klade za cíl na příkladech z češtiny, angličtiny a španělštiny ukázat několik častých asymetrických situací napříč různými právními odvětvími, na něž právní překladatelé naráží, a představit možné nástroje a strategie, které jsou překladatelům k dispozici, a jejich praktické použití. Konkrétně se příspěvek věnuje situacím, kdy ve výchozím jazyce existuje sémanticky širší pojem než v cílovém jazyce, což může vést ke dvěma překladatelským problémům. Buďto pojmu odpovídá vícero ekvivalentů a je třeba na základě pojmových znaků vybrat jeden z nich jako ekvivalent v daném kontextu (např. české směnce v angličtině odpovídá termín *promissory note*, pokud se jedná o směnku vlastní, a *bill of exchange*, pokud se jedná o směnku cizí, ale nadřazený pojem chybí), nebo je naopak třeba pojem přeložit neologismem, který bude sémanticky stejně široký, byť v daném jazyce neexistuje nebo se v daném kontextu nepoužívá (např. překlad české smlouvy o dílo). Dalším častým a překladatelsky problematickým případem asymetrie jsou pak situace, kdy je sice pojmová realita obdobná, ale v každém jazyce je použito jiné dělící kritérium (např. rozdíl mezi přečiny a zločiny v českém právu vs. *summary* a *indictable offences* v Anglii). Přestože se v souvislosti s asymetrií hovoří zejména ve vztahu k právní terminologii, příspěvek poukáže i na analogické příklady z oblasti frazeologické, které mohou být překladatelsky potenciálně složitější, jak popisuje např. Klabal (2019), jelikož k jejich vyřešení nestačí slovník či pojmová analýza, ale je třeba pracovat s autentickými dokumenty např. ve formě korpusů (Biel 2017).



## **Bibliografie**

Biel, Łucja. (2017). “Enhancing the Communicative Dimension of Legal Translation: Comparable Corpora in the Research-Informed Classroom.” *Interpreter and Translator Trainer* 11 (4): 316–36.

Chromá, Marta. (2014). *Právní překlad v teorii a praxi: Nový občanský zákoník*. Praha: Karolinum.

Duběda, Tomáš (2021). Direction-Asymmetric Equivalence in Legal Translation. *Comparative Legilinguistics* 47, 57-72.

Klbal, Ondřej. (2019). “Corpora in Legal Translation: Overcoming Terminological and Phraseological Asymmetries between Czech and English.” *CLINA* 5 (2): 165–86.

Prieto Ramos, Fernando (2021) Translating legal terminology and phraseology: between inter-systemic incongruity and multilingual harmonization, *Perspectives*, 29 (2), 175-183.

**Rozvíjanie prekladateľských zručností na báze paralelných korpusov – koncepcia cvičebnice**

Jana Levická, Miroslav Zumrík (Jazykovedný ústav Ľ. Štúra SAV)

Cieľom príspevku je predstaviť koncepciu a spôsob využitia viacjazyčných paralelných korpusových dát ako materiálového a metodologického základu pri príprave cvičebnice korpusovej práce určenú na výučbu budúcich i aktívnych prekladateľov. Aplikačný rozmer príspevku tým prepája oblasť didaktiky prekladu s aplikačnou polohou korpusovej lingvistiky. Príspevok bude odkazovať na použité paralelné korpusy (slovensko-anglický a slovensko-nemecký paralelný korpus) oddelenia Slovenského národného korpusu Jazykovedného ústavu Ľ. Štúra SAV, v. v. i., pričom poukáže na ich potenciál i limity využitia v prekladateľskej a translatologickej praxi, ako aj pri príprave publikácie. V príspevku sa budú prezentovať zvolené tematické okruhy a najmä štruktúra prekladateľských úloh tvoriacich plánovaný výstup. V znení úloh sa akcentuje šírka, variabilita a otvorený charakter možných prekladateľských riešení, a tiež sa zdôrazňuje analytický prístup k riešeniu prekladateľských problémov, ktorý sa opiera o niektoré kvantitatívne charakteristiky dostupné v korpusovom manažéri. Zadania úloh majú prekladateľov a translatológov motivovať k dôkladnej analýze existujúcich a hľadaniu vlastných riešení. Pripravovaná publikácia sa sekundárne dotýka aj identifikácie nepriamych prekladateľských postupov (moduláciu, transpozíciu atď.). Časť zadaní úloh sa orientuje na odhaľovanie interferencií aj pomocou porovnania riešení v neprekladových pôvodných slovenských textoch jednojazyčného korpusu. Publikácia teda kladie dôraz na kreatívne rozvíjanie prekladateľských zručností.

## Bibliografie

- Beeby, Allison – Inés, Patricia Rodríguez – Sánchez-Gijón, Pilar (eds.). 2009. *Corpus Use and Translating. Corpus use for learning to translate and learning corpus use to translate.* Amsterdam/Philadelphia: John Benjamins Publishing Company. 151 s.
- Benko, Vladimír – Butašová, Anna – Puchovská, Zuzana – Segretain, Arnaud – Zeleňáková, Mariana. 2019. *Webové korpusy Aranea: učebnica pre učiteľov cudzích jazykov, prekladateľov, tlmočníkov, filológov a študentov filologických odborov.* Bratislava: Univerzita Komenského v Bratislave. 181 s.
- Chlumská, Lucie. 2018. *Překladová čeština a její charakteristiky.* Praha: Nakladatelství Lidové noviny. 150 s.
- Malamatidou, Sofia. 2019. *Corpus Triangulation. Combining Data and Methods in Corpus-based Translation Studies.* London and New York: Routledge Taylor and Francis Group. 190 s.
- Mikhailov, Mikhail – Cooper, Robert. 2016. *Corpus Linguistics for Translation and Contrastive Studies. A guide for research.* London and New York: Routledge Taylor and Francis Group.
- Slovenský národný korpus – par-sken-all-4.0. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2022. Dostupný na <https://korpus.sk>.
- Slovenský národný korpus – par-skde-all-3.0. Bratislava: Jazykovedný ústav Ľ. Štúra SAV 2022. Dostupný na <https://korpus.sk>.
- Šimková, Mária – Gajdošová, Katarína – Kmeťová, Beáta – Debnár, Marek. 2017. *Slovenský národný korpus. Texty, anotácie, vyhľadávania.* Bratislava: Jazykovedný ústav Ľ. Štúra SAV – Vydavateľstvo Mikula. 168 s.
- Šimková, Mária – Gajdošová, Katarína. 2020. *Slovenský národný korpus. Používanie, príklady, postupy.* Bratislava: Jazykovedný ústav Ľ. Štúra SAV – Vydavateľstvo Mikula. 336 s.

**Tvorba optimálních referenčních překladů:  
aplikace Levého konceptů překlad jako rozhodovací proces a maximimimová strategie**

David Mraček (Univerzita Karlova, Filozofická fakulta)

Jako referenční se označují překlady, jež pocházejí od lidského překladatele a využívají se při hodnocení překladů strojových. S tím, jak se kvalita strojových překladů blíží úrovni překladů lidských, respektive ji v některých aspektech dokáže předstihnout (Popel et al. 2020), rostou též požadavky na kvalitu referenčních překladů. Optimální referenční překlady, tedy produkty usilující o nejvyšší možnou kvalitu, podle všeho nemohou záviset na jednom jediném člověku, budou spíše výsledkem konsenzu určité skupiny překladatelů a translatořů (Kloudová et al. 2023).

V naší studii sledujeme spolupráci dvojice překladatelů, kteří zároveň působí jako badatelé v oblasti kvality překladu a vyučující teoretických i praktických předmětů souvisejících s překladem, na tvorbě optimálních referenčních překladů. Projekt zahrnoval překlady celkem 50 anglických textů do češtiny, výchozími texty byly zpravodajské články o průměrné délce 1,5 normostrany. Dvojice odborníků měla za úkol vyjít z trojice překladů dodaných profesionálními překladateli a jejich syntézou dospět k optimální verzi, jež bude využitelná pro hodnocení strojových překladů, badatelské účely, případně trénování strojových překladačů. Již na začátku práce však konstatovali překvapivou suboptimalitu dodaných překladů, která byla na mnoha místech tak výrazná, že namísto syntézy původních řešení musela dvojice přistoupit k jejich výraznému přepracování, ba velmi často vytvářela překlad zcela nový.

Suboptimalita původních profesionálních překladů se týkala všech kritérií, jež si s odborným hodnocením kvality překladu spojujeme: vedle významových nepřesností a výpustek, gramatických chyb a ryze formálních lapsů jako např. překlepů bylo nejvíce nedostatků shledáno na rovině stylizační. Ty byly z velké části zapříčiněny interferencí výchozího jazyka, která se projevuje jednak přímo, přítomností struktur považovaných v cílovém jazyce za nesystémové, jednak nepřímo, absencí prvků, jež přispívají k idiomatičnosti cílového jazyka, ale ve výchozím textu pro ně překladatel nenašel podnět (Emmerová 2001; Kufnerová 2003 a 2009; Newmark 1991).

Celý proces tvorby optimálních referenčních překladů tak byl složitější a pro zúčastněné časově i psychologicky náročnější, než jak se původně předpokládalo. Psychologii (spolu)práce sledované dvojice přiblížíme pomocí klasických teoretických konceptů, jež do uvažování o překladu průkopnický zavedl Jiří Levý: pojetí překladu jako rozhodovacího procesu a maximimimová povaha překládání.

Z hlediska překladatelovy pracovní situace je překládání v každém svém okamžiku procesem rozhodovacím, kdy se překladatel rozhoduje mezi určitým počtem alternativ. Výběr z množiny prvků přitom řídí tzv. instrukce pro výběr (Levý 1971: 32), mezi něž lze v případě uměleckého překladu řadit kontext, strukturu překladatelovy jazykové paměti a estetickou normu. Jak vypadá rozhodovací situace a jaké další instrukce do ní vstupují, je-li dvojice odborníků konfrontována nejen s předlohou, ale i několika již existujícími překladovými řešeními a jejím úkolem je syntézou stávajících překladů, diskusí nad nimi a postupným redigováním dospět do bodu, kdy nikdo z dvojice nepřijde s lepším řešením?

Překladatelská praxe je podle Levého (1971: 38) převážně pragmatická. Překladatel intuitivně sahá po tzv. maximimimové strategii, „volí ze škály možností tu, která slibuje při minimálním úsilí maximální efekt“. Jinými slovy „překládá tak, aby i při nejnevýhodnější čtenářské reakci neklesla hodnota jeho řešení pod minimální mez, kterou připouští jazyková nebo estetická norma“. Do jaké míry se od této obvyklé strategie liší tvorba optimálních referenčních překladů, jež mají snést ta nejpřísnější kvalitativní měřítka?

Naší výzkumnou metodou je analýza písemného protokolu, jenž dokumentuje všechny fáze vzniku optimálního referenčního překladu (počáteční řešení všech tří překladatelů, poznámky k jejich nedostatkům, návrhy alternativních řešení, poznámky k překladatelským strategiím a postupům, poznámky k rešerším, zápis diskuse mezi tvůrci apod.). Prostřednictvím vybraných úseků procesního

protokolu se pokusíme platnost Levého myšlenek ověřit v oblasti neliterárního překládání, případně jeho koncepty problematizovat či rozšířit.

## Bibliografie

Emmerová, Jarmila (1991): Odras jazyka originálu v jazyce překladu. In: M. Hrdlička – R. Redek (eds.), *Hovory o překladu a tlumočení*. Praha: Interlingua Servis, 79–87.

Kloudová, Věra – Mraček, David – Bojar, Ondřej – Popel, Martin (2023): Možnosti a meze tvorby tzv. optimálních referenčních překladů: po stopách „překladařštiny“ v profesionálních překladech zpravodajských textů. *Slovo a slovesnost*. Praha: Melantrich, 84(2), 122–156.

Kufnerová, Zlata a kol. (2003): *Překládání a čeština*. Jinočany: H&H.

Kufnerová, Zlata (2009): *Čtení o překládání*. Praha: H&H.

Levý, Jiří (1971): Geneze a recepce literárního díla. In: *Bude literární věda exaktní vědou?* Praha: Československý spisovatel, 71–146.

Newmark, Peter (1991): *About Translation*. Philadelphia: Multilingual Matters.

Popel, Martin – Tomková, Markéta – Tomek, Jakub – Kaiser, Łukasz – Uszkoreit, Jakob – Bojar, Ondřej – Žabokrtský, Zdeněk (2020): Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature Communications*, vol. 11, no. 4381, 1–15.

## Slovosled ve strojovém a lidském překladu právního textu

Vanda Obdržálková (Ústav translatologie, FF UK)

Příspěvek shrne závěry kontrastivně lingvistické a translatologické studie, která se zaměřuje na nedostatky ve slovosledu identifikované v překladech právního textu z angličtiny do češtiny pořízených třemi překladači (DeepL, Google Translate a LINDAT Translation) a nástrojem ChatGPT 3.5 v porovnání s překlady lidskými. Studie částečně navazuje na předchozí výzkum zabývající se kvalitou překladů pořízených studenty a profesionálními překladateli (Duběda et al. 2018).

Ačkoli se dosavadní výzkum právního překladu zaměřuje zejména na lexikální/terminologickou, kolokační a gramatickou ekvivalenci, řada autorů vyzdvihuje ekvivalenci textovou (Baker 1992) založenou na kohezi a koherenci. Podle Chromé (2014) je tento přístup k interpretaci právního textu klíčový, protože se zaměřuje na jeho smysl. Pořadí slov ve větných celcích je jedním z prostředků, od nichž se koheze a pragmatika (smysl) sdělení odvíjí. V tomto směru příspěvek navazuje na předchozí práce věnované principům funkční větné perspektivy a tematicko-rematické návaznosti a jejich realizaci v různých jazycích (Aurová 2017, Dušková 2015). Narušení těchto principů lze považovat za projev interference a nežádoucí překladovosti.

Některé práce zaměřené na strojový překlad zmiňují pořádek slov mezi jeho hlavními nebo významnými problémy, zejména v případě jazyků, které v tomto ohledu vykazují zásadní rozdíly. Často zároveň poukazují na to, že se jedná o oblast nedostatečně prozkoumanou (Murthy et al. 2019). Jedním z důvodů, proč je slovosledu věnována menší pozornost než jiným aspektům, je fakt, že tradičně využívané systémy hodnocení kvality strojového překladu často nejsou schopny narušený pořádek slov postihnout. Pro jeho adekvátní posouzení je třeba provést hodnocení na úrovni celého textu (tzv. *document-level MT evaluation*), které je časově a technicky náročné (viz např. Rysová et al. 2019).

Z hlediska dvou základních kategorií používaných při hodnocení strojového překladu, tedy tzv. *accuracy* (míra zachování významu originálu, také nazývaná *adequacy*, popř. *fidelity*) a *fluency* (plynulost, gramatická a stylistická správnost, jednoznačnost a srozumitelnost) (viz např., Lommel et al. 2014, Popel et al. 2020, Lim et al. 2024), bývá slovosled zpravidla považován za jeden z faktorů ovlivňujících druhou z nich (Kubička et al. 2017, Toledo Báez a Carl 2019).

Metody a cíle: Anglický právní text (část obchodních podmínek) o rozsahu jedné normostrany jsme nechali přeložit do češtiny výše zmíněnými nástroji strojového překladu a nástrojem ChatGPT 3.5 (po zadání odpovídající instrukce). Lidské překlady (10 pořízených studenty a 10 profesionálními překladateli) byly získány v rámci předchozího výzkumu v období leden 2016 – únor 2017. Ve všech překladech identifikujeme případy narušení principů českého slovosledu a provedeme jejich kvantitativní a kvalitativní analýzu.

Cílem analýzy takto získaného materiálu je: 1. popsat nedostatky v oblasti slovosledu ve strojových překladech a jejich možné příčiny, 2. porovnat výstupy z uvedených nástrojů strojového překladu z hlediska dodržení zásad českého slovosledu, 3. porovnat z tohoto hlediska kvalitu studentských, profesionálních a strojově pořízených překladů, 4. zhodnotit dopad zjištěných nedostatků na plynulost a eventuálně významovou adekvátnost (viz výše) a zhodnotit míru úsilí, které vyžadují při post-editaci (tzv. *post-editing effort*). V této závěrečné části analýzy se rovněž zaměříme na rozdíly při hodnocení jednotlivých segmentů oproti hodnocení na úrovni celého textu.

## Bibliografie

- Aurová, Miroslava. 2017. *Srovnání aktuálního členění v češtině a španělštině: textově-lingvistický přístup*. České Budějovice: Jihočeská univerzita v Českých Budějovicích. Episteme. Theoria.
- Baker, Mona. 1992. *In other words: a coursebook on translation*. London: Routledge.
- Duběda, Tomáš, David Mraček, Vanda Obdržálková. 2018. *Překlad do nemateršského jazyka: fakta, otázky, perspektivy*, Univerzita Karlova, nakladatelství Karolinum.
- Dušková, Libuše. 2015. *From syntax to text: the Janus face of functional sentence perspective and intra- and interlingual study of English*. Prague: Charles University in Prague, Karolinum Press.
- Chromá, Marta. 2014. *Právní překlad v teorii a praxi. Nový občanský zákoník*. Praha: Karolinum.
- Klubička, Filip, Antonio Toral Ruiz, a M. Víctor Sánchez-Cartagena. 2017. Fine-Grained Human Evaluation of Neural Versus Phrase-Based Machine Translation. *The Prague Bulletin of Mathematical Linguistics*, 108(1), 121-132. <https://doi.org/10.1515/pralin-2017-0014>
- Lim, Zheng Wei, Ekaterina Vylomova, Trevor Cohn, Charles Kemp. 2024. Simpson's Paradox and the Accuracy-Fluency Tradeoff in Translation. *arXiv preprint arXiv:2402.12690*. <https://arxiv.org/pdf/2402.12690.pdf>
- Lommel, Arle, Hans Uszkoreit, Aljoscha Burchardt. 2014. Multidimensional Quality Metrics (MQM): A Framework for Declaring and Describing Translation Quality Metrics. *Tradumàtica* 12, 455-463.
- Murthy, Rudra, Anoop Kunchukuttan, Pushpak Bhattacharyya. 2019. [Addressing word-order Divergence in Multilingual Neural Machine Translation for extremely Low Resource Languages](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3868–3873, Minneapolis, Minnesota. Association for Computational Linguistics.
- Popel, Martin, Markéta Tomková, Jakub Tomek, Łukasz Kaiser, Jakob Uszkoreit, Ondřej Bojar, Zdeněk Žabokrtský. 2020. Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature Communications*, vol. 11, no. 4381, 1–15.
- Rysová, Kateřina; Magdaléna Rysová, Tomáš Musil, Lucie Poláková a Ondřej Bojar. 2019. [A Test Suite and Manual Evaluation of Document-Level NMT at WMT19](#). In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 455–463, Florence, Italy. Association for Computational Linguistics.
- Toledo Báez, Cristina a Carl, Michael. 2019. Machine translation errors and the translation process: A study across different languages. *The Journal of Specialised Translation*.



## Koreferenční vztahy v odborných textech a strojový překlad (na příkladu španělštiny a češtiny)

Jana Pešková (Jihočeská univerzita v Českých Budějovicích)

Strojové překladače se postupně stávají nedílnou součástí překladatelského procesu, a to zejména pro oblast překladu neuměleckých textů. Jejich kvalita se nadále vylepšuje jak stránce gramatické, tak po stránce lexikální. Způsob práce strojového překladače je však jiný než práce překladatele člověka, u něhož lze předpokládat silnější tendenci ke vnímání textu jako celku. Lidský překladatel je také schopen v textu snáze identifikovat textové vztahy zasahující nad rámec mezipropoziční syntaxe a tradiční gramatiky, a vhodnými prostředky je převádět do cílového jazyka. Tyto textové vztahy, jež označujeme jako vztahy koreferenční, mohou být v různých jazycích formalizovány odlišným typem prostředků zajišťujících logické a kohezní propojení všech součástí textu.

Některé z těchto prostředků jsou povahy lexikální (« *Jeho pes se jmenuje Rex. To zvíře je mně nesympatické.* » - koreferenční vztah založený na hyperonymii), jiné gramatické (« *Včera jsme se přestěhovali. To je dobře.* » - koreferenční vztah využívající anaforicko resumptivní funkce demonstrativa), jak pro španělštinu uvádějí např. Cuenca (2010), Parodi & Burdiles (2016). Část těchto prostředků z kontrastivního hlediska neimplikuje žádný problém pro překlad, neboť je lze přeložit formálně identickým prostředkem - např. «*Jeho pes se jmenuje Rex. To zvíře je mně nesympatické.*», kde možným překladatelským respondentem nominálního syntagmatu v anaforické funkci «*to zvíře*» může být identicky konstruované syntagma ve španělštině – např.: «*este/ese/el animal*». Bližším porovnáním češtiny a španělštiny však zjistíme, že některé typy koreferenčních vztahů nelze vždy do druhého jazyka převést analogicky respondentní jednotkou. Při strojovém překladu některých struktur může docházet k tomu, že text gramaticky a jazykově „správný“ tak bude vykazovat nedostatky vyplývající z odlišného uspořádání exponentů textové koheze. Jedním příkladem mohou být struktury typické pro španělské právní texty, které označujeme jako juxtaponované konceptualizační anafory. Příkladem tohoto typu anaforických prostředků je například nedeterminované neživotné substantivum v apozičním postavení „*circunstancia*“, jež zároveň v příkladu uvedeném níže plní funkci anaforicko resumptivní. Příklad byl excerpován z výkladové části španělského rozsudku. Překlad do českého jazyka byl pořízen prostřednictvím překladače Bing, přičemž v příspěvku budou představena řešení různých nástrojů strojového překladu.

«[...] *iba caminando por el barrio de Ripagaina de Pamplona sin llevar la mascarilla, cuyo uso es obligatorio en virtud de las medidas sanitarias impuestas, circunstancia observada por una patrulla de la Policía Foral [...]*».

«[...] *procházel se čtvrtí Ripagaina v Pamploně bez nasazené roušky, jejíž používání je povinné na základě nařízených zdravotních opatření, což je okolnost, kterou zaznamenala hlídka provinční policie [...]*».

Podobných příkladů lze nalézt více. Na jejich představení se zaměříme v našem příspěvku. Za tímto účelem budou analyzovány vybrané paralelní odborné texty ve španělštině a češtině. Naším hlavním cílem bude upozornit na problematická místa v koreferenčních vztazích při strojovém překladu mezi češtinou a španělštinou. Všimnout si přitom budeme zejména těch endoforických prostředků, které v obou jazycích fungují odlišně.

## Bibliografie

BIENÍ, Janusz (2019). «Encapsuladores nominales en la prensa polaca y española: calificación y valoración». *Neophilologica. Etudes semantico-syntaxiques des langues romanes. Prace Naukowe Uniwersytetu Slaskiego w Katowicach*, pp. 32-48.

CUENCA, Maria Josep (2010). *Gramática del texto*. Madrid: Arco/Libros.

CUENCA, Maria Josep (2018). *Comentario de textos: Los mecanismos referenciales*. Madrid: Arco/Libros.

DUQUE, Eladio, et al. (2019). *Palabras clave para organizar textos en español: Recursos pragmáticos y discursivos (1st ed.)*. Nueva York: Routledge.

DUQUE, Eladio (2021). «Las etiquetas de discurso en la enseñanza del español como lengua extranjera». En: Sáracho-Arnáiz, M.y H. Otero-Duval (eds.) *Internacionalización y enseñanza del español como LE/L2*, Madrid: ASELE, pp. 854-873.

ESPARZA-TORRES, Miguel Ángel (2006). «Pautas para el análisis de la cohesión y la coherencia en textos españoles». *Rilce: Revista de Filología Hispánica* 22/1, pp. 59-89.

GONZÁLEZ RUIZ, Ramón (2009). «Algunas notas en torno a un mecanismo de cohesión textual: la anáfora conceptual». In Rosario González & Azucena Penas (eds.), *Estudios Sobre El Texto: Nuevos Enfoques y Propuestas*, Frankfurt am Main: Peter Lang, pp. 247-278.

IZQUIERDO ALEGRÍA, Dámaso; GONZÁLEZ RUIZ, Ramón (2013a). «Encapsulación y etiquetas discursivas en el discurso parlamentario: función argumentativa a partir de un corpus paralelo». *Oralia*, 16, pp. 185-219.

IZQUIERDO ALEGRÍA, Dámaso; GONZÁLEZ RUIZ, Ramón (2013b). «La evaluación a través de un mecanismo cohesivo: las etiquetas discursivas en un corpus de editoriales». In C. Llamas Saíz, C. Martínez Pasamar & M. Casado Velarde (eds.). *Léxico y argumentación en el discurso público actual*. Frankfurt am Main: Peter Lang, pp.149-185.

LÓPEZ SAMANIEGO, Ana (2010). «Documentos profesionales con destinatarios no expertos. El empleo de los mecanismos referenciales en la sentencia del 11M». *Revista Signos*, 2010, 43(72), pp. 99-123.

PÉREZ ÁLVAREZ, Bernardo Enrique (2014). «Funciones discursivas de esto, eso y aquello en un corpus oral». *Anuario de Letras*, vol. 2, núm. 1, 2014, pp. 103-140.

PARODI, Giovanni; BURDILES, Gina (2016). «Encapsulación y tipos de coherencia referencial y relacional: el pronombre ello como mecanismo encapsulador en el discurso escrito de la economía». *Onomázein*. pp. 107-129.

TORNER CASTELLS, Sergi, et al. (2011). «Problemas en el uso de las anáforas en producciones escritas de español como lengua extranjera». *Revista Española de Lingüística* 41(2), pp. 147-174.

PEÑA MARTÍNEZ, Gemma et al. (2014). «La anáfora conceptual en el discurso de divulgación científica en francés». *Revista Española de Lingüística*, 39(2), pp. 177-194.

RYSOVÁ, Magdaléna; RYSOVÁ, Kateřina (2021). «Primární a sekundární diskurzni konektory». In: *Slovo a slovesnost*, vol. 82, no. 3, pp. 179-208.

**Slovotvorba v básnickém překladu aneb jak dosáhnout prostřednictvím nových kompozit funkční ekvivalence (na příkladu českých překladů šibeniční poezie Christiana Morgensterna)**

Jana Pokojová (Univerzita Karlova)

V našem příspěvku se pokusíme připomenout, jak se mohou dva zdánlivě odlišné obory, jako jsou básnický překlad a jazykověda, doplňovat a vzájemně se o sebe opírat. Na příkladu překladů šibeniční (groteskní) poezie německého básníka Christiana Morgensterna (1871-1914) poukážeme na hranice přeložitelnosti, která je i v tomto specifickém případě možná, pokud se opřeme o lingvistické teorie, které vycházejí z myšlenky univerzality jakéhokoli přirozeného jazyka. Tato myšlenka předpokládá, že veškeré obsahové jevy mohou být vyjádřeny jazykovými prostředky, které nemusejí mít tentýž význam, ale fungují ve stejné situaci. Znamená to, že veškeré obsahové jevy mohou být jazykově vyjádřeny, protože každý jazyk má prostředky k tomu, aby plnil komunikační funkci. Funkční pojetí překladu pracuje s tím, že jednotlivé jazykové prostředky plní stejnou funkci, a to po všech stránkách, tedy nejen ve významově věcné (denotační, referenční), konotační (expresivní, asociační), ale i pragmatické (Knittlová 2003). Jazyk je vnímán jako nositel informace o mimojazykové skutečnosti a nositelem informace je sdělení jako celek. Jazykové prostředky všech rovin slouží k zakódování informace. Cílem překladu není reprodukovat konstitutivní prvky sdělení, nýbrž jeho funkci jako celku (Malý 2014). Pokud překlad zachovává funkci vzhledem ke komunikační situaci, lze překládat i takové lexikální jednotky, které se nekryjí, nebo vůbec nemají v druhém jazyku protějšek (Vilíkovský 2002).

Jiří Levý vyvozuje svou funkční teorii překladu ze zvláštní kategorie, kterou nazývá noetická kompatibilita. Tou zastřešuje obecné požadavky na moderní překlad jako např. výstižnost a pravdivost překladu. Překladatel vytváří iluzi, že čtenář čte původní dílo, přičemž mu jde o zachování jeho hodnoty pro příjemce (Levý 1983). Vyžadována je věrná reprodukce jednotlivých složek a zachování kompozičního principu díla, avšak při závazném zachování jeho estetické kvality, čímž se daří propojit lingvistický i literární přístup k překladu (Malý 2004).

Šibeniční básně kladou na překladatele značné tvůrčí a jazykové nároky, neboť jsou úzce spjaty s výrazovými možnostmi němčiny, ve které byly vytvořeny. Šibeniční poezie využívá různé techniky práce s jazykem, mezi něž patří homonymie, dvojsmysly, hra s gramatikou, zvuková podobnost slov, humorné interpretace slov nebo vytváření komických novotvarů často na principu kompozice.

Pro zachování hodnot německého originálu je třeba volit odpovídající domácí prostředky, což se pokusíme ukázat na příkladu tvorby nových kompozit, nejčastěji spadajících do podskupiny substantivních kompozit skladu S-K-S, popř. A-K-S sloužících především terminologickým potřebám. Navracíme se tím k produktivitě obrozenecké epochy, která přispěla řadou umělých terminologických kompozit jako např. lososopstruh, losokoz, z nichž se udržel dodnes např. mravkolev nebo lidoop (Šlosár 1999). Jsme přesvědčeni, že když překladatel vyjde ze znalostí kontrastivní česko-německé lexikologie a využije možnosti českých slovotvorných principů, lze vytvořit funkčně ekvivalentní překlad, který zachovává ideově-estetický obsah originálu (Levý 1983).

Na příkladu dvou básní (Gruselett a Neue Bildungen der Natur vorgeschlagen) a jejich českých překladů (Hiršal a Grögerová, Pokojová a Janula, Hašek, Malý) názorně ilustrujeme jazykové tvůrčí principy vytváření nových slov pomocí kompozice, které v konečném výsledku vedou k vytvoření funkčně ekvivalentního básnického překladu.

## **Bibliografie**

KNITTLOVÁ, Dagmar (2003): *K teorii i praxi překladu*. 2. vyd. Olomouc, Univerzita Palackého.

LEVÝ, Jiří (1983): *Umění překladu*. 2. vyd. Praha, Panorama.

MALÝ, Radek. (2014): *Příběhy básní a jejich překladů*. Olomouc: Univerzita Palackého v Olomouci.

MARTINCOVÁ, Olga. (2005): *Neologismy v dnešní češtině*. Praha: Ústav pro jazyk český Akademie věd České republiky.

ŠLOSAR, Dušan (1999): *Česká kompozita diachronně*. Brno: Masarykova univerzita v Brně.

ŠTÍCHA, František. *Česko-německá srovnávací gramatika*. Praha: Argo, 2003.

VILIKOVSKÝ, Ján: *Překlad jako tvorba*. 1. vyd. Praha, Ivo Železný 2002.

## Komparace české a ruské jazykové verze počítačové hry „Craftomation 101: Programming & Craft“ vyhotovené OpenAI GPT-4 v porovnání s lidským překladatelem

Mgr. Tereza Poledníková (Ústav slavistiky FF MU)

Umělá inteligence ve velké míře pronikla také do videoherního průmyslu – počítačových her. Ty jsou svou podstatou interaktivní software s prvky audiovizuálního díla. Ve srovnání s pasivním divákem neinteraktivního audiovizuálního produktu, tedy filmu či seriálu, zaujímá uživatel počítačové hry aktivní roli, přičemž je hrou vybízen k postupu a konkrétním akcím. Nevhodný překlad může hráči až znemožnit dokončit danou hru, například vlivem špatného překladu může hráči zůstat skryta podstatná informace, kterou potřebuje ke splnění určitého úkolu a dalšímu postupu ve hře [Mangiron, 2013: 45].

Proces lokalizace počítačových her zahrnuje i jiné činnosti než překladatelské, a proto je finančně nákladný. Česká republika však oproti světu představuje poměrně malý trh a vývojáři (tvůrci hry) či vydavateli by se případná investice do oficiální české lokalizované verze nemusela finančně vyplatit. V současné době proto existuje tendence od oficiálně vyhotovených překladů her do češtiny upouštět. Někteří vývojáři do svých her implementovali velké jazykové modely (LLM), pomocí nichž hra nabízí několik jazyků.

Takovým případem je i počítačová hra *Craftomation 101: Programming & Craft* ruského vývojáře vydaná v roce 2024. Díky použití OpenAI GPT-4 má hráč možnost zvolit si z třiceti dvou jazyků. Oficiální lokalizace jsou ovšem vyhotoveny pouze pro anglický a ruský jazyk, u některých jazyků má hráč možnost volby tzv. *komunitního* (fanouškovského) překladu.

Předkládaný příspěvek představuje komparativní práci zaměřující se především na výslednou jazykovou kvalitu ruské a české lokalizované verze předmětné počítačové hry. Komparaci jsou podrobeny ruská oficiální jazyková verze s ruskou jazykovou verzí vyhotovenou pomocí umělé inteligence a česká jazyková verze vyhotovená umělou inteligencí s českým komunitním překladem, který však vznikl pouhou posteditací.

Je důležité zmínit, že „expresivní substance počítačové hry začíná existovat až ve chvíli, kdy někdo vyvíjí ludickou aktivitu; pokud není hrána, zůstává ve virtuálním stavu a neaktualizuje žádný výstup. Jinými slovy, hru nelze analyzovat, aniž bychom ji hráli“ [Švelch, 2012: 35]. F. Čermák [1993: 38] upozorňuje, že „jeden problém nastává, máme-li jazykový materiál zbavený kontextu, který nutně specifikuje informaci potřebnou pro analýzu (význam a funkci) z mnoha hledisek; [...] Zvláštním, krajním případem je izolovaná věta, která je kontextu zbavena úplně [...]“. Tímto se potvrzuje skutečnost, že analýza jakékoliv počítačové hry bez vlastního průchodu (odehrání hry) není možná. Kontext při překladu videoher zaujímá důležité postavení, neboť text není překládán jako jednotný celek, ale jako jednotlivé tzv. *stringy* (textové řetězce) obsahující slovo, slovní spojení, větu či odstavce. Z tohoto důvodu je v předkládaném příspěvku pozornost zaměřena především na nevětné a polovětné textové řetězce, protože se předpokládá, že právě v těchto případech se nachází vysoká četnost pravopisné a sémantické chybovosti ruské a české jazykové verze předmětné hry vyhotovené OpenAI GPT-4.

Výsledná jazyková kvalita ruské a české lokalizované verze vyhotovená umělou inteligencí je hodnocena na základě funkčnosti překladu, tedy zachování funkce přeloženého konkrétního jazykového výrazu.

V případě komunitního překladu české lokalizované verze, posteditace jazykové verze vyhotovené OpenAI GPT-4, byla stanovena výzkumná otázka, do jaké míry je umělá inteligence skutečně schopna nahradit lidského překladatele.

## Bibliografie

ČERMÁK, František, 1993. *Základy lingvistické metodologie: Nástin hlavních principů na pozadí obecné teorie vědy*. Praha: Karolinum. ISBN 80-7066-551-3.

MANGIRON, Carmen, 2013. Subtitling in game localisation: a descriptive study. *Perspectives: Studies in Translatology* [online]. Č. 1, s. 42–56 [cit. 15.3.2024]. ISSN 1747-6623. Dostupné z: <https://www.tandfonline.com/doi/pdf/10.1080/0907676X.2012.722653?needA%20ccess=true>.

ŠVELCH, Jaroslav, 2012. Co nám říká hra: Teoretické a metodologické přístupy k počítačové hře jako expresivnímu médiu. *Illuminace: Časopis pro teorii, historii a estetiku filmu* [online]. Č. 2 (86), s. 33–48. ISSN 2570-9267. Dostupné z: <https://www.iluminace.cz/contents/ilu/2012/02.pdf>.

13

*Jazykový management  
ve vícejazyčných podnicích  
a jejich okolí*

organizují

Marek Nekula

Jiří Nekvapil

(CZE)

# Popis panelu

Cílem panelu je přispět k poznání komunikace v podnicích, v nichž se užívá několik jazyků (Nekvapil & Nekula 2006). Jednotlivé příspěvky se zaměřují jak na samotné interakce (Sieglová 2022), tak na plánování jejich charakteru (Nekvapil & Sherman 2009, 2018). Komplexnost a variabilita interní a externí komunikace těchto podniků má co do činění mj. s jejich velikostí, odvětvím, povahou aktivit i geografickým ukotvením, spojeným s určitými jazyky a kulturami. Tyto rozdíly se projevují i v rozdílech při šíření cizích jazyků a kultur do okolí podniků a následných proměnách lokální jazykové situace v obcích či městech, kde takové podniky působí (Sherman et al. 2010; Park 2022; Nekvapil & Sherman 2022). Sociolingvisticky orientovaný panel si klade otázky charakteristické pro interakční sociolingvistiku (např. distribuce moci manifestované v interakci) a sociologii jazyka (např. management vícejazyčnosti), přičemž zahrnuje i otázky prostorovosti a vizuality jazyka (linguistic landscapes). Příspěvky se orientují hlavně, i když ne výlučně na teorii jazykového managementu a konverzační analýzu.

K hlavním výzkumným otázkám patří:

- Jak souvisejí každodenní interakční události s jazykovými opatřeními cílenými na charakter interakčních událostí – reagují jazyková opatření na dění v konkrétních interakcích?
- Jak je v interakcích dosahována intersubjektivita v případě snížené jazykové kompetence komunikujících pracovníků?
- Jak se v komunikaci projevuje rozdílný společenský status různých vrstev pracovníků?
- Jak a v jaké míře ovlivňují vícejazyčné podniky jazykovou situaci ve svém okolí?
- Lze identifikovat vývojové tendence ve výše uvedených aspektech v průběhu posledních třiceti let? (historická dimenze výzkumu).
- Jaká je úloha následných interview při analýze interakčních událostí? (metodologická dimenze výzkumu).

Seznam možných témat příspěvků např. zahrnuje:

- Jazyková situace v konkrétním vícejazyčném podniku, vztah mikro a makro faktorů v něm.
- Jazykové problémy v podnikové komunikaci a způsoby jejich zvládnání; různé druhy předinterakčního managementu.
- Rozdílné úrovně zdvořilosti ve vícejazyčném podniku; jak se zde uplatňují rozdíly mezi různými typy interakčních událostí a obecně rozdíl mezi transakčním a interpersonálním jednáním.
- Jazyková charakteristika různých typů interakčních událostí v podniku; souvztažnost jazykových, tělesných a materiálních aspektů v nich.
- Analýza jazykové krajiny v konkrétním vícejazyčném podniku a jazykové krajiny v jeho okolí.



## Literatura

Nekvapil, J. & Nekula, M. (2006). On language management in multinational companies in the Czech Republic. *Current Issues in Language Planning*, 7 (2&3), 307–327.

Nekvapil, J. & Sherman, T. (2022). How multinational corporations affect their local surroundings: Sociolinguistic insights from the Czech Republic. *Working Papers in Language Management*, 6.

Nekvapil, J. & Sherman, T. (2018). Managing superdiversity in multinational companies. In A. Creese & A. Blackledge (eds.) *The Routledge Handbook of Language and Superdiversity. An Interdisciplinary Perspective*. London and New York: Routledge, Taylor & Francis Group, 329–344.

Nekvapil, J. & Sherman, T. (2009). Pre-interaction management in multinational companies in Central Europe. *Current Issues in Language Planning*, 10, 181–198.

Park, M. (2022). Metrolingualism in Prague: The adaptation and management of multilingualism in a multiethnic bakery shop. *Studie z aplikované lingvistiky / Studies in Applied Linguistics*, 13 (2), 99–122.

Sherman, T., Carl, J., Engelhardt, O., Balogh, E. & Balla, Á.T. (2010). The interaction of language schools and multinational companies in the management of multilingual practices. *European Journal of Language Policy*, 2 (2), 229–250.

Sieglová, D. (2022). Critical incidents as a window into simple language management. *Language Learning in Higher Education*, 12 (1), 105–127.

# Obsah

*Jednoduchý a organizovaný jazykový management v práci cizinců v českých herních firmách*  
Jan Houška (Univerzita Karlova)

*Difuze vícejazyčné jazykové krajiny podniků v pohraničí do okolí*  
Marek Nekula (University of Regensburg)

*Sociolingvistický výzkum nadnárodních podniků v České republice*  
Jiří Nekvapil (Univerzita Karlova)

*Jazykový management v korejsko-českých firmách: vliv nerovných mocenských vztahů*  
Minyoung Park (Hankuk University of Foreign Studies)

*Polostrukturované rozhovory jako zdroj dat v sociolingvistickém výzkumu o nadnárodních podnicích*  
Tamah Sherman (Ústav pro jazyk český AV ČR)

*Kritické incidenty v pracovní praxi studujících vysokých škol: rozvoj komunikačních kompetencí*  
Dagmar Siegllová (Škoda Auto University)

*Ukazování na technické výkresy při poradách na mnohojazyčných pracovištích*  
František Tůma (WU – Vienna University of Economics and Business)

**Jednoduchý a organizovaný jazykový management v práci cizinců v českých herních firmách**

Jan Houška (Univerzita Karlova)

Herní průmysl existuje v průniku mezi lokálním a globálním (Kerr, 2017) i s ohledem na jazyk. Ačkoliv jsou videohry kvůli mezinárodním uživatelům v angličtině, herní produkce existuje v průmyslech s národními jazyky a odlišným dějinným vývojem. Herní firmy v České republice jsou tradičně vlastněny českými zakladateli (Bigas, 2023). Stejně firmy zaměstnávají na produkční pozice cizince, působící na místě, a vzdálené pracovníky, spolupracující s týmy na dálku. Cílem příspěvku je zmapovat interakci obou skupin cizinců s místními pracovníky s ohledem na jazykové problémy jednoduchého a organizovaného managementu. Tato interakce probíhá v angličtině, češtině a rodných jazycích cizinců.

Ve výzkumu je zahrnuto 28 jedinců z 10 herních firem. 14 je ze střední a východní Evropy a 14 ze západní Evropy či Severní a Jižní Ameriky. Aktivní znalost češtiny je vyšší mezi Východoevropany, všichni jsou také mluvčí slovanských jazyků. 5 studií má lokální vlastníky a 5 zahraniční (2 z nich mají místní zakladatele). Vedoucí týmů napříč studií jsou Češi, cizinci jsou junioři a senioři. Metoda výzkumu jsou 3 polostrukturované rozhovory s každým respondentem, uskutečněné v celkovém rozmezí 6 až 8 měsíců. Následné interview nebylo realizováno kvůli profesním a etickým překážkám přístupu do studií, které znemožnily pořídit nahrávky firemních interakcí (Nekvapil & Nekula, 2006). Longitudinální studie byla vhodná kvůli fluktuaci zaměstnanců v projektově zaměřené herní produkci. Opakované rozhovory mi umožnily se doptávat na jazykové problémy s časovým odstupem (Asada, 2018).

U jednoduchého managementu jsem se soustředil na jazykové problémy ve formální a neformální komunikaci, na pracovních schůzkách, ale i teambuildingových akcích. U organizovaného managementu jsem zkoumal jazyk interní komunikace firem (e-mailů, produkční dokumentace). S ohledem na různou míru internacionalizace a odlišné problémy obou typů managementu vzešly z výzkumu 3 kategorie firem: (1) dvě firmy s dominantním užíváním češtiny, (2) čtyři firmy s jazykovým mixem mezi češtinou a angličtinou či mezi češtinou a užitím slovanských jazyků, (3) čtyři společnosti s užitím angličtiny jako lingua franca. Kategorizace je inspirována výzkumem Kraft & Lønsmann (2018). V první kategorii byla čeština korporátní jazyk. Byla jazykem dokumentace a studiových i týmových schůzek a prostředkem vylučování cizinců. Ti volili postinterakční management (Nekvapil & Sherman, 2009) a absentovali na schůzkách, které jim shrnovali čeští kolegové. Čeština v interakcích byla ale gratifikována (Neústupný, 2003) Východoevropany odkazem na její rychlejší naučení.

V druhé kategorii studií byla čeština používána na schůzkách v přítomnosti cizinců, spíše v kombinaci s angličtinou v rámci střídání jazykových kódů. Oddělené komunikační bubliny zahraničních a lokálních pracovníků vznikaly hlavně v neformálním kontextu, například při obědech. Ve dvou firmách byla respondenty zmiňovaná důležitost polštiny či ruštiny pro interpersonální komunikaci. Ve třetí kategorii studií (tj. zejména v mezinárodních pobočkách) byla korporátním jazykem angličtina. Čeština byla stále používána v interpersonální komunikaci, zejména mezi vedoucími týmů na separátních schůzkách. Studia se ale řídila pravidlem angličtiny jako jediného jazyka pracovních schůzek, když jich byl účasten cizinec. Vzhledem k mezinárodnosti firem cizinci zmiňovali nedostatečnou znalost angličtiny jako problém, který spojovali často s českými kolegy. Vzdálení pracovníci se společnostmi spolupracovali virtuálně a jejich mluvená a psaná komunikace probíhala výlučně v angličtině.

Čeština je jazyk vysokého managementu a seniorních programátorských týmů. Z rozhovorů plyne, že čeština je symbol kontinuity a úspěchů většiny zahrnutých studií, ale také komunikační problém pro západní pracovníky. Těm brání v účasti na schůzkách či v povýšení, jako forma skleněného stropu (Piekkari et al., 2005). V mezinárodních pobočkách roste důležitost angličtiny, u vzdálených pracovníků sloužící jako monolingvní rámec zkušeností, odloučených od českých studií. Český herní průmysl tak zůstává semiperiferním kontextem (Švelch, 2021), který se vztahuje skrze mezinárodní pracovníky a pobočky k centřům, ale zachovává si lokální charakter. Studie je limitovaná nezahrnutím místních pracovníků, to ale odpovídá užšímu a explorativnímu zaměření disertačního výzkumu.

## Bibliografie

Asada, H. (2018). The longitudinal management of speech styles by overseas and local students in a Japanese university context. V L. Fairbrother, J. Nekvapil, & M. Sloboda (Eds.), *The Language Management Approach: A Focus on Research Methodology* (pp. 283–303). Peter Lang.

Bigas, J. (2023, prosinec 11). *V Česku působí 155 herních studií*. Vortex. <https://www.vortex.cz/vcesku-pusobi-155-hernich-studii/>

Kerr, A. (2017). *Global games: Production, circulation and policy in the networked era*. Routledge/Taylor & Francis Group.

Kraft, K., & Lønsmann, D. (2018). A language ideological landscape: The complex map in international companies in Denmark. V T. Sherman & J. Nekvapil (Eds.), *Volume 5 English in Business and Commerce* (pp. 46–72). De Gruyter Mouton.

Nekvapil, J., & Nekula, M. (2006). On Language Management in Multinational Companies in the Czech Republic. *Current Issues in Language Planning*, 7(2–3), 307–327. <https://doi.org/10.2167/cilp100.0>

Nekvapil, J., & Sherman, T. (2009). Pre-interaction management in multinational companies in Central Europe. *Current Issues in Language Planning*, 10(2), 181–198. <https://doi.org/10.1080/14664200802399133>

Neustupný, J. V. (2003). Japanese students in Prague. Problems of communication and interaction. *International Journal of the Sociology of Language*, (162), 125–143. <https://doi.org/10.1515/ijsl.2003.033>

Piekkari, R., Vaara, E., Tienari, J., & Säntti, R. (2005). Integration or disintegration? Human resource implications of a common corporate language decision in a cross-border merger. *The International Journal of Human Resource Management*, 16(3), 330–344. <https://doi.org/10.1080/0958519042000339534>

Švelch, J. (2021). Promises of the Periphery: Producing Games in the Communist and TransformationEra Czechoslovakia. V O. Sotamaa & J. Švelch (Eds.), *Game Production Studies* (pp. 237–256). Amsterdam University Press.

**Difuze vícejazyčné jazykové krajiny podniků v pohraničí do okolí**

Marek Nekula (University of Regensburg)

Zdroje vícejazyčnosti v podnicích, včetně vícejazyčnosti její jazykové krajiny, jsou různé: management expandujících podniků ekonomických mocností (Matras 2009: 47) či mezinárodně aktivních podniků, migrace či přeshraniční „pendlování“ zaměstnanců nebo orientace na mezinárodní či přeshraniční klientelu (literaturu pro český kontext shrnuje mj. Vasiljev 2017). Takto členitá je i jazyková situace v česko-německém pohraničí, kterou charakterizuje asymetrie v osvojování a užívání jazyků sousedních států (mj. Holly et al. 2003, Müller & Fráně 2023, Nekula 2024). Příspěvek detailněji analyzuje jazykovou krajinu v jednom podniku se státní účastí v bavorské části česko-bavorského pohraničí i šíření vícejazyčné jazykové krajiny do okolí (Nekvapil & Sherman 2022), resp. do okolních obcí. Jazyková krajina (Mensel et al. 2016) v podniku je vyložena jako výsledek organizovaného jazykového managementu zdola (Nekvapil 2009), argumentovaného jazykovými ideologiemi. Při jejím etablování hraje roli finanční podpora evropských strukturálních fondů i fraktální rekurzivita jejich jazykových ideologií vícejazyčnosti (jazykové ideologie a jejich terminologii shrnuje mj. Busch 2017). Tento model se „přelévá“ do realizace komunálních projektů v okolí zmíněného podniku a uplatňuje se prostřednictvím projektové infrastruktury „ostrůvkovitě“ i v jiných infrastrukturních projektech v pohraničí. Studie vychází z terenního výzkumu a interview s aktéry proměny jazykové krajiny a je pojata jako příspěvek k výzkumu tzv. linguistic borderscapes, překračování jazykové hranice a otázek jazykové deteritorializace (mj. Kimura 2017, Nekula 2021).

## Bibliografie

Busch, B. (2017): *Mehrsprachigkeit*. Wien: Facultas.

Holly, W., Nekvapil, J., Scherm, I. & Tišerová, P. (2003). Unequal neighbours: coping with asymmetries. *Journal of Ethnic and Migration Studies* 29, 819–834.

Kimura, G. Ch. (2017): Signs of de-territorialization? Linguistic landscape at the German-Polish border. *Eurasia Border Review* 8, 45–58.

Matras, Y. (2009). *Language Contact*. Cambridge: Cambridge University Press.

Müller, K. B. & Fráně, L. (2023): Borders and language. Minor misunderstandings, big troubles, and the fruits of multilingualism. In: K. B. Müller (Hg.), *Active Borders in Europe: Identity and Collective Memory in the Cross-Border Space*. Cham: Springer, 59–99.

Mensel, L. v., Vandenbroucke, M. & Blackwood, R. (2016). Linguistic landscapes. In: O. García, N. Flores, M. Spotti (eds), *Oxford Handbook of Language and Society*. Oxford: Oxford University Press, 423-449.

Nekula, M. (2021). Grenzen und Sprachgrenzen in der Sprachwissenschaft. In: M. Klessmann, D. Gerst, H. Krämer (eds), *Grenzforschung: Handbuch für Wissenschaft und Studium*. Baden-Baden: Nomos, 400–418.

Nekula, M. (2024). Tschechisch und Tschechien. In: S. Höder, Ph. Krämer (eds), *Nachbarsprachen – Sprachnachbarn: Mehrsprachigkeit und Sprachpolitik in Deutschlands Grenzregionen*. Baden-Baden: Nomos Verlag, [https://www.researchgate.net/publication/373236911\\_Nachbarsprache\\_Tschechisch](https://www.researchgate.net/publication/373236911_Nachbarsprache_Tschechisch)

Nekvapil, J. (2009). The integrative potential of Language Management Theory. In: J. Nekvapil, T. Sherman (eds), *Language management in contact situations: Perspectives from three continents*. Frankfurt am Main: Peter Lang, 1–11.

Nekvapil, J. & Sherman, T. (2022). How multinational corporations affect their local surroundings: Sociolinguistic insights from the Czech Republic. *Working Papers in Language Management*, 6.

Vasiljev, I. (2017): Jazyk a ekonomika. In: Petr Karlík, Marek Nekula, Jana Pleskalová (eds), *CzechEncy - Nový encyklopedický slovník češtiny*.

URL: [https://www.czechency.org/slovník/JAZYK\\_A\\_EKONOMIKA](https://www.czechency.org/slovník/JAZYK_A_EKONOMIKA) (poslední přístup: 22. 3. 2024)

## Sociolingvistický výzkum nadnárodních podniků v České republice

Jiří Nekvapil (Univerzita Karlova)

V druhé polovině roku 1994 jsem spolu s dalšími dvěma kolegy uskutečnil empirický výzkum zaměřený na jazykovou, komunikační a sociokulturní situaci v mladoboleslavské automobilce Škoda (viz Höhne & Nekula, 1997; Nekvapil, 1997a, 1997b, 1999, 2000). Od té doby uplynulo 30 let. Ve své prezentaci se chci věnovat bilanci toho, co se na poli sociolingvistického výzkumu nadnárodních podniků operujících v ČR dosáhlo, a upozornit na to, co se nadále jeví jako desideratum. Zabývat se budou společenskými podmínkami, které tento výzkum umožnily a stimulovaly – totiž samotným vznikem a existencí nadnárodních podniků v ČR, dále utvářející se novou jazykovou a komunikační situací a konečně stavem sociolingvistického bádání v ČR. Tyto tři analytické dimenze sleduji i ve svém výkladu orientovaném na budoucí výzkum.

**Bibliografie**

Höhne, S. & Nekula, M. (eds.) (1997). *Sprache, Wirtschaft, Kultur: Deutsche und Tschechen in Interaktion*. München: Iudicum.

Nekvapil, J. (1997a). O komunikačním překonávání česko-německé etnické polarizace. *Přednášky z XL. běhu LŠSS*. Praha: Univerzita Karlova, 43-57.

Nekvapil, J. (1997b). Die kommunikative Überwindung der tschechisch-deutschen ethnischen Polarisierung. *Deutsche, deutsche Kollegen, Expatriates und andere soziale Kategorien im Automobilwerk Škoda*. In: S. Höhne & M. Nekula (eds.), *Sprache, Wirtschaft, Kultur: Deutsche und Tschechen in Interaktion*. München: Iudicum, 127–145.

Nekvapil, J. (1999). *Sprechen über Personen: Soziale Kategorisierung im tschechisch-deutschen Kontakt*. Trainingsmodul. Hof: Berufliche Fortbildungszentren der Bayerischen Wirtschaft.

Nekvapil, J. (2000). *Mluvení o osobách: sociální kategorizace v česko-německém kontaktu*. Tréninkový modul. Hof: Berufliche Fortbildungszentren der Bayerischen Wirtschaft.

**Jazykový management v korejsko-českých firmách: vliv nerovných mocenských vztahů**

Minyoung Park (Hankuk University of Foreign Studies)

Cílem této přednášky je prozkoumat různé formy jednoduchého a organizovaného managementu v mezinárodních korejsko-českých firmách v České republice. Konkrétním cílem je pozorovat, jak mocenské vztahy (Neustupný, 2002) mezi zaměstnanci ovlivňují užívání jazyka a jeho management v kontaktních situacích (Neustupný, 1985), v nichž uživatelé jazyka nesdílejí mateřský jazyk, ale užívají angličtinu jako lingua franca. V této přednášce jsou jako výzkumné subjekty vybrány čtyři mezinárodní korejsko-české firmy, které jsou rozděleny do čtyř skupin v souladu s kategorizací mezinárodních společností, kterou představili Peltokorpi a Vaara (2012). K objasnění toho, jak jednotliví zaměstnanci a zahraniční dceřiné firmy jako agenti s autoritou manažují a ovlivňují každodenní užívání jazyka (např. užívání jazyka ve zprávách), jsou využity metody kvalitativního výzkumu a teorie jazykového managementu, jako jsou pozorování, fotodokumentace, nahrávání přirozeně probíhajících interakcí a různé druhy rozhovorů (např. polostrukturované rozhovory, interakční interview a následné rozhovory). Pět výzkumných otázek je položeno: 1) Jak je manažována interakce druhého jazyka v nevyvážené komunikační situaci (např. místní zaměstnanci vs. expatriati nebo poskytovatelé služeb a produktů vs. zahraniční zákazníci) na mikroúrovni? 2) jak je organizována a formulována oficiální jazyková politika ve firmě? 3) jak firma manažuje potenciální komunikační problémy, s nimiž se setkávají korejští a místní zaměstnanci? 4) jak firma manažuje potenciální komunikační problémy v interakci zaměstnanců korejských i místních se zahraničními zákazníky? 5) jak firma manažuje potenciální sociokulturní problémy v interakci mezi korejskými a místními zaměstnanci? Tyto výzkumné otázky jsou zodpovězeny v rámci teorie jazykového managementu, která se zabývá řešením komunikačních problémů z hlediska jazykového, sociokulturního a socioekonomického.

**Bibliografie**

Neustupný, J. V. (1985). Problems in Australian-Japanese contact situations. In J. B. Pride (Ed.), *Cross-Cultural Encounters: Communication and Mis-communication* (s. 44-64). Melbourne: River Seine Publication.

Neustupný, J. V. (2002). Language and power into the 21<sup>st</sup> century. A paper for the International Conference on Language and Empowerment, Malaysian Association of Modern Languages, Kuala Lumpur, Malaysia, April 11-13, 2002, s. 1-19.

Peltokorpi, V. & Vaara, E. (2012). Language policies and practices in wholly owned foreign subsidiaries: A recontextualization perspective. *Journal of International Business Studies*, 43, s. 808-833.



**Polostrukturované rozhovory jako zdroj dat  
v sociolingvistickém výzkumu o nadnárodních podnicích**

Tamah Sherman (Ústav pro jazyk český AV ČR)

Možnost provést výzkum o jazykových praktikách na pracovišti je do značné míry závislá na přístupu, který výzkumník má k zaměstnancům i k pracovnímu prostoru. Z tohoto důvodu je jednou z nejčastěji využívaných výzkumných metod při výzkumu o podnicích polostrukturovaný rozhovor. Při tomto typu rozhovoru výzkumník získává informace od zaměstnanců v prostředí, které je v různé míře odděleno od skutečného průběhu pracovních činností, a může být vedeno online nebo po telefonu (srov. Sanden 2020) nebo dokonce fyzicky mimo pracoviště (Park 2013). Ačkoli obvykle bývá polostrukturovaný rozhovor jen jednou součástí smíšeného souboru metod (srov. mj. Kim & Angouri 2023, Lønsmann & Kraft 2018), které mohou zahrnovat také nahrávky interakcí, dotazníky, zúčastněné pozorování a analýzu dokumentů, je také robustním a všestranným zdrojem dat, který lze analyzovat pomocí více kvalitativních metod.

Na základě souboru více než 40 polostrukturovaných rozhovorů vedených mnou a mými kolegy v nadnárodních podnicích s hlavním sídlem v Německu a Jižní Koreji v letech 2007 až 2022 si tento příspěvek klade následující otázky: 1) Jaké jsou možnosti a limity polostrukturovaného rozhovoru při popisu a analýze používání jazyka a jeho managementu v nadnárodních podnicích? 2) Které situační faktory tyto možnosti ovlivňují? 3) Jaké jsou nejužitečnější a nejdostupnější metody triangulace? Nejprve se zabývám některými otázkami spojenými se samotným rozhovorem, včetně sociální, etnické, genderové, jazykové a profesní identity tazatele i tázaného, ad hoc volby jazyka rozhovoru, časových, prostorových a soukromých omezení a práce s tlumočníky. Poté se věnuju řadě výzev, s nimiž se analytik následně potýká, např. prezentace a kontextualizace výpovědí z rozhovoru, rozpory mezi postupy deklarovanými v rozhovoru a postupy pozorovanými v terénu a rozdílné pohledy na tutéž situaci, jak ji popisuje více dotazovaných.

### **Bibliografie**

- Kim, K. & Angouri, J. (2023). "It's hard for them to even understand what we are saying" (.). Language and power in the multinational workplace. *Critical Perspectives on International Business* 19(1), 27-45.
- Lønsmann, D. & Kraft, K. (2018). Language policy and practice in multilingual production workplaces. *Multilingua* 37(4), 403-427.
- Park JS-Y. (2013). Metadiscursive regimes of diversity in a multinational corporation. *Language in Society* 42(5), 557-577.
- Sanden, G. R. (2020). Domain loss theory revisited: From multinational corporations to multicorporate nations. *International Journal of Applied Linguistics* 30(3), 510-529.

## Kritické incidenty v pracovní praxi studujících vysokých škol: rozvoj komunikačních kompetencí

Dagmar Sieglová (Škoda Auto University)

V souvislosti se společenským a technologickým vývojem globálně propojeného světa se mění kvalifikační požadavky trhu práce na absolventy vysokých škol. Jazykové a komunikační dovednosti jsou proto spolu s technickými dovednostmi považovány za klíčové pro pracovní trh 21. století (Council of the EU, 2014). Na tuto skutečnost reagují i tvůrci národních i mezinárodních vzdělávacích politik při úpravě vzdělávacích strategií a plánů (MŠMT ČR, 2020). Rozvoji jazykových a komunikačních dovedností studujících vysokých škol v prostředí povinných pracovních stáží je věnován předložený příspěvek. Jeho cílem je představit pojem kritický incident (Flanagan 1954, in Butterfield et al., 2005; Fitzgerald et al., 2007) a prostřednictvím analýzy vybraných ukázek poukázat na potenciál kritických událostí jako nástroje pro výzkum a propojení teorie s praxí.

V prezentaci vycházím z výzkumu realizovaného v rámci jazykového vzdělávání Škoda Auto Vysoké školy od roku 2016 do současnosti. Pozornost soustředím na práci s kritickými událostmi zpracovanými formou písemných narativů, kterých byli přímými účastníky studující vysoké školy. Na základě utřídění a provedené tematické analýzy (Braun & Clarke, 2006) datového souboru čítajícího přes 350 nashromážděných narativů představím rozbor situací z kategorie studentských pracovních stáží v nadnárodních korporacích v ČR a v zahraničí (Sieglová & P. Příbramská, 2020). Akcentovat budu především komunikační aspekty vybraných situací. Prostřednictvím rozboru vybraných ukázek po stránce emoční, kognitivní a behaviorální popíšu kritickou událost jako cyklický proces (Sieglová, 2023) přispívající k rozvoji komunikačních a interkulturních dovedností (Sieglova et al., 2022). Popsaný cyklus kritických incidentů zasadím do kontextu teorie jazykového managementu (Sieglová, 2022) a navrhnou možné implikace získaných poznatků v celospolečenských souvislostech.

### Bibliografie

- Braun, V. a & V. Clarke. (2006). **Using Thematic Analysis in Psychology**. *Qualitative Research in Psychology*, 3 (1), 77–101.
- Butterfield, L. D., Borgen, W. A., Amundson, N. E. & Maglio, A. S. T. (2005). **Fifty Years of the Critical Incident Technique: 1954–2004 and Beyond**. *Qualitative Research*, 5 (4), 475–497.
- Council of the EU. (2014). **Conclusions on multilingualism and the development of language competences**. *Education, Youth, Culture and Sport Council meeting Brussels 20 May 2014*.
- Fitzgerald, K., Seale, S., Kerins, C. A. & R. McElvaney. (2007). **The Critical Incident Technique: A Useful Tool for Conducting Qualitative Research**. *Journal of Dental Education*, 72 (3), 299–304.
- MŠMT ČR (2020). **Strategie vzdělávací politiky České republiky do roku 2030+**. *Ministerstvo školství, mládeže a tělovýchovy ČR*.
- Sieglová, D. (2022). **Critical incidents as a window into simple language management**. *Language Learning in Higher Education*, 12 (1), 105–127.
- Sieglová, D. (2023). **A multi-disciplinary use of critical incidents: A didactic model for personal growth**. *Studies in Applied Linguistics 2023*, 14 (1), 41–61.
- Sieglová, D. & P. Příbramská. (2021). **Work Internship Abroad as an Accelerator of Social Mobility and Life-Long Learning**. In T. Löster & T. Pavelka (Eds.). *The 15th International Days of Statistics and Economics. Conference Proceedings*. Praha: Melandrium, 913–924.
- Sieglová, D., Rammer, V., Vuokila, A., Soukupová, V., Gaisch, M., Ala-Sankila, R., Stejskalová, L., Fetscher, D., de Martino, M., Klein, S. & Pompella, M. (2022). **The Potential of Critical**

**Incidents: A Cross-Cultural Comparison.** In *Proceedings CCBC 2022*. Düren, Germany: Shaker Verlag, 235–246.

## Ukazování na technické výkresy při poradách na mnohojazyčných pracovištích

František Tůma (WU – Vienna University of Economics and Business)

Migrace, globalizace a potřeba podniků zaměstnat kvalifikované pracovníky a zefektivnit výrobu způsobují, že se na pracovištích běžně setkávají lidé s různými úrovněmi osvojení různých jazyků a s různorodým kulturním zázemím. I přesto předmětem většiny publikovaných analýz interakcí na pracovišti zůstává jednojazyčná interakce (srov. Hazel & Svennevig, 2018; Tietze et al., 2021). Existující studie zabývající se interakcemi na pracovišti navíc cílí převážně na používání samotného jazyka, přestože stále více roste povědomí o tom, že účastníci k utváření svého jednání používají vedle verbálních prostředků také prostředky tělesné (např. ukazovací gesta, viz Kendon, 2004) a materiální (např. nástroje spojené s výkonem specifické činnosti na pracovišti, srov. Goodwin, 2018). Předkládaný příspěvek přispívá do odborné diskuse o povaze interakcí na pracovišti tím, že představuje zjištění na podkladě nahrávek z mnohojazyčných prostředí se zaměřením na multimodální aspekty zaznamenaných interakcí.

Studie vychází z datového souboru čtyř hodin videonahrávek dyadických nebo skupinových interakcí probíhajících především v kancelářských prostorách dvou malých podniků zaměřených na materiální výrobu. V obou firmách pracují zaměstnanci z různých zemí; mezi jazyky zachycené na nahrávkách patří němčina, čeština, angličtina a srbština. Pro plánování a realizaci výroby používají zaměstnanci v obou podnicích technické výkresy, což se projevuje také na jejich společných schůzkách. Prvotní prohlížení nahrávek odhalilo, že účastníci se k výkresům vztahují použitím řady ukazovacích gest (např. ukazování nataženým ukazováčkem nebo propiskou, obojí s dlaní v supinaci nebo pronaci, poklepání na výkres, krouživé pohyby nad oblastí výkresu, výrazné změny směru pohledu apod.), která někdy doplňují vpisováním do výkresu.

Aby bylo možné rozkrýt, jak účastníci koordinují používání různých jazykových prostředků, ukazování na technické výkresy a dalších tělesných prostředků k utváření svého jednání v konkrétních situacích, byly sekvence, kde mluvčí na technické výkresy ukazují, přepsány a analyzovány za použití principů multimodální konverzační analýzy (Goodwin, 2018). Podrobná sekvenční analýza byla vedena následujícími výzkumnými otázkami: Jak účastníci v interakcích na mnohojazyčném pracovišti koordinují ukazovací gesta a další sémiotické prostředky? Jakou roli v těchto interakcích hrají technické výkresy?

Analýza dokládá, že účastníci (včetně mluvčích s omezenou úrovní osvojení používaného jazyka) produkováním velmi prostých jazykových prostředků, jako jsou např. deiktické výrazy a jednoduché věty, utvářejí relativně komplexní jednání, které dotvářejí použitím vhodného ukazovacího gesta. Zatímco ukazovací gesta s dlaní v pronaci mluvčí používají k určení oblasti nebo objektu, ke kterým v dané sekvenci poskytují nebo požadují informace, ukazovací gesta s dlaní v supinaci dodávají výslednému jednání evaluační rozměr: i pokud jsou tato gesta produkována společně s oznamovací větou, obsahuje následné jednání komunikačního partnera hodnocení ve vztahu k části výkresu, na kterou první mluvčí ukazuje a o které hovoří.

Uvedená zjištění jsou důležitá pro další diskuse o interakcích na jazykově a kulturně pestrém pracovišti (k přehledům výzkumu viz např. Piekkari et al., 2022; Tietze et al., 2021). Technické výkresy usnadňují probíhající interakce tím, že poskytují účastníkům referenční body, ke kterým se mohou různými ukazovacími gesty vztahovat a společně s relativně jednoduchými jazykovými prostředky se tak dorozumět. Dalším specifickým zjištěním je, že účastníci se v interakci obvykle vizuálně zaměřují na technické výkresy, přičemž oční kontakt navazují minimálně. Jedná se tedy o specifický typ interakce, který je značně odlišný od každodenních rozhovorů, kde oční kontakt hraje podstatnou roli v koordinaci střídání mluvčích (např. Auer, 2021). Lze shrnout, že v případě mnohojazyčného prostředí, kde používaný jazyk nemusejí všichni mluvčí plně ovládat, hrají materiální a tělesné prostředky velmi důležitou roli.

## Bibliografie

- Auer, P. (2021). Turn-allocation and gaze: A multimodal revision of the “current-speaker-selects-next” rule of the turn-taking system of conversation analysis. *Discourse Studies*, 23(2), 117–140. <https://doi.org/10.1177/1461445620966922>
- Goodwin, C. (2018). *Co-operative action*. Cambridge University Press.
- Hazel, S., & Svennevig, J. (2018). Multilingual workplaces – Interactional dynamics of the contemporary international workforce. *Journal of Pragmatics*, 126, 1–9. <https://doi.org/10.1016/j.pragma.2017.11.005>
- Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge University Press.
- Piekkari, R., Gaibrois, C., & Johansson, M. (2022). A review of language-sensitive research in international business: A multi-paradigmatic reading. *Journal of Comparative International Management*, 25(1), 144–174. <https://doi.org/10.55482/jcim.2022.32906>
- Tietze, S., Back, H., & Piekkari, R. (2021). Managing communication in multilingual workplaces. In F. Cooren & P. Stücheli-Herlach (Ed.), *Handbook of Management Communication* (s. 373–388). De Gruyter. <https://doi.org/10.1515/9781501508059-020>

14

*Aktuální otázky české  
lexikologie a lexikografie*

organizují

Božana Niševa

Katja Brankačkec

Michal Vašíček

(CZE)

# Popis panelu

Panel oslovuje jak lexikologicky, tak lexikograficky zaměřené badatele. Tématem mohou být teoretické a metodologické otázky týkající se diachronně i synchronně orientovaného popisu slovní zásoby češtiny a / nebo jiných jazyků; teoreticko-metodologické koncepce různých typů (jedno- a vícejazyčných, všeobecných výkladových, speciálních, valenčních aj.) slovníků (srov. např. Niševa – Blažek – Skwarska 2017; Niševa et al. 2018; Krejčířová 2023); přednosti a úskalí využití umělé inteligence při tvorbě slovníků; technická řešení zveřejnění slovníku na internetu (tvorba nových elektronicky založených slovníkových databází, retrodigitalizace tištěných vydání aj.; srov. např. Bongalon et al. 2021); problematika recepce slovníků různého typu (viz např. Mžourková – Křivan 2019). Ke konkrétnějším lexikografickým otázkám, na které bychom chtěli zaměřit pozornost, patří např. principy budování elektronických korpusů „šitých na míru“, tedy pro určitý lexikografický projekt, nebo tvorba neologických materiálových základů s přihlédnutím k specifickým rysům tištěných slovníků neologismů a / nebo neologických databází. Rádi též poskytneme prostor pro představení lexikografických projektů s gramatickým zaměřením (např. valenčních slovníků), pro obeznámení širší odborné veřejnosti s klasickými a moderními přístupy v teorii a praxi etymologické a překladové lexikografie, frazeografie, terminografie aj. S ohledem na prudký vývoj jazykových aplikací založených na umělé inteligenci je čím dál důležitější i tzv. zpětná vazba poskytující údaje o práci uživatelů s informacemi ve slovnících a o způsobech, kterými je dále využívají. Vítány jsou proto též metalexikografické příspěvky a příspěvky o recepci slovníků a vývoji uživatelského chování (viz např. Abel – Klosa 2014). V souvislosti s tím se soustředíme i na problematiku komunikace lexikografa s uživatelem, resp. zda ji pojímat jako výhodu, nebo břímě – a obecněji – o úloze lexikografie v současném světě.

Cílem je propojení lexikologů a lexikografů z českých akademických a/nebo univerzitních pracovišť, příp. též představitelů komerčních lexikografických projektů. Očekávaným výsledkem je vzájemná výměna zkušeností z řešení konkrétních lexikologických a lexikografických úkolů, na nichž účastníci panelu momentálně pracují, popularizace jejich stávající činnosti a navázání další spolupráce.

Téma navrhovaného panelu jsme volili se zřetelem k činnosti oddělení slavistické lingvistiky a lexikografie Slovanského ústavu AV ČR, v. v. i. a Lexikologicko-lexikografické sekce Jazykovědného sdružení ČR.

## Literatura

ABEL, Andrea – KLOSA, Annette (eds.). *Der Nutzerbeitrag im Wörterbuchprozess. 3. Arbeitsbericht des wissenschaftlichen Netzwerks „Internetlexikografie“*. (OPAL 4/2014) – Mannheim: Institut für Deutsche Sprache, 2014.

BONGALON et al. Using Open-Source Tools to Digitise Lexical Resources for Low-Resource Languages. In: Kosem, Iztok et al. (eds.): *Electronic lexicography in the 21st century: post-editing lexicography. Proceedings of the eLec 2021 conference*. Brno: Lexical Computing CZ s. r. o., 2021.

KREJČÍŘOVÁ, Iveta. *Dvojjazyčná lexikografie a SLÚ AV ČR*. Dostupné z WWW <http://www.vedakolemnas.cz/sys/galerie-download/VKN-128.pdf>

MŽOURKOVÁ, Hana – KŘIVAN, Jan. Pust' me uživatele do slovníku! Aneb o neprobádané cestě v české lexikografii. *Naše řeč* 102 (2019), č. 1–2.

NIŠEVA, Božana – SKWARSKA, Karolína – BLAŽEK, David (eds.). *Moderní slovanská lexikografie*. Praha: Slovanský ústav AV ČR, v. v. i., 2017.

NIŠEVA, Božana et al. (eds.). *Slovanská lexikografie počátkem 21. století: sborník příspěvků z mezinárodní konference, Praha 20.–22. 4. 2016*. Praha: Slovanský ústav, v. v. i., 2018.



# Obsah

*Metodika exemplifikace ve Slovinsko-českém slovníku*

David Blažek (Slovanský ústav AV ČR)

*Víceslovné lexémy v hornolužicko-českém slovníku Jiřího Mudry*

Katja Brankačec (Slovanský ústav AV ČR), Michal Škrabal (Ústav Českého národního korpusu FF UK)

*Staročeská účelová adjektiva a jejich významové definice ve slovníku*

Irena Fuková (Ústav pro jazyk český AV ČR)

*Praktické zkušenosti s převodem slovníků do formátu TEI Lex-0*

Boris Lehečka (Moravská zemská knihovna v Brně)

*Retrodigitalizace Bulharsko-českého slovníku Karla Hory a kolektivu SLÚ ČSAV: teoretické a metodologické aspekty přípravy podkladů pro bulharsko-českou slovníkovou databázi*

Božana Niševa<sup>1</sup>, Štefan Pilát<sup>1</sup>, Michaela Kuzmova<sup>2</sup> (<sup>1</sup>Slovanský ústav AV ČR, <sup>2</sup>Ústav bulharského jazyka prof. L. Andrejčina, BAV)

*Exemplifikace ve Slovníku nářečí českého jazyka: zdroje, postupy, problémy*

Marta Šimečková (Ústav pro jazyk český AV ČR)

*Komunikační funkce ve slovníku? Tak určitě! (Ke zpracování jistotních částic v češtině)*

Barbora Štěpánková, Lucie Poláková, Jana Šindlerová (Charles University)

*CzeDLex – Slovník českých diskurzivních konektorů*

Pavčina Synková<sup>1,2</sup>, Jiří Mírovský<sup>1</sup>, Lucie Poláková<sup>1</sup> (<sup>1</sup>ÚFAL MFF UK, <sup>2</sup>ÚČJTK FF UK)

*Specifika a perspektivy nářeční lexikografie při tvorbě lexikálních databází*

Michal Vašíček (Slovanský ústav AV ČR), Darja Vašíčková (Filozofická fakulta UK v Praze, ÚVES)

## Metodika exemplifikace ve Slovinsko-českém slovníku

David Blažek (Slovanský ústav AV ČR)

Při přípravě nového Slovinsko-českého slovníku středního rozsahu (25 000 hesel) ve Slovanském ústavu (2002–2023) byl kladen velký důraz mj. na dostatečnou exemplifikaci s cílem poskytnout uživateli slovníku pomoc při zasazování informací z heslového odstavce do konkrétního kontextu a naznačit zapojení lexému do slovních kombinací či syntaktických celků. Lze hovořit také o receptivní a produktivní funkci *exempel* (např. Kosem: 321), přičemž u slovníku dekodovacího, jakým Slovinsko-český slovník pro české uživatele je, vystupuje do popředí zejména role receptivní. Ve slovníku jsou proto nedílnou součástí *exempel* i jejich ekvivalenty v češtině (mj. i z důvodu příbuznosti obou jazyků a z ní plynoucích rizik mezijazykové homonymie či paronymie).

Elektronická verze slovníku umožní uvedení téměř neomezeného množství příkladů, avšak u verze knižní je nutné brát v potaz určitou prostorovou úspornost. Hlavní snahou při výběru příkladů tak bylo hledisko dostatečné reprezentativnosti příkladů, což bylo posuzováno podle nejobvyklejší kolokability (zejména) v korpusových zdrojích. Tímto autoři zároveň vyhověli i hledisku autentičnosti příkladů a vyhnuli se zařazení *exempel* umělých, resp. ryze teoretických. Dalším hlediskem při volbě příkladů byla snaha o pokrytí nejvyššího možného počtu aspektů lexému, tj. dostatečné ilustrování jejich ohýbání, valence či komunikační funkce. Pro dosažení tohoto cíle zde jsou v citovaném slovníku uváděna slovní spojení volná a pevná, případně polovětné i větné konstrukce. Od *exempel* ilustrujících využití lexémů v jejich základním významu se přechází k příkladům použití ve významu přeneseném (např. terminologická či frazeologická spojení).

Dalším aspektem výběru *exempel* byla jejich jazyková správnost a srozumitelnost. Autoři tedy usilovali o eliminaci příkladů, které by pro uživatele ve svém omezeném kontextu mohly být matoucí nebo nejednoznačné. V neposlední řadě byla při volbě ilustrativního jazykového materiálu ve slovníkových heslech zohledněna otázka společensky citlivých exemplifikačních aspektů (Lišková: 27).

Autoři se proto snažili v *exemplech* vyhnout negativním společenským stereotypům (genderovým, národnostním, náboženským ap.), existovala-li k nim dostatečně autentická a frekventovaná alternativa (např. *neapolská mafie* vs. *Neapolský záliv* u hesla *neapolský*).

V předneseném příspěvku budou jednotlivá hlediska výběru *exempel* do slovníku doplněna konkrétními příklady hesel, která byla při korekturách slovníku vyřazena nebo musela být upravena z důvodu nesplnění některého z výše popsaných aspektů.

## **Bibliografie**

Čermák, F.: Překladová lexikografie. In: Čermák, F. – Blatná, R. (eds.): Manuál lexikografie. H&H, Jinočany 1995, s. 230–248.

Kosem, I.: Slovarski zgledi. In: Gorjanc, V. et al. (eds.): Slovar sodobne slovenščine: Problemi in rešitve. Znanstvena založba Filozofske fakultete Univerze v Ljubljani, Ljubljana 2015, s. 321–338.

Lišková, M.: Exemplifikace v Akademickém slovníku současné češtiny. In: Niševa, B. et al. (eds.): Slovanská lexikografie počátkem 21. století. Slovanský ústav, Praha 2018, s. 25–31.

Tichá, Z.: Funkce příkladů v slovníku neologizmů. In: Ondrejovič, S. – Považaj, M. (eds.): Lexicographica '99. Veda, Bratislava 2001, s. 141–145.

## Víceslovné lexémy v hornolužicko-českém slovníku Jiřího Mudry

Katja Brankačec (Slovanský ústav AV ČR), Michal Škrabal (Ústav Českého národního korpusu FF UK)

Digitalizace a dokončení slovníkového torza Jiřího Mudry (podrobněji k cílům projektu viz Škrabal – Brankačec 2021) přináší řadu metodologických i jazykových otázek. V původním Mudrově materiálu jsou, mezi jinými problémy, velmi nekonzistentně řešeny víceslovné lexémy. Ojedinele je tu najdeme jako samostatná hesla, jde vesměs o jednotky depropriální povahy: např. *bunsenowy palak* ‚Bunsenův hořák‘. Většinu víceslovných hesel však Mudra zpracovává jakožto hnízdovaná podhesla, aniž by je lišil od různých příkladů, valenčních vazeb, frazémů a kolokací (obr. 1, 2).

To je vcelku pochopitelné vzhledem k době vzniku i chystanému papírovému vydání slovníku, dnes však není zapotřebí se těchto omezení držet. Rozhodli jsme se proto v rámci digitalizace Mudrova materiálu rozlišovat v mikrostruktuře hesla, potažmo makrostruktuře slovníku 1) kolokace (krátké příklady kolokability příslušných lexémů), 2) exemplifikace (delší citace, zejména z korpusů: Serbski Institut Budyšin 2021; Brankačec – Vavřín – Zasina 2023) a 3) víceslovné lexémy (lexikalizovaná sousloví); frazeologii – jakožto specifický segment lexikonu – ponecháváme zatím v aktuální fázi stranou. První dvě skupiny zachováváme v rámci konkrétního hesla, zatímco víceslovné lexémy zpracováváme jako samostatná hesla, na něž v heslech relevantních komponentů figuruje odkaz. K tomuto kroku jsme byli vedle praktičnosti a konzistence (změny ve víceslovném hesle stačí zanést na jedno místo, nikoliv do více hesel) motivováni též lepší vyhledatelností těchto pojmů ve slovníkové aplikaci.

Otázkou však zůstává, do jaké míry jsou nejrůznější kolokace v lužické srbštině lexikalizované. V jediném dostupném korpusu Hotko v2 (Serbski Institut Budyšin 2021), podobně jako v lužickosrbských slovnících, najdeme často různé varianty víceslovných nebo i jednoslovných derivátů pro pojmy, které mají ve větších jazycích pevně lexikalizovaný jednotný tvar. Tak jsou např. doložena tato adjektiva se stejnou sémantikou: *akcijowy / akcijny / akciski*. Situace u mnohých adjektiv se navíc komplikuje tím, že všechna jsou doložena v různých významech, konkrétně v tomto případě ‚akční‘ a ‚akciový‘ – od všech jsou doloženy příklady, nicméně s různou frekvencí. Následně má i český pojem *akciová společnost* v lužické srbštině dvě synonymní varianty (*akcijowe towarstwo / akcijne towarstwo*), u kterých kvůli nedostatku dat nelze určit, která z nich je typičtější; první se v Hotko v2 vyskytuje 14krát, druhá 9krát. Nastává tedy dilema, zda tyto varianty ponechat jako kolokace, anebo z nich vytvořit dvě samostatná hesla, která mimoto budou na sebe vzájemně odkazovat. Zatím se u takovýchto případů kloníme k prvnímu řešení, neboť se necítíme kompetentní (navzdory tomu, že jeden z autorů je rodilým mluvčím – sic!) rozhodnout kodifikační otázku, kterou však nejsou ochotni rozsoudit ani původci lužickosrbských slovníků z klíčových jazykovědných pracovišť. Takových adjektiv s vícerymi variantami je velké množství, zejména u přejímaných odvozenin (např. *ambulancowy / ambulanční* ‚ambulanční‘, *atomowy / atomarny* ‚atomový‘, *baletny / baletowy* ‚baletní‘ aj.), ale částečnou synonymií se občas vykazují též indigenní slovní zásoba (např. *běrnjacy / běrnowy* ‚bramborový‘, *běžacy / běžaty* ‚tekoucí, běžící‘, *bobrowy / bobrjacy* ‚bobří‘ aj.).

božij 23 boží; ~ narod(k)/žlobik betlém / jesličky; -e wotkazanje, žohnowanje přijímání, požehmání; -a nóc štedrovečerní noc, půlnoční bohoslužba; -a služba bohoslužby; -e džěčo Ježíšek, Jezulátko; što dariš swojim swójbny k božemu džěsću co daruješ swým přibuzným k ježíšku; -e spěče namebevstoupení Páně; Bože čělo Boží tělo; zool -e slónčko sluněčko sedmitečné; -a rosa bot rosnatka okrouhlostá (Drosera); -a martra bot nár kopretina řimbaba • -a ručka jeho zaja ranila ho mrtvice • slónčko boží domčk dže sluněčko zapadá

**Obr. 1.** Zpracované heslo *boži* z Mudrovy pozůstalosti

bazja 9z 1 základna; wojerska, energetiska -a vojenská, energetická ~; -a a nadtwar ~ a nadstavba; wobnowjenje Domowiny wot -y sem obnowení Domowiny zdola • na -y być něšto sčinić byť s to něčo udělat 2 chem zásadja; -y a kisaliny -y a kyseliny

**Obr. 2.** Zpracované heslo *baza* z Mudrovy pozůstalosti

## Bibliografie

Brankačkec, K., Vavřín, M., Zasina, A. J.: *Korpus InterCorp – hornolužická srbština, verze 16 z 12. 10. 2023*. Ústav Českého národního korpusu FF UK, Praha 2023. Dostupný z WWW: <<http://www.korpus.cz>>.

Serbski Institut Budyšin: HOTKO: hornolužický textový korpus, verze 2 z 6. 3. 2021. Ústav Českého národního korpusu FF UK, Praha 2021. Dostupný z WWW: <<http://www.korpus.cz>>.

Škrabal, M., Brankatschk, K. (2021). Mudra's Upper Sorbian-Czech dictionary – what can be done about this lexicographic „posthumous child“? In *Electronic lexicography in the 21st century (eLex 2021): Post-editing lexicography*. Proceedings of the eLex 2021 conference. Brno: Lexical Computing CZ, 325–338.

**Staročeská účelová adjektiva a jejich významové definice ve slovníku<sup>3</sup>**

Irena Fuková (Ústav pro jazyk český AV ČR)

Příspěvek se zabývá kategorií účelových adjektiv ve staročeském období. Postihuje její specifickou vztahu k ostatním typům staročeských deverbativních adjektiv (srov. MČ 1, s. 321–350). Podává vnitřní charakteristiku skupiny staročeských účelových přídavných jmen, a to z různých hledisek a s důrazem na její neohraničenost, prolínání s ostatními adjektivními typy a její vnitřní diferencovanost (srov. Veselý – Veselý, 2019, s. 54–61). Specifikuje typy nositele vlastnosti, řídicího substantiva, stč. účelových adjektiv jak z hlediska typu substance, tak z hlediska sémantické role (srov. Štícha et al., 2018, s. 842–845). Eviduje formanty, kterými se staročeská účelová adjektiva odvozovala nejčastěji i zřídka. Dále zjišťuje nejdůležitější odlišnosti jednotlivých slovtvorných typů užívaných pro vyjádření účelu.

Materiálovým východiskem jsou jednak staročeské slovníky: Elektronický slovník staré češtiny (ESSČ), Staročeský slovník (StčS) a Malý staročeský slovník (MSS), jednak lexikální kartotéka Staročeského slovníku (ExcStčS) a Staročeská textová banka (StčTB). Ze slovníků byl získán soubor adjektivních lemmat s (potenciálním) účelovým významem, doklady na ně pak byly získány z kartotéky a StčTB.

Slovníky ESSČ a StčS mají jako celek pro slovtvorně orientovaný výzkum historické slovní zásoby důležitý význam, neboť usilují o úplný popis dochované slovní zásoby staročeského období (StčS 1968, s. 19, ESSČ). Slovtvorný výzkum na nich založený pak zpětně prověřuje slovníkové zpracování a může poskytnout podněty pro jeho zpřesnění. Příspěvek se proto zaměřuje také na problematiku slovníkových definic účelových adjektiv: Získané poznatky o kategorii účelových přídavných jmen umožňují systematický přístup při definování významu lexikálních jednotek tohoto typu v historickém slovníku.

---

<sup>3</sup> Autorka v příspěvku rozvíjí poznatky z vlastní disertační práce (Fuková, 2023).

## Bibliografie

ESSČ: *Elektronický slovník staré češtiny*. Praha: Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka. Online: <<https://vokabular.ujc.cas.cz/informace.aspx?t=ESSC&o=slovníky>>. Cit. 24. 6. 2024.

ExcStčS: *Kartotéka excerpce ze staročeských textů vzniklá pro potřeby práce na Staročeském slovníku*. Praha: Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka.

FUKOVÁ, Irena (2023): *Adjektiva tvořená od infinitivního/minulého slovesného kmene příponou -c(i) v češtině staré a v češtině doby střední*. Disertační práce. Praha: Filozofická fakulta Univerzity Karlovy.

MČ 1: DOKULIL, Miloš et al. (1986): *Mluvnice češtiny. 1, Fonetika, Fonologie, Morfonologie a morfejmika, Tvoření slov*. Praha: Academia.

MSS: BĚLIČ, Jaromír – KAMIŠ, Adolf – KUČERA, Karel (1978): *Malý staročeský slovník*. Praha: SPN.

StčS: *Staročeský slovník (1968–2008)*. Praha: Academia.

StčS 1968: *Staročeský slovník. Úvodní stati, soupis pramenů a zkratk* (1968). Praha: Academia.

StčTB: *Staročeská textová banka*. Praha: Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka. Online: <<http://vokabular.ujc.cas.cz/banka.aspx?idz=STB>>. Cit. 24. 6. 2024.

ŠTÍCHA, František et al. (2018): *Velká akademická gramatika spisovné češtiny. I*. Praha: Academia.

VESELÝ, Luboš – VESELÝ, Vojtěch (2019). Nekodifikujme, raději popisujme a vysvětlujme normu a úzus. Pěstujme jazykovědu, nepěstěme jazyk. *Korpus – gramatika – axiologie*, 20, s. 47–62.

## Praktické zkušenosti s převodem slovníků do formátu TEI Lex-0

Boris Lehečka (Moravská zemská knihovna v Brně)

Příspěvek představí základní principy formátu TEI Lex-0 (Tasovac, 2018), na jehož vývoji se podílí pracovní skupina „Lexical Resources“ z infrastruktury DARIAH. Tento formát vychází ze standardu TEI (TEI: P5 Guidelines, 2024), přičemž se snaží z širokého spektra definovaných prvků, atributů a jejich kombinací vybrat takový repertoár, který zajistí, že s nimi bude možné zachytit existující i budoucí lexikografická díla způsobem, který umožní jejich další využití v digitální podobě: publikaci, sdílení, extrakci ap. a zejména srovnávání jednotlivých děl mezi sebou. Např. místo tří prvků pro heslovou stať, které definuje TEI P5 ve verzi 4.7.0, tj. <entry>, <entryFree> a <superEntry>, využívá TEI Lex-0 pouze element <entry>, který lze do sebe vnořovat a rozlišovat pomocí atributu @type.

Stejný přístup se týká i oblasti metadat, tj. obsahu elementu <teiHeader>, jehož skladba by měla vhodným způsobem reprezentovat informace o elektronickém nosiči zachyceného lexikografického díla, ale i o jeho předchozích podobách (např. tištěných) a verzích nebo o osobách, které se podílely na jeho vzniku v různých rolích. Zároveň by zde měl být prostor pro zachycení specificky lexikografických dat o obsahu díla, jako je např. časové období popisované slovní zásoby, výchozí a cílový jazyk, repertoár údajů uváděných ve zpracování (výslovnost, etymologie, doklady, překlad ap.).

Vedle základních informací o formátu TEI Lex-0 se příspěvek zaměří také na jeho praktické uplatnění v případě dvou typů slovníků: dobového retrodigitalizovaného a historického, který vzniká pouze v elektronické podobě.

Slovník Dicionario da Lingua Portuguesa (de Morais Silva, 1789) vyšel ve třech vydáních v letech 1789, 1813 a 1823. Jedná se o výkladový slovník portugalštiny s doklady z literatury a úzu. Pro digitalizaci bylo zvoleno první vydání, které bylo naskenováno a text rozpoznán pomocí programu ABBYY FineReader 12 a následně exportován do formátu PDF. I když by pro další zpracování byl vhodnější formát DOCX, na rozdíl od PDF nedokázal export do tohoto formátu v některých případech zachovat dvousloupcovou sazbu původního díla (řádky ze dvou sloupců se spojovaly do jednoho). Pro převod z formátu PDF do formátu XML, který bude zachovávat i typografické vlastnosti textu, jsme využili programovou knihovnu iText Core (Apryse, 2024). Na takto vytvořené XML jsme následně pomocí programovacího jazyka XProc 3.0 (2022) aplikovali sadu transformací XSLT, na jejichž základě vznikl dokument, který se blížil standardu TEI Lex-0, přičemž další úpravy bylo nutné provádět ručně. Druhým příkladem je Elektronický slovník staré češtiny (2006–), který vzniká v oddělení vývoje jazyka Ústavu pro jazyk český a je publikován na webových stránkách Vokabuláře webového (2006–). Jedná se o historický výkladový jednojazyčný slovník, zachycující slovní zásobu češtiny od jejích počátků do konce 15. století. Vzniká pouze v elektronické podobě (v textovém procesoru Microsoft Word). Plná heslová stať se skládá z heslového záhlaví, obsahujícího základní podobu heslového slova a jeho tvaroslovnou charakteristiku, z údaje o slovtvorné motivaci, z popisu významu, dokladové části a fakultativně ze závěrečné poznámky. V tomto případě byl k dispozici slovník ve specifickém formátu XML, který se využíval pro publikování ve webové aplikaci. Rovněž zde se využila postupná transformace do standardu TEI Lex-0, přičemž ruční úpravy výsledného textu nebyly nutné, neboť již samotné zpracování v dokumentu Word obsahovalo vyznačení jednotlivých částí heslových statí pomocí odstavcových a znakových stylů. Obohacení původních dat o nové údaje (např. identifikace místa v citovaném prameni, tj. folia, strany či verše) jsme docílili využitím regulárních výrazů a technologie Invisible XML (bez data).

Rozdílné lexikografické zpracování slovníků převáděných do formátu TEI Lex-0 i rozdílná podoba výchozích podkladů (PDF a DOCX, resp. XML) vedly k tomu, že ke každému slovníku bylo nutné přistupovat individuálně, a to pomocí řetězení postupných úprav a transformací. Po zkušenostech s převodem dalších slovníků je však možné konstatovat, že existuje určitý repertoár společných kroků,



které lze pomocí parametrů, např. v podobě repertoáru používaných zkratek, přizpůsobit konkrétnímu slovníku a umožnit tím snazší přechod k formátu TEI Lex-0 pro více děl.

## Bibliografie

APRYSE, 2024. IText Core. APRYSE. *The Leading PDF Library for Developers | iText* [online]. [cit. 2024-03-23]. Dostupné z: <https://itextpdf.com/products/itext-core>

de MORAIS SILVA, António, 1789. *Diccionario da Lingua Portuguesa*. Lisboa: Officina de Simão Thaddeo Ferreira.

*Elektronický slovník staré češtiny* [online], 2006–. Praha: Ústav pro jazyk český AV ČR, v. v. i., oddělení vývoje jazyka [cit. 2024-03-23]. Dostupné z: <http://vokabular.ujc.cas.cz>

*Invisible XML* [online], bez data. [cit. 2024-03-23]. Dostupné z: <https://invisiblexml.org>

TASOVAC, Toma, Laurent ROMARY, Piotr BANSKI, et al. 2018. *TEI Lex-0: A baseline encoding for lexicographic data* [online]. DARIAH WORKING GROUP ON LEXICAL RESOURCES. Version 0.9.3. [cit. 2024-03-23]. Dostupné z: <https://bit.ly/tei-lex-0>

TEI: P5 Guidelines, 2024. *The Text Encoding Initiative Consortium* [online]. [cit. 2024-03-23]. Dostupné z: <http://www.tei-c.org/Guidelines/P5/>

*Vokabulář webový*. Online. 2006–. Dostupné z: <http://vokabular.ujc.cas.cz>. [cit. 2015-03-21].

*XProc 3.0* [online], 2022. [cit. 2024-03-23]. Dostupné z: <https://xproc.org>

**Retrodigitalizace Bulharsko-českého slovníku Karla Hory a kolektivu SLÚ ČSAV: teoretické a metodologické aspekty přípravy podkladů pro bulharsko-českou slovníkovou databázi**

Božana Niševa<sup>1</sup>, Štefan Pilát<sup>1</sup>, Michaela Kuzmova<sup>2</sup>

(<sup>1</sup>Slovanský ústav AV ČR, <sup>2</sup>Ústav bulharského jazyka prof. L. Andrejčina, BAV)

V příspěvku se zaměřujeme na teoreticko-metodologické přístupy a postupy uplatňované při retrodigitalizaci Bulharsko-českého slovníku K. Hory a kolektivu Slovanského ústavu ČSAV. Tento základní úkol jsme si vytyčili v rámci Mobilitního projektu Plus s názvem *The Use of Modern Technologies in Opening Lexicographical Heritage to the Public – The Theoretical and Methodological Aspects of Preparing the Source Materials for a Bulgarian-Czech Dictionary Database (2023 – 2024)*, který probíhá v letech 2023–2024 a účastní se ho pracovníci Slovanského ústavu AV ČR, v. v. i., a Ústavu bulharského jazyka prof. L. Andrejčina BAV. V souladu se současným pojetím problematiky (viz např. Kallas et al. 2019) retrodigitalizaci chápeme jako proces převodu staršího/tištěného slovníku do digitální, počítačem čitelné a zpracovatelné podoby. Vzhledem k retrodigitalizaci Bulharsko-českého slovníku vydělujeme konkrétně tyto tři základní fáze: 1) skenování a OCR; 2) vytvoření schématu hierarchické datové struktury odrážející vnitřní strukturu hesel (s využitím obecného značkovacího jazyka XML) a transformace textu slovníku do této struktury; 3) vytvoření prostředí pro on-line prezentaci slovníkových dat za použití kaskádových stylů (CSS), vytvoření vyhledávacích a filtrovacích nástrojů. K specifickým lexikografickým metodám uplatněným při popisu na 60 000 heslových slov (od archaismů po lexémy z různých odborných oblastí) v něm obsažených přistupujeme rovněž z následujících pozic. Při retrodigitalizaci je vedle mikrostruktury lexikografického díla, která se vyznačuje lineárností, nehierarchičností (Dziemianko 2012), transformován též účelný a hierarchicky odstupňovaný repertoár interpunkčních a grafických prostředků. Požadavek největší možné přesnosti převodu tištěného Bulharsko-českého slovníku do digitální podoby (např. pro potřeby lingvistické analýzy, překladatelské práce apod.) se tedy vztahuje k citlivému zacházení jednak s bulharským a českým jazykovým materiálem popsaným na základě jednotné (tradiční) lexikografické koncepce a jednak s interpunkčními a grafickými prostředky, které se dostávají do nového (digitálního) prostředí, v němž se jejich primární funkce, resp. význam, podstatně liší (Bouda & Cysouw 2012). Se zřetelem k tomu poukazujeme na úskalí retrodigitalizace dvojjazyčného typu slovníku, v jehož levé části je obsažen jazykový materiál psaný cyrilicí a v pravé latinkou. Snažíme se mj. i o typologizaci chyb, se kterými je třeba v obdobných případech počítat, a to už v prvotní fázi vyhotovení OCR dokumentu, resp. při jeho redigování. Na neposledním místě při přípravě podkladů pro budoucí online zpřístupnění Bulharsko-českého slovníku K. Hory a kolektivu SLÚ ČSAV širší veřejnosti vycházíme ze skutečnosti, že se jedná o archivaci a elektronizaci nedílné – a na poli bulharsko-české lexikografie dosud nepřekonané – součásti českého lexikografického dědictví.

## Bibliografie

- [1] Atkins, B. T. S. (1996). Bilingual dictionaries: Past, present and future. In M. Gellerstam, J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström, & K. Røjder Papmehl (Eds.), *Euralex '96 proceedings* (pp. 515-546). Göteborg: University, Department of Swedish.
- [2] Atkins, B. T. S., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- [3] Bouda, P., & Cysouw, M. (2012). Treating Dictionaries as a Linked-Data Corpus. In C. Chiarcos, S. Nordhoff, & S. Hellmann (Eds.), *Linked Data in Linguistics. Representing and Connecting Language Data and Language Metadata* (pp. 15-23). Berlin, Heidelberg: Springer.
- [4] Čermák, F. (1995). Překladová lexikografie. In F. Čermák, & R. Blatná (Eds.), *Manuál lexikografie* (s. 230-248). Praha: Nakladatelství H & H.
- [5] Dziemianko, A. (2012). On the use(fulness) of paper and electronic dictionaries. In S. Granger, & M. Paquot (Eds.), *Electronic Lexicography* (pp. 319-342). Oxford: Oxford University Press.
- [6] Kallas, J., Koeva, S., Langemets, M., Tiberius, C., & Kosem, I. (2019). Lexicographic Practices in Europe: Results of the ELEXIS Survey on User Needs. In *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference. 1-3 October 2019, Sintra, Portugal* (pp. 519-536).
- [7] Knoll, V. (2017). Trendy v elektronické diachronní lexikografii a jejich využitelnost ve slavistice. In *Slavia*, 86(2-3), s. 301-334.
- [8] Müller-Spitzer, C. (2014). 11. Textual structures in electronic dictionaries compared with printed dictionaries: A short general survey. In: Gouws, R., Heid, U., Schweickard, W., & Wiegand, H. (Eds.), *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. (pp. 367-381). Berlin, Boston: De Gruyter Mouton.
- [9] Rundell, M. (2015). From print to digital: Implications for dictionary policy and lexicographic conventions. *Lexikos*, 25(1), pp. 301-322.
- [10] Sekaninová, E. (1993). *Dvojjazyčná lexikografia v teorii a praxi*. Bratislava: Veda.
- Svensén, B. (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- [12] Tarp, S. (2012). Theoretical challenges in the transition from lexicographical p-works to e-tools. In S. Granger, & M. Paquot (Eds.), *Electronic Lexicography* (pp. 107-118). Oxford: Oxford University Press.

### Exemplifikace ve *Slovníku nářečí českého jazyka*: zdroje, postupy, problémy

Marta Šimečková (Ústav pro jazyk český AV ČR)

*Slovník nářečí českého jazyka* je prvním slovníkem, který zpracovává slovní zásobu všech nářečí češtiny, a to ve vztahu ke spisovnému jazyku nediferenčním způsobem. Vzniká od r. 2011 v dialektologickém oddělení Ústavu pro jazyk český AV ČR, v. v. i. Jde o slovník elektronický, který je od r. 2016 zpřístupňován na webu [snj.ujc.cas.cz](http://snj.ujc.cas.cz). Dosud byla zveřejněna hesla s náslovím *A-* až *D-* a také řada hesel mimo abecední pořadí.

Nedílnou součástí slovníkových odstavců jsou exemplifikace, které vzhledem k již uvedenému tvoří mnohdy rozsáhlé statě (viz např. hesla *bába*; *bůh*; *chléb*, *chleba*) – je to dáno jednak skutečností, že je slovník tezaurem českého nářečního lexika, jednak že autoři slovníku nejsou omezeni rozsahem, neboť jde o slovník v online prostoru. Zatímco o samotném slovníku už bylo napsáno mnohé (např. Kloferová 2013, 2016a, 2016b, 2017; Šipková 2012, 2016; Ireinová 2016, 2019, 2021), problematika exemplifikací zůstává spíše upozaděna, což víceméně platí o metalexikografické literatuře obecně (s některými výjimkami, v českém prostředí např. Lišková–Procházková 2016; Lišková 2017a, 2017b; Tichá 2012).

Práce s exemplifikacemi je přitom v daném slovníku v řadě ohledů specifická – exemplifikace mají oproti výkladovým slovníkům spisovné češtiny nejen objasnit užívání slov pomocí uzuálních kontextů (uzuálnost však může být někdy sporná, jak bude ukázáno na konkrétních příkladech), ale také lokalizovat slova a jejich významy do určitých nářečních či regionálních oblastí (čimž slovník částečně zastupuje funkci atlasu), též dokládat podrobně hláskovou či morfologickou variantnost slova ap. Při výběru exemplifikací jsou tak zohledňována různá kritéria. Mnohé obtíže jsou spojeny i se samotnými zdroji, z nichž jsou exemplifikace čerpány, a také s úpravami dokladů (nutný převod dokladu do zástupné, slovníkové podoby; rekonstrukce předpokládané nářeční formy ze zápisu, jenž nebyl pořízen dialektologickou transkripcí, vedoucí však někdy k nechtěné idealizaci podávaného obrazu nářečí). I to bude předmětem příspěvku, který bude v některých místech představovat kritické ohlédnutí za odvedenou prací ve snaze o zkvalitnění prací budoucích.

## Bibliografie

- IREINOVÁ, M. Materiálová základna a zdroje Slovníku nářečí českého jazyka. In: REMBISZEWSKA, D. K., ed. *Słowiańskie słowniki gwarowe – tradycja i nowatorstwo*. Warszawa– Łomża: Instytut Slawistyki PAN, Łomżyńskie Towarzystwo Naukowe im. Wagów, 2016, s. 75–82. ISBN 978-83-86175-98-7.
- IREINOVÁ, M. O Slovníku nářečí českého jazyka. In: CSÁSZÁRI, É. a M. IMRICHOVÁ, eds. *Király Péter 100. Tanulmánykötet Király Péter tiszteletére I*. Budapest: ELTE BTK, Szlav Filológiai Tanszék, 2019, s. 223–229. ISBN 978-963-489-066-9.
- IREINOVÁ, M. Úkoly české dialektologie na počátku 21. století. *Gwary Dziś*. 2021, 14, s. 11–16. ISSN 1898-9276.
- KLOFEROVÁ, S. Současná česká dialektologie: po Českém jazykovém atlasu Slovník nářečí českého jazyka. *Slavia*. 2013, roč. 82, s. 169–174. ISSN 0037-6736.
- KLOFEROVÁ, S. Slovník nářečí českého jazyka – úskalí lexikografické a dialektologické abstrakce. In: LIŠKOVÁ, M., V. VODRÁŽKOVÁ a Z. DĚNGEOVÁ, eds. *Akademický slovník současné češtiny a software pro jeho tvorbu aneb Slovníky a jejich uživatelé v 21. století. Sborník abstraktů z workshopu konaného ve dnech 29.–30. 11. 2016 v Praze*. Praha: Ústav pro jazyk český AV ČR, v. v. i., 2016a, s. 38–39. ISBN 978-80-88211-02-0.
- KLOFEROVÁ, S. Ke koncepci celonárodního nářečního slovníku. In: MYZNIKOV, S. A., O. N. KRYLOVA a I. V. BAKLANOVA, eds. *Slavjanskaja dialektnaja leksikografija 2. Matěrialy konferencii*. Sankt Petěrburg: Nestor-Istorija, 2016b, s. 100–116. ISBN 978-5-4469-1064-9.
- KLOFEROVÁ, S. Ein neues dialektologisches Projekt des Tschechischen: Das Wörterbuch der tschechischen Dialekte in Böhmen, Mähren und Schlesien. In: WANDL-VOGT, E., S. WINTERSTEIN, CH. SCHRÖDL a A. SINGER, eds. *Dialekt / dialect 2.0. Langfassungen / Long papers. 7. Kongress der Internationalen Gesellschaft für Dialektologie und Geolinguistik (SIDG) / 7. Congress of the International Society for Dialectology and Geolinguistics (SIDG)*. Wien: Praesens Verlag, 2017, s. 213–220. ISBN 978-3-7069-0962-4.
- LIŠKOVÁ, M. – PROCHÁZKOVÁ, B. Exemplifikace. In: KOCHOVÁ, P. a Z. OPAVSKÁ, eds. *Kapitoly z koncepce Akademického slovníku současné češtiny*. Praha: Ústav pro jazyk český AV ČR, v. v. i., 2016, s. 154–160. ISBN 978-80-86496-89-4.
- LIŠKOVÁ, M. Výměna stárnoucí manželky za hezkou mladou kočku aneb K exemplifikaci v Akademickém slovníku současné češtiny. *Naše řeč*. 2017a, 100, č. 1, s. 28–36. ISSN 0027-8203.
- LIŠKOVÁ, M. Slovníková exemplifikace jako zdroj komiky. Nad příklady elektronické podoby Slovníku současné češtiny (nechybujte.cz, verze 2.0, 2016). *Jazykovědné aktuality*. 2017b, 54, č. 3–4, s. 47–50. ISSN 1212-5326.
- TICHÁ, Z. Příklady ve výkladovém slovníku. In: K. BUZÁSSYOVÁ, B. CHOCHOLOVÁ a N. JANOČKOVÁ, eds. *Slovo v slovníku. Aspekty lexikální sémantiky – gramatika – stylistika (pragmatika). Na počest Alexandry Jarošovej*. Bratislava: Veda, 2012, s. 282–293.
- ŠIPKOVÁ, M. Možné typy nářečního slovníku. In: ČMEJRKOVÁ, S., J. HOFFMANNOVÁ a J. KLÍMOVÁ, eds. *Čeština v pohledu synchronním a diachronním. Stoleté kořeny ústavu pro jazyk český*. Praha: Karolinum, 2012, s. 327–332. ISBN 978-80-246-2121-0.
- ŠIPKOVÁ, M. Slovník nářečí českého jazyka: Charakteristika, vymezení, materiálová základna, výběr heslových slov. In: LIŠKOVÁ, M., V. VODRÁŽKOVÁ a Z. DĚNGEOVÁ, eds. *Akademický slovník současné češtiny a software pro jeho tvorbu aneb Slovníky a jejich uživatelé v 21. století. Sborník abstraktů z workshopu konaného ve dnech 29.–30. 11. 2016 v Praze*. Praha: Ústav pro jazyk AV ČR, v. v. i., 2016, s. 68–69. ISBN 978-80-88211-02-0.

### Komunikační funkce ve slovníku? Tak určitě! (Ke zpracování jistotních částic v češtině)

Barbora Štěpánková, Lucie Poláková, Jana Šindlerová (Charles University)

Základem výkladového slovníku jsou autosémantická (pojmová) slova s výkladem opřeným o jejich lexikální význam (srov. i *lexikálně gramatická slova*, Filípec–Čermák 1985: 77). Okrajovější postavení mají ve slovnících slova, u nichž nad lexikálním významem převažuje gramatická funkce, tzv. *syntaktika* (u Filípce *gramaticko-sémantická slova*), tj. předložky, spojky a částice, v některých pojetích i pomocná slovesa či zájmena. Výklad významu těchto výrazů bývá opřen právě o jejich gramatickou funkci. Vzhledem k tomu, že částice představují „slova s primárně pragmatickou povahou a územ“ (Čermák 1992: 257), domníváme se, že vyžadují specifické lexikografické řešení, opřené nikoli o gramatickou, ale právě o pragmatickou složku významu.

V novějších teoretických přístupech a následně i ve výkladových slovnících lze v rámci výkladu významu pozorovat posun k pragmatice (srov. Dolník 2012, Kochová–Opavská 2016, Nagy 2013), který bývá spojován s tzv. pragmatickým obratem a posunem zaměření lingvistiky také na komunikaci. Tento posun je soustředěn především na oblast autosémantik (srov. Geeraerts 2010), ale setkáváme se s ním i u jiných slovních druhů. U částic<sup>4</sup> se obecně projevuje zájem o jejich úlohu v diskurzu, ve slovníku např. snahou zaznamenat postoj mluvčího či komunikační funkci, kterou indikují (srov. Kochová–Opavská: 140). Co se týče jistotních částic, nacházíme však v českých slovnících<sup>5</sup> jen ojedinělé pokusy (srov. *asi* v ASSČ, *zajisté* v SSJČ). Tento stav si vysvětlujeme několika skutečnostmi spojenými se stávající lexikografickou praxí: 1. obecný problém velkých lexikografických děl, kdy abecední postup a velké množství zpracovávaných hesel neumožňují jednotnou analýzu celé sémantické skupiny výrazů; 2. částice jako výrazy často vzniklé pragmatikalizací<sup>6</sup> se i dnes dále vyvíjejí a pragmatikalizují, tradiční výklad jistotních částic založený na popisu míry jistoty (což je rys blízký pojmovosti autosémantik, a tudíž relativně snadno uchopitelný) je ale jen opatrně doplňován jinými informacemi, pro které nejsou k dispozici konvencionalizované lexikografické popisy.

V tomto příspěvku představujeme návrh slovníkového zpracování jistotních částic, v němž komunikační funkce tvoří opěrný bod popisu, založený na detailní anotaci korpusových dat.<sup>7</sup> Jistotní částice pojmáme v širokém smyslu, v souladu např. s Mluvnicí češtiny 2 (s. 233), tj. kromě epistemických částic (jednoslovných i víceslovných – *jistě*, *s jistotou*), mezi ně zařazujeme i výrazy s evidenciálními rysy (typické *prý*, *údajně*, ale i přechodové *zřejmě* a víceslovné *podle všeho*<sup>8</sup>) a výrazy „dotvrzovací“ (*samozřejmě*, *opravdu*). Anotace zachycuje míru jistoty, relevantní prvky v okolí výrazu, ale především komunikační funkci v kontextu (srov. Štěpánková et al. 2023). Na základě anotovaných dat se popis významu pomocí komunikační funkce jeví jako výhodný zejména vzhledem k tomu, že míra jistoty výrazu se vlivem kontextu proměňuje a v některých případech se stává spíše obecně jistotním/modálním příznakem potvrzujícím komunikační funkci dané věty (např. ...*vy to, pane Banako, jistě znáte.*; *Asi bych si od něho měla něco přečíst.*). Při anotaci se zaměřujeme i na jiná (nejistotní) užívání těchto výrazů, tj. na jejich původní plnovýznamové použití či dále pragmatikalizovaná použití (v přitakání, ve zdvořilostních konstrukcích apod.). Získáváme tak komplexní přehled o chování jednotlivých výrazů, využitelný nejen pro tvorbu slovníků, ale i pro představu o procesu pragmatikalizace v češtině vůbec.

<sup>4</sup> Detailní analýzu pojetí *částic* / *particles* ve slovanských jazycích a jejich ekvivalentů v anglofonní lingvistice viz Rozumko 2016.

<sup>5</sup> Na rozdíl od některých zahraničních slovníků, srov. např. OLD – *perhaps*, nebo DWDS – *vielleicht*.

<sup>6</sup> Pragmatikalizaci chápeme jako výrazné posílení pragmatické složky významu výrazu na úkor části pojmové, srov. např. Aijmer 1997.

<sup>7</sup> Jako zdroj dat používáme česko-anglickou verzi korpusu InterCorp; překladové ekvivalenty využíváme jako oporu při identifikaci významů částic (srov. např. Dyvik 2004, Aijmer et al. 2006).

<sup>8</sup> V rámci diskuse o vztahu mezi epistemičností a evidencialitou (přehledově např. Ivanová 2017) se na základě testovacích analýz (srov. Šindlerová et al. 2023) kloníme k prostupnosti obou kategorií.

## Bibliografie

- Aijmer, K. (1997). "I think – an English modal particle". In: T. Swan – W. O. Jansen (eds.), *Modality in Germanic languages (Trends in Linguistics, Studies and Monographs 99)*. Berlin/New York: Mouton de Gruyter, 1–47.
- Aijmer, K., Foolen, A., Simon-Vandenberg, A. M. (2006): Pragmatic markers in translation: a methodological proposal. In: K. Fischer (ed.), *Approaches to discourse particles*. Oxford: Elsevier, 101–114.
- Akademický slovník současné češtiny* [online] (2017–2023). Praha: ÚJČ AV ČR (ASSČ)
- Čermák, F. (1992). Paradigmatika a syntagmatika slovníku: problémy a možnosti. *Slovo a slovesnost* 53, 249–264.
- Dolník, J. (2012). Lexikální pragmatika. In: K. Buzássyová – B. Chocholová – N. Janočková (eds.), *Slovo v slovníku: aspekty lexikální sémantiky – gramatika – stylistika (pragmatika): na počest Alexandry Jarošovej*. Bratislava: Veda, 41–49.
- Dyvik, H. (2004): *Translations as semantic mirrors: from parallel corpus to wordnet*. In *Advances in corpus linguistics*. Brill, 309–326.
- „vielleicht“, in: *Digitales Wörterbuch der deutschen Sprache*, hrsg. v. d. Berlin-Brandenburgischen Akademie der Wissenschaften, <https://www.dwds.de/wb/vielleicht> (DWDS)
- Filipec, J. & Čermák, F. (1985). *Česká lexikologie*, Praha: Academia.
- Geeraerts, D. (2010). *Theories of Lexical Semantics*. Oxford: Oxford University Press.
- Havránek, B. et al. (1960–1971). *Slovník spisovného jazyka českého*. Praha: ČSAV. (SSJČ)
- Ivanová, M. (2017). *Modálnost' a modálne verbá v slovenčine*. Prešov: FFPU.
- Klégr, A. et al. (2022). *Korpus InterCorp – angličtina*, verze 14 z 31. 1. 2022. Praha: Ústav Českého národního korpusu FF UK. <http://www.korpus.cz>
- Kochová, P. & Opavská, Z. (eds.) (2016). *Kapitoly z koncepce Akademického slovníku současné češtiny*. Praha: Ústav pro jazyk český AV ČR.
- Mluvnice češtiny 2. Tvarosloví*. (1986). J. Petr (ed.). Praha: Academia.
- Nagy, M. (2013). Lexikální význam, nebo encyklopedické heslo: metodologické úvahy. In: P. Adámková – M. Martinková – O. Uličný (eds.), *Studie k moderní mluvnici češtiny 4. Dynamika českého lexika a lexikologie*. Olomouc: Filozofická fakulta Univerzity Palackého v Olomouci, 24–51.
- „perhaps“, in: *Oxford Learner's Dictionary* <https://www.oxfordlearnersdictionaries.com/definition/english/perhaps?q=perhaps> (OLD)
- Rosen, A. et al. (2022). *Korpus InterCorp – čeština*, verze 14 z 31. 1. 2022. Praha: Ústav Českého národního korpusu FF UK. <http://www.korpus.cz>
- Rozumko, A. (2016). Linguistic Concepts across Languages: The Category of Epistemic Adverbs in English and Polish. *Yearbook of the Poznan Linguistic Meeting 2* (1), 195–214.
- Šindlerová, J., Štěpánková B. & Andrén, I. L. (2023). Epistemická částice *zřejmě* pohledem paralelního korpusu. *Korpus – gramatika – axiologie*, 27, 37–52.
- Štěpánková, B., Šindlerová, J. Poláková, J. (2023). The Epistemic Marker *určitě* in the Light of Corpus Data. *Jazykovedný časopis*, 74, 130–139.

## CzeDLex – Slovník českých diskurzivních konektorů

Pavlaína Synková<sup>1,2</sup>, Jiří Mírovský<sup>1</sup>, Lucie Poláková<sup>1</sup> (<sup>1</sup>ÚFAL MFF UK, <sup>2</sup>ÚČJTK FF UK)

CzeDLex (Mírovský et al., 2021a) je elektronický, online dostupný slovník českých diskurzivních konektorů vybudovaný na základě anotace diskurzivních vztahů v Pražském závislostním korpusu (dále PDT, Hajič et al., 2006, Rysová et al., 2016) a později obohacený z dalších zdrojů (Mírovský et al., 2017, 2021b). PDT s anotací diskurzivních vztahů zahrnuje asi 50 tis. vět a asi 21 tis. diskurzivních vztahů. Diskurzivními vztahy v tomto korpusu rozumíme významové vztahy podílející se na koherenci textu, a to takové, které se vyskytují mezi samostatnými výpověďmi nebo většími úseky textu (Poláková et al., 2012). Takovým vztahem je např. konfrontace – vyskytuje se mezi samostatnými výpověďmi (*Na výzvu už reagovala Indie. Naopak Čína ještě neodpověděla.*), ale zároveň také v rámci podřadných (*Zatímco Indie už na výzvu reagovala, Čína ještě neodpověděla.*) i souřadných souvětí (*Na výzvu už reagovala Indie, naopak Čína ještě neodpověděla.*). Konektory těchto vztahů jsou pak prostředky, které otevírají pozice pro dvě takové jednotky textu (v PDT byly tyto textové jednotky omezeny na ty, které obsahují verbum finitum) a vyjadřují významový vztah mezi nimi. Ze slovnědruhového hlediska se jedná o spojky (např. *nebo, zatímco*), příslovce (např. *proto, naopak, potom*), částice (např. *vždyť, přece*), zájmena (např. *což, tím*) a dále o víceslovné výrazy (např. *kromě toho; z toho důvodu; s tím, že*) a také některá interpunkční znaménka (: ; -). Ustálené spojovací prostředky nazýváme primárními konektory, pro neúplně ustálené spojovací prostředky (např. *z toho důvodu; za podmínky, že; to znamená, že*) byl zvolen termín sekundární konektory (Rysová a Rysová, 2018). Sekundární konektory jsou součástí slovníku v omezené míře, primární konektory, které se vyskytly ve zkoumaném materiálu, byly zahrnuty všechny.

Při hledání struktury hesla byly inspirací slovníky konektorů pro jiné jazyky – nejpropracovanější se ukázaly slovníky pro němčinu (DiMLex, Stede, 2002) a francouzštinu (LexConn, Roze et al., 2012). Bylo nutné vypořádat se zejména s tím, že: 1) nalezených spojovacích výrazů byly stovky a bylo nutné rozhodnout, která slova a řetězce slov sloučit pod jednotlivá hesla, 2) konektory se jen málokdy užívají k vyjádření pouze jediného diskurzivního významu (viz funkční homonymie u *když, ale* apod.), 3) konektory se liší z hlediska slovnědruhového, slovosledného i z hlediska toho, v jaké části vztahu se používají (významově i lineárně), 4) naprostá většina spojovacích výrazů je homonymní s nekonektivními užitími daných slovních forem (např. výraz *ale* jako spojka a jako postojová částice). Heslo ve slovníku má tuto podobu: na první úrovni je tvořeno lemmatem (*protože, ale, jak – tak, jinými slovy*) a charakterizováno jako primární, nebo sekundární a jako jednoslovné, nebo víceslovné. V druhé úrovni je heslo členěno dle jednotlivých diskurzivních významů (např. *když* – současnost, *když* – podmínka) a pro každý z nich jsou uvedeny další charakteristiky: slovnědruhá příslušnost; všechny podoby výskytu konektoru v datech (např. *ale* ve významu opozice má v datech 23 různých realizací – *sice ale, ale přesto, zároveň ale* ...) a jeho modifikace (takové varianty konektoru, které ho specifikují co do jistoty nebo intenzity, např. *hlavně proto*); údaj, zda je vyjadřovaný diskurzivní typ symetrický a v případě, že není, kterou část vztahu konektor vyjadřuje (např. zda se vyskytuje v části, kde se vyjadřuje příčina/důvod, nebo v části, kde se vyjadřuje důsledek); slovosledná pozice v rámci klauze; údaj, zda se konektor vyskytuje v lineárně prvním, nebo druhém argumentu (nebo v obou). Jako doplněk jsou uvedena i nekonektivní užití hesla první úrovně. Korpusové příklady (konektivních i nekonektivních užití) obsažené ve slovníku jsou ručně přeloženy do angličtiny a také všechny charakteristiky konektorů jsou uvedeny tak, aby byly srozumitelné mezinárodnímu publiku. Slovník má strukturu využitelnou v NLP aplikacích a tvoří významný podklad pro český diskurzivní parser (Mírovský et al., 2024).

Současná verze slovníku obsahuje 200 jednotek první úrovně, zahrnuje celá data PDT a dále konektory z české části korpusu Prague Czech-English Dependency Treebank (Hajič et al., 2011) a několika dalších menších zdrojů (<https://ufal.mff.cuni.cz/czedlex>).

Cílem příspěvku je představit slovník českému publiku a diskutovat o tom, zda použité charakteristiky popisují konektory dostatečně nejen pro účely NLP, ale i z hlediska lingvistické analýzy.

Příspěvek byl vytvořen s podporou projektů GAČR GA22-03269S a projektu MŠMT LM2023062 LINDAT/CLARIAH-CZ.



## Bibliografie

Hajič, J. et al. (2006): *Prague Dependency Treebank 2.0*. Data/software, Institute of Formal and Applied Linguistics.

Hajič, J. et al. (2011): *Prague Czech-English Dependency Treebank 2.0*. Data/software, Institute of Formal and Applied Linguistics.

Mírovský, J., Synková, P., Poláková, L., Rysová, M. (2017): CzeDLex – A Lexicon of Czech Discourse Connectives. *The Prague Bulletin of Mathematical Linguistics*, 109, pp. 61–91.

Mírovský, J., Synková, P., Poláková, L., Kloudová, V., Rysová, M. (2021a): *CzeDLex 1.0*. Data/software, Charles University, Prague, Czech Republic, <http://hdl.handle.net/11234/1-4595>.

Mírovský, J., Synková, P., Poláková, L. (2021b): Extending Coverage of a Lexicon of Discourse Connectives Using Annotation Projection. *The Prague Bulletin of Mathematical Linguistics*, 117, pp. 5–26.

Mírovský, J., Synková, P., Poláková, L., Paclíková, M. (2024): Cost-Effective Discourse Annotation in the Prague Czech–English Dependency Treebank. In *Proceedings of LREC-COLING 2024*, European Language Resources Association, Torino, Italy.

Poláková, L. et al. (2012): Manual for annotation of discourse relations in the Prague Dependency Treebank. *Technická zpráva*. Praha: Univerzita Karlova v Praze, Matematicko-fyzikální fakulta. Dostupné online na adrese: <<http://ufal.mff.cuni.cz/techrep/tr47.pdf>>.

Rysová, M. et al. (2016): *Prague Discourse Treebank 2.0*. Data/software, Institute of Formal and Applied Linguistics.

Rysová, M., Rysová, K. (2018): Primary and secondary discourse connectives: Constraints and preferences. *Journal of Pragmatics*, 130, pp. 16–32.

Stede, M. (2002): DiMLex: A Lexical Approach to Discourse Markers In: *A. Lenci, V. Di Tomaso(eds.): Exploring the Lexicon - Theory and Computation*. Alessandria (Italy): Edizioni dell'Orso.

Roze, Ch., Danlos, L., Muller, P. (2012): LEXCONN: A French Lexicon of Discourse Connectives. *Discours [En ligne]*, 10, <http://discours.revues.org/8645>.

### Specifika a perspektivy nářeční lexikografie při tvorbě lexikálních databází

Michal Vašíček (Slovanský ústav AV ČR), Darja Vašíčková (Filozofická fakulta UK v Praze, ÚVES)

V příspěvku se zaměříme na některé specifické otázky nářeční lexikografie vyvstávající při zpracování archivních slovníkových materiálů. Vycházet přitom budeme ze zkušeností se zpracováním slovníkové kartotéky Ivana Paňkevyče, významného slavisty, který od 20. let 20. stol. sbíral lexikum rusínských, resp. ukrajinských nářečí na území tehdejšího Československa (Dolnycká 1997; Dolnycká–Šišková 1992; Dzendzelivs'kyj 1969). Tato kartotéka byla naskenována a na jejím základě byla vytvořena lexikální databáze jihokarpatských nářečí (Skwarska et al. 2020), v současné době je databáze veřejně přístupná (Šišková et al. 2023; Vašíček 2023). Jedná se zatím o digitalizovanou slovníkovou kartotéku, jež by se měla stát základem pro tvorbu elektronického slovníku (teoreticky vícejazyčného), může však sloužit jako zdroj materiálu pro řadu dalších digitálních nástrojů. Z kontextů heslových slov lze např. vytvořit diachronní nářeční korpus překračující čistě lexikografické zaměření původního materiálu. Hlavní potenciál současných databázových technologií však spatřujeme v tom, že v nich lze shromažďovat, třídit a srovnávat materiály z různých starších zdrojů, zejména nářečních slovníků a atlasů. Za velmi perspektivní pro další rozvoj dialektologických bádání považujeme propojení lexikografických a lingvogeografických prací, tedy nářečních slovníků a jazykových map. Jako velmi žádoucí vnímáme i tvorbu nových map na základě údajů v databázi a existujících nářečních slovníků. Dialektologická bádání jsou vždy ukotvena v konkrétním prostoru, nejmenší prostorovou jednotkou je lokální nářečí, které představuje kompletní a do jisté míry uzavřený jazykový systém, který lze srovnávat s okolními systémy. Lokalizace nářečního lexému má tak ve slovníku klíčový význam. Současné dialektologické výzkumy se navíc stále více sblížují s areálovou a kontakto­vou lingvistikou (k tomu např. Wiemer–Seržant 2014), nabývají tak na významu výzkumy zaměřené na původ lexika, což je zároveň impulsem pro rozvoj nářeční etymologie (např. Králik 2020; Varbot 2021). Specifika karpatských nářečí byla utvářena v první řadě kontaktem se sousedními jazyky, proto je pro studium jazyků karpatského areálu toto propojení různých odvětví lingvistiky zvláště aktuální. Právě přesné geografické vymezení výskytu zkoumaného lexému (se zahrnutím jeho různých významů a jejich areálové distribuce) často hraje zásadní roli pro správné určení jeho etymologie. Nesnadným úkolem nářeční lexikografie je správné zachycení mnohoznačnosti slov, zvláště odlišení polysémie a homonymie. Obzvláště důležité je proto vytváření sémantických map (podle principu „od slova k významu“, OKDA), což potvrdilo dosavadní zpracování některých dílčích skupin lexika v databázi, např. ornitonym (Vašíčková 2022). Část materiálů databáze bychom proto chtěli experimentálně zpracovat následujícím způsobem: vybraná slova dohledat v maximálním množství nářečních slovníků a atlasů z oblasti Východních Karpat (zejména OKDA; OLA; AUM; LAUH; Latta 1991; Onyškevyč 1984; Sabadoš 2008; Varchol–Ivčenko 1990 aj.), vytvořit mapu ke každému slovu a všem jeho doloženým významům, následně získané údaje konfrontovat a až na tomto základě vytvářet slovníkové heslo. Několik takto zpracovaných hesel bychom chtěli představit na BČL. Popsaný způsob zpracování se zdá jako možná cesta k novým velkým lexikografickým projektům – po zpracování veškerého dostupného materiálu by vznikl elektronický nářeční tezaurus karpatorusínské/karpatokrajinské jazykové oblasti. Pro Rusíny by se v podstatě jednalo o celonárodní nářeční slovník, při jehož sestavování lze uplatnit metody používané pro národní dialektologické slovníky (pro SNČJ viz např. Kloferová 2016). To by však byla práce pro mnohočlenný autorský kolektiv na několik desetiletí. Z možností, jak omezit zpracovávaný materiál, se zdá nejracionálnější princip tematických slovníků, tedy zpracování menších tematických skupin lexika (některé tematické slovníky z karpatského areálu již existují jako součást lexikologických studií, např. Jastrems'ka 2008; Čižmárová 2008).

## Bibliografie

- AUM – *Atlas ukrajins'koji movy: u 3 t.* Kyïv 1984–2001.
- ČIŽMÁROVÁ, M. *Nominačné modely v botanickom názvosloví.* Prešov 2008.
- DOLNYCKÁ, M., ŠIŠKOVÁ, R. Z archívu Ivana Paňkevyče. *Slavia* 61, 1992, 255–264.
- DOLNYCKÁ, M. Materiál ke slovníku ukrajinských jihokarpatských nářečí Ivana Paňkevyče. *Slavia* 66, 1997, 81–84.
- DZENDZELIVS'KYJ, J. Ivan Pan'kevyč jak dialektoloh. In: *Naukovyj zbirnyk Muzeju ukrajins'koji kul'tury v Svydnyku* 4 (1), 1969, 167–209.
- JASTREMS'KA, T. *Tradycijne hucul's'ke pastuchuvannja.* Lviv 2008.
- KLOFEROVÁ, S. Ke koncepci celonárodního nářečního slovníku. In: Myznikov, S., Krylova, O., Baklanova, I., eds. *Slavjanskaja dialektnaja leksikografija 2. Materialy konferencii.* Sankt Peterburg 2016, 100–116.
- KRÁLIK, L. *Etymológia a nářečová lexikografia (na materiáli Slovníka slovenských nářečí).* Bratislava 2020.
- LATTA, V. *Linhvistyčnyj atlas ukrajins'kich hovoriv schidnoji Slovaččyny.* Bratislava 1991.
- LAUH – Dzendzelivs'kyj J. O. *Linhvistyčnyj atlas ukrajins'kich narodnych hovoriv Zakarpat'skoji oblasti URSR: u 3 t.* Užhorod 1958–1993.
- OKDA – *Obščekarpatskij dialektologičeskij atlas.* 1989–2003.
- OLA – *Obščeslavjanskij lingvističeskij atlas.* 1988–
- ONYŠKEVYČ, M. J. *Slovnyk bojkivs'kich hovirok.* Kyjiv 1984.
- SABADOŠ, I. *Slovnyk zakarpats'koji hovirky sela Sokyrnycja Chusts'koho rajonu.* Užhorod 2008.
- SKWARSKA, K., KREJČÍŘOVÁ, I., VAŠÍČEK, M. Cesta k digitalizaci slovníků v Oddělení slavistické lingvistiky a lexikografie SLÚ. In: Šlaufová, E.; Skwarska, K.; Čermák, V. (edd.): *Proměny české slavistiky po roce 1989.* Praha 2020, 137–148.
- ŠIŠKOVÁ, R. et al. Lexikální databáze jihokarpatských nářečí z materiálů I. Paňkevyče. Praha 2023. <<https://miracle.slu.cas.cz/ukrajina2/viewer.html>>
- VARBOT, Ž. Ž. O roli etimologii v dialektnoj leksikografii. *Rocznik slawistyczny* 60, 2021, 251–257.
- VARCHOL, N., IVČENKO, A. *Frazeolohičnyj slovnyk lemktiv'skich hovirok schidnoji Slovaččyny.* Prjašiv 1990.
- VAŠÍČEK, M. The *Lexical Database of South Carpathian Dialects Based on Ivan Pankevych's Materials Made Available.* *Slavia* 92 (4), 2023, 458–467.
- VAŠIČKOVA, D. O nekotorych mnogoznačnych ornitonimach v karpatorusinskich dialektach. In: Koporová, K. (ed.) *Studium Carpatho-Ruthenorum 2022.* Prjašiv 2022.
- WIEMER, B., SERŽANT, I. A. East Slavic Dialectology: Achievements and Perspectives of Areal Linguistics. In: Seržant, I. A. and B. Wiemer (eds.), *Contemporary Approaches to Dialectology: The area of North, Northwest Russian and Belarusian vernaculars.* Slavica Bergensia 13. Bergen 2014, 11–80.

15

*Data-based research  
in word formation*

organizers

Jurgis Pakerys

Magda Ševčíková

(ENG)

# Workshop description

The use of language resources has become a standard in many areas of linguistics, including word formation. Recent data-based research in this field has addressed a variety of issues, such as a comparison of derivation across languages (Körtvélyessy et al. 2020), discussions of paradigmatic accounts of word formation (Ruz et al. 2022, Fernández-Domínguez et al. 2020, Hathout & Namer 2019 and other papers in *Morphology* 29:2), investigations of affix rivalry (see Huyghe & Varvara's 2023 paper and other papers in *Word Structure* 16:1; Santana-Lario & Valera 2017), or advantages and disadvantages of modelling word formation without overt derivational material as conversion vs. zero-affixation (Iordăchioaia & Melloni 2023 and the special issue of *Zeitschrift für Sprachwissenschaft* 42:1).

Building on the increasing body of data-based studies on word formation, which have been presented at specialized conferences (such as the *Word-Formation Theories & Typology and Universals in Word-Formation Conference* in Košice) and at conferences with a broader scope (cf. IMM *International Morphology Meeting*, MMM *Mediterranean Morphology Meeting*, ISMo *International Symposium of Morphology*, or the annual meetings of SLE *Societas Linguistica Europaea*), the objective of the proposed workshop is to provide an opportunity for a focused discussion about any type of linguistic research into word formation that exploits text corpora, lexical databases, ontological resources, datasets compiled specifically for a particular purpose, or other sorts of data.

We aim to gather contributions that explore:

- word formation in individual languages or word formation across multiple languages from a cross-linguistic and typological perspective,
- major word-formation processes (in particular, derivation, compounding, and conversion) as well as processes that are minor in the sense that they are used rarely across languages and/or their usage is limited to a small number of languages (reduplication, blending, etc.),
- synchronic or diachronic aspects of word formation,
- qualitative and quantitative evidence on particular word-formation issues, etc.

Differently from events that are primarily interested in the compilation of linguistic data and tools and in computational aspects (such as the *DeriMo Workshop on Resources and Tools for Derivational Morphology*, *SIGMORPHON Workshop on Computational Phonology, Morphology, and Phonetics*, or the wide-scope conferences on *Language Resources and Evaluation LREC*), the actual focus of this workshop is on linguistic insights into the word-formation system of natural languages. However, we expect that in light of these findings a debate will emerge about the data itself. Highly relevant questions are, among others, whether there are general data suitable for different word-formation questions, what kind of word-formation issues require unannotated rather than annotated data, and vice versa, or what challenges need to be overcome to get linguistically interpretable results from the data.

The list of suggested topics is neither exhaustive, nor restrictive. Papers on other aspects of data-based research in word formation or on borderline cases between word formation and other linguistic issues are most welcome.

## References

- Fernández-Domínguez, J. & Bagasheva, A. & Lara-Clares, C. (eds.; 2020): *Paradigmatic Relations in Word Formation*. Leiden: Brill.
- Hathout, N. & Namer, F. (2019): Paradigms in word formation: what are we up to? *Morphology* 29:2, pp. 153–165.
- Huyghe, R. & Varvara, R. (2023): Affix rivalry: Theoretical and methodological challenges. *Word Structure* 16:1, pp. 1–23.
- Iordăchioaia, G. & Melloni, C. (2023): Zero affixes in derivational morphology: Introduction. *Zeitschrift für Sprachwissenschaft* 42:1, pp. 1–11.
- Körtvélyessy, L. & Bagasheva, A. & Štekauer, P. (eds.; 2020): *Derivational Networks across Languages*. Berlin: De Gruyter Mouton.
- Ruz, A. E. & Fernández-Alcaina, C. & Lara-Clares, C. (eds.; 2022): *Paradigms in Word Formation*. Amsterdam: John Benjamins.
- Santana-Lario, J. & Valera, S. (eds.; 2017): *Competing Patterns in English Affixation*. Bern: Peter Lang.

# Table of Contents

*Word-formation in minority languages: corpus linguistics for low-density varieties*

Livio Gaeta (University of Turin)

*From partial to full conversion in English deadjectival nominals: what and where*

María Granados-Serrano, Salvador Valera (University of Granada)

*Denominal prefixed verbs and their valency structure: A corpus study on Czech, English, German and Spanish*

Hana Hledíková, Cristina Fernández-Alcaina (Charles University)

*Syntactic behavior of Czech compounds*

Veronika Kolářová (Charles University)

*Diversity of schemas in English bahuvrihi compounds*

Petr Kos (University of South Bohemia)

*What is go-ing on? Morphological sensitivity (and lack thereof) in the incorporation of -ing ending loanwords in French and Hebrew*

Lior Laks (Bar-Ilan University), Fiammetta Namer (Université de Lorraine & ATILF)

*Testing ChatGPT for the semantic annotation of (competing) doublets*

Cristina Lara-Clares (University of Jaén), Cristina Fernández-Alcaina (Charles University)

*Measuring the degree of transparency of English derivational suffixes*

Verginica Barbu Mititelu<sup>1</sup>, Gianina Iordăchioaia<sup>2</sup>, Svetlozara Leseva<sup>3</sup>, Ivelina Stoyanova<sup>3</sup>  
(<sup>1</sup>Romanian Academy Research Institute for Artificial Intelligence, <sup>2</sup>University of Graz, <sup>3</sup>Institute of the Bulgarian Language, Bulgarian Academy of Sciences)

*Using the Démonette derivational database for research in French morphophonology*

Fiammetta Namer<sup>1</sup>, Nabil Hathout<sup>2</sup>, Stéphanie Lignon<sup>1</sup> (<sup>1</sup>Université de Lorraine & ATILF, <sup>2</sup>CLLE, Université de Toulouse)

*How the measures of derivational productivity depend on lemmatization quality (the case of Lithuanian deverbal nouns)*

Jurgis Pakerys (Vilnius University), Virginijus Dadurkevičius (Vytautas Magnus University), Agnė Navickaitė-Klišauskienė (Vilnius University)

*Potential and actual word-formation lacunas in the system of Slavic Verbal Nouns*

Olena Pchelintseva (Johannes Gutenberg University)

*Inflectional and derivational predictions in morphological families. A case study on Latin*

Matteo Pellegrini, Federica Iurescia, Eleonora Litta, Marco Passarotti (CIRCSE Research Centre, Università Cattolica del Sacro Cuore, Milan)

*Exploring the diachrony of Italian word formation through the Google Books corpus: The case of NN compounds and their rival syntagmatic NPN counterparts*

Jan Radimský (University of South Bohemia in České Budějovice), Silvia Micheli, Pavel Štichauer (Charles University, Prague)

*Creating and exploiting a lexical database of deverbal nouns in French*

Rossella Varvara, Justine Salvadori, Richard Huyghe (University of Fribourg)

## Word-formation in minority languages: corpus linguistics for low-density varieties

Livio Gaeta (University of Turin)

Corpus linguistics normally deals with large text corpora drawn from standardized languages for which different text types can be collected and accordingly made searchable in a database. Far less usual is working on minority languages providing small sets of data scarcely or not standardized, for which the label ‘low-density languages’ has recently been suggested (Maxwell & Hughes 2006). This term identifies those languages which – in neat contrast to the few ‘high-density languages’ – lack in computational resources, in particular texts provided with linguistic annotation, which are a prerequisite for Natural Language Processing. Besides the difficulty of collecting data, building a corpus in this context must deal with the dramatic variation present in the sources. This is the case of Walser German varieties spoken in North-Western Italy on which an ongoing research project is based. Besides extensive contact with Romance (including Italian, Franco-Provençal and French), the Walser German varieties are exposed to a dramatic process of language decay (Dal Negro 2004, Zürrer 2009). In the paper, the problems raised by the specific case of endangered languages will be discussed which require different solutions about what is normally done in corpus linguistics. In particular, a specific platform has been developed for hosting and elaborating the data which will be illustrated in detail.

On this basis, a significant number of word-formation processes will be investigated drawn from the big bunch of derivational processes typically occurring in a language. Focusing particularly on Titsch and Töitschu, spoken respectively in Gressoney and Issime, we will show similarities and differences between the two varieties with regard to abstract (1a) and agent nouns (1b), verb-forming patterns including conversion (1c) and suffixation (1d), verbal prefixation (1e) and phrasal verb formation (1f), adjective-forming patterns (1g). Thanks to concrete examples extracted from the corpus, we will show the different uses and meanings displaying the different patterns, which partially reflect contact-based models like action nouns (or *nomina actionis*) (2a) vs. semelfactive action nouns (or *nomina vicis*, Tovenà & Donazzan 2017) (2b), particle verbs (2c) vs. phrasal verbs (2d), etc. Besides pointing out contact-induced phenomena like for instance the role of Italian and Franco-Provençal semelfactive abstracts like respectively *occhio* ‘eye’ → *occhiata* ‘glance’ and *pià* ‘foot’ → *piataye* ‘kick’, we will also consider relevant differences regarding the Standard German variety. In particular, we will focus on compounds insofar as they reflect different linking elements as in Titsch *maschinfierer* ‘car driver’ vs. German *Maschinenführer* ‘engine driver’. Considering the broad spectrum of data collected from the corpus and the dictionary, which consistently alters the picture obtained by merely comparing the standard varieties as carried out in typological research (Haspelmath 2001), this research fosters the development of similar research projects on minority languages in Europe with the aim of enriching the picture of word formation phenomena on the European continent compared with what has been gathered for standard languages (Müller et al. 2016).



	<b>Titsch</b>	<b>Töitschu</b>
1a.	<i>fiere</i> ‘to lead’ → <i>fieròng</i> ‘direction’ <i>toufe</i> ‘to baptize’ → <i>toufé</i> ‘baptism’	<i>rechtnun</i> ‘to count’ → <i>rechtnunh</i> ‘count’ <i>gschwellen</i> ‘to blow’ → <i>gschwulsch</i> ‘bulge’
1b.	<i>afenge</i> ‘to begin’ → <i>afenger</i> ‘beginner’ <i>chnätte</i> ‘to reproach’ → <i>chnätter</i> ‘contentious person’	<i>bachen</i> ‘to bake’ → <i>bacher</i> ‘baker’ <i>chringen</i> ‘to fight’ → <i>chriger</i> ‘brawler’
1c.	<i>loufe</i> ‘to run’ → <i>louf</i> ‘run’ <i>redò</i> ‘to talk’ → <i>red</i> ‘talk’	<i>chringen</i> ‘to fight’ → <i>chrig</i> ‘fight’ <i>trommun</i> ‘to cut’ → <i>trom</i> ‘cut’
1d.	<i>adresse</i> ‘address’ → <i>adressiere</i> ‘to address’ <i>chälb</i> ‘calf’ → <i>chälbrò</i> ‘to calve’	<i>ei</i> ‘egg’ → <i>eiurun</i> ‘to whey-bubble’ <i>kubbiu</i> ‘pair’ → <i>kubbiurun</i> ‘to pair’
1e.	<i>abbräche</i> ‘to break off’ <i>dòrchbräche</i> ‘to break through’	<i>ousbrechen</i> ‘to perforate’ <i>unnerbrechen</i> ‘to interrupt’
1f.	–	<i>brechen ab / ous</i> ‘to warn off / out’
1g.	<i>gòscht</i> ‘taste’ → <i>gòschtléch</i> ‘tasty’ <i>foare</i> ‘to drive’ → <i>foarbar</i> ‘drivable’	<i>bljüten</i> ‘to blood’ → <i>bljütig</i> ‘bleeding’ <i>birhu</i> ‘birch’ → <i>birhen</i> ‘birchwood’

Tab. 1: Derivational patterns in Titsch and Töitschu

2a.	<i>was</i> <i>anbelangt</i> <i>de</i> <i>bode</i> , <i>z'schnue</i> , <i>z'entlecke</i> <i>vòn</i> <i>de</i> what regards DEF earth DEF=snow DEF=dismember.INF of DEF <i>loubene</i> <i>òn</i> <i>d'ändrònge</i> <i>vòm</i> <i>klima</i> , avalanche and DEF=change von.DEF clima 'regarding the earth, the snow, the dismemberment of the avalanche and the climate change'
2b.	<i>em</i> <i>louf</i> <i>vòn</i> <i>de</i> <i>joare</i> <i>hät</i> <i>de</i> <i>tracht</i> <i>em</i> <i>grösse</i> in.DEF course of DEF years has DEF costume in.DEF big <i>òn</i> <i>em</i> <i>ganze</i> <i>en</i> <i>ändertògmacht</i> and in.DEF whole in.DEF change made 'in the course of the years the costume underwent on the whole a change'
2c.	<i>Wenn</i> <i>d'nässe</i> <i>wéreder</i> <i>z'heissé</i> , <i>leck</i> <i>bròt</i> <i>òn</i> if DEF=soup were.2SG.DAT to=hot put bread and <i>chésch dré</i> , <i>sotte</i> <i>erkoaltetsch</i> cheese inside, so cool.3SG.REFL'. 'if the soup were too hot for you, put bread and cheese inside, so that it cools down'.
2d.	<i>un</i> <i>d'Schat</i> <i>llëis</i> <i>llëis</i> <i>müntrut</i> <i>ouf</i> <i>wider</i> <i>den</i> <i>Groat</i> and DEF=shade slow slow climbs up against DEF mountain 'and the shade slowly climbs up towards the mountain'

Tab. 2: Corpus-based examples of Titsch and Töitschu word-formation patterns

## Bibliography

Dal Negro S. (2004). *The Decay of a Language*. Bern.

Haspelmath M. (2001). "The European linguistic area: Standard Average European", in M. Haspelmath et al. (eds.). *Language Typology and Language Universals*. Vol. 2, 1492–1510. Berlin.

Maxwell M. & Hughes B. (2006). "Frontiers in Linguistic Annotation for Lower-Density Languages", in *Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora 2006*. Association for Computational Linguistics: 29-37.

Müller P. O., Ohnheiser I., Olsen S. & Rainer F. (eds.) (2016). *Word-Formation*. Vol. 4. Berlin.

Tovena L. & Donazzan M. (2017). "Italian *-ata* event nouns and the *nomen vicis* interpretation". *Italian Journal of Linguistics* 29.1: 75-100.

Zürrer P. (2009). *Sprachkontakt in Walser Dialekten: Gressoney und Issime im Aostatal (Italien)*. Stuttgart.

## From partial to full conversion in English deadjectival nominals: what and where

María Granados-Serrano, Salvador Valera (University of Granada)

The separation partial vs. complete, total or full conversion is a rather traditional distinction that has been superseded in several approaches as different types of conversion (syntactic vs. morphological), especially in the formation of deadjectival nominals of the type *the rich* or *the impossible* (for the view of partial vs. full conversion, cf. Sweet 1891–98, I: 38–39; Quirk *et al.* 1985: 1559; Tournier 1985: 175; for the view of syntactic vs. morphological conversion, cf. Plag 2013, Iordăchioaia 2023).

This paper is an overview of deadjectival nominals that can be considered full conversion on account of their nominal inflection according to the *British National Corpus* (hereafter, BNC) and the *Corpus of Contemporary American English* (hereafter, COCA). In both corpora, the paper relies on the syntax query [*the + adj suffix + -s*] for data retrieval with a view to reliable contrast of as many cases of both partial and full conversion as possible, i.e. with and without nominal inflection (cf. Payne & Huddleston 2002). The paper then presents the occurrence and distribution of formations that arise from morphological complex adjectives (e.g. formations in *-ed* or *-ful*) with and without nominal inflection (e.g. *the coloureds* vs. *the deceased*, or *the hopefuls* vs. *the powerful(s)*), and with any relevant Determiners and/or Modifiers:

- (1) Some of *these false coloureds* are coming out of the closet (COCA)
- (2) *Many songwriting hopefuls* never get that far (BNC)

The paper takes into consideration all the formations with adjectival suffixes, including those derivatives with nominal counterparts, such as formations in *-ian* (e.g. *the utopians*) or *-oid* (e.g. *the dramatoids*) (see Bauer *et al.* 2013: 288), and compares the results obtained for an outline of:

- i) which formations allow or disallow nominal inflection, specifically with regard to the morphological type of the presumed adjectival base and of its suffix,
- ii) in which of the two varieties,
- iii) in which period (considering the COCA covers a wider timespan than the BNC and the differences, if any, may be later developments than the ones recorded in the BNC), and
- iv) in which of the corpus sample types.

This paper does not go into the question whether this should be described in one or the other of the terms presented in the opening paragraph of this abstract, as they probably capture a part of the picture each, but it is intended as a step towards the identification of where full lexicalization takes place, according to the corpus data and to what major dictionaries may contribute, e.g. the *Oxford English Dictionary* (OED) (for the review of lexicalization, cf. Lipka 2004 or Hohenhaus 2005).

The results display similarities and dissimilarities in usage in units like *coloureds* and *hopefuls*, respectively, even if both appear to display similar properties as regards lexicalization as nouns. At the same time, the results put a spotlight on the units in *-ed* or *-ful* (to use the same morphological type of examples as above) that display nominal inflection (i.e. what the traditional approach calls *full conversion*) compared with those that do not and, in principle, could do so too (i.e. what the traditional approach calls *partial conversion*), and it does both quantitatively and qualitatively.

## Bibliography

- Bauer, L., R. Lieber & I. Plag. 2013. *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press.
- Davies, M. 2004–. *British National Corpus* (from OUP). Available online at <https://www.english-corpora.org/bnc/>.
- Davies, M. 2008–. *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.
- Hohenhaus, P. 2005. Lexicalization and institutionalization. In P. Štekauer & R. Lieber (eds.), *Handbook of Word-Formation*. Dordrecht: Springer, 353–373.
- Iordăchioaia, G. 2023. Morphological vs. syntactic conversion: Diagnostics and analytical means. *SKASE Journal of Theoretical Linguistics*, 20(4): 57–73.
- Lipka, L. 2004. *SKASE Journal of Theoretical Linguistics*, 1(1): 2–19.
- Plag, I. 2013. *Word-formation in English*. Cambridge: Cambridge University Press.
- Payne, J. & R. Huddleston. 2002. Nouns and noun phrases. In R. Huddleston & G. K. Pullum (eds), *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press; 323–523.
- Proffitt, M. 2019–. *The Oxford English Dictionary*. Available online at <http://www.oed.com>.
- Quirk, R., S. Greenbaum, G. Leech & J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Sweet, H. 1891–98. *A New English Grammar: Logical and Historical*. Oxford: Clarendon Press.
- Tournier, J. 1985. *Introduction Descriptive à la Lexicogénétique de l'Anglais Contemporain*. Genève: Editions Slaktine.

**Denominal Prefixed Verbs and Their Valency Structure:  
A Corpus Study on Czech, English, German and Spanish**

Hana Hledíková, Cristina Fernández-Alcaina  
(Institute of Formal and Applied Linguistics, Charles University)

Prefixed verbs like those in example (1) present a special case of verb formation in languages where verbs are most often formed by prefixation from other verbs or by conversion from other word classes. Because in these cases the intermediate unprefixed verb (*\*lodit*, *\*to horn*, *\*florar*, *\*sklaven*) is not attested, the formation of such verbs may also be described as the simultaneous use of two formal means, prefixation and the addition of verbal inflectional markers (e.g., Fleischer 2012; RAE 2009).

- |                                                                                                                                              |                                                                                                                                    |
|----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
| <p>(1) <i>na-lod'-i-t</i> &lt; <i>lod'</i><br/>         PREF-ship-THEME-INF &lt; ship<br/>         'put on a ship' &lt; 'ship'</p>           | <p><i>de-horn</i> &lt; <i>horn</i><br/>         PREF-horn &lt; horn<br/>         'remove the horn' &lt; 'horn'</p>                 |
| <p><i>en-flor-a-r</i> &lt; <i>flor</i><br/>         PREF-flower-THEME-INF &lt; flower<br/>         'decorate with flowers' &lt; 'flower'</p> | <p><i>ver-sklav-en</i> &lt; <i>Sklave</i><br/>         PREF-slave-INF &lt; slave<br/>         'make into a slave' &lt; 'slave'</p> |

One point of interest which has been considered in connection with these formations is how to account for the observed regularities in the verbs' valency structure, i.e., the number and types of syntactic arguments they occur with. These regularities may be attributed both to the form of the prefix and the semantic relation between the verb and the base noun. The verb typically denotes a situation in which the motivating noun plays a certain role, and based on this, it is possible to distinguish different semantic types of verbs – as can be seen in the different paraphrases in (1). The semantic types are connected with the verb's valency characteristics, because each type of generalized situation is typically associated with a different set of participants (e.g., Baeskow 2023).

This contribution investigates the regularities in the valency behaviour of denominal prefixed verbs in four Indo-European languages: Czech, English, German and Spanish. In all four languages, it is possible to form a prefixed verb directly from a noun and the formations are comparable in their structure (although in English, there is no other formal change that accompanies the addition of the prefix due to the absence of inflectional markers on the citation forms of verbs).

In each language, prefixed denominal verbs for which the unprefixed verb is not attested were extracted from a list of verbal lemmas annotated for their morphemic structure. The verbs were annotated for their semantic type based on the semantic relationship between the verb and the motivating noun. For each verb, its corpus concordances were extracted from comparable corpora and twenty (or all available) concordances were manually annotated for the type of syntactic arguments the verb occurs with, using functor labels from the functional generative description framework (Lopatková et al. 2016). The corpora used and the size of the final sample is shown in Table 1.

Based on the annotated concordances, we investigate to what degree the verb's prefix and semantic type determines the types of valency patterns it occurs with. We show that in all four languages, the verb's semantic type provides more information about the valency patterns than the form of the prefix. The sample also documents that across all four languages, verbs of certain semantic types often occur with a valency pattern in which the base noun's semantic role is also expressed as a syntactic argument (e.g., *nalodit zboží na loď* / *na člun* 'to on-ship goods on the ship / on the boat'). These instances are further investigated in terms of whether the syntactic argument is expressed using the same lexical root as the base noun and in terms of their function to explain the similarities observed across the examined languages.

Language	Corpus	Size	Time period	Genres	Denominal prefixed verbs (lemmas)	Annotated concordances
Czech	SYN2000 (Čermák et al. 2000)	100 mil.	1990–1999	fiction: 33% non-fiction: 33% newspaper: 33%	240	3 494
English	BNC (BNC Consortium 2007)	100 mil.	1960s–1993	written: 90% – imaginative: 19% – informative: 81% spoken: 10%	82	887
German	DWDS (Geyken 2007)	120 mil.	1900–1999	fiction: 26.35% newspaper: 27.29% academic: 24.59% non-fiction: 21.77%	211	2 133
Spanish	CREA (Real Academia Española 2014)	120 mil.	1975–2000	fiction: 25% academic + newspaper: 75%	812	7 445

**Table 1:** List of used corpora and their description, along with the size of the final sample of verbs

### Bibliography

- Baeskow, H. (2023). A cognitive approach to the allegedly left-headed prefix verbs in German: Arguments for the interaction of prefixation and event-schema metonymy. *Linguistik Online*, 120(2), 3–40.
- Fleischer, W. (2012). *Wortbildung der deutschen Gegenwartssprache*. De Gruyter.
- Lopatková, M., Kettnerová, V., Bejček, E., Vernerová, A., & Žabokrtský, Z. (2016). *Valenční slovník českých sloves—VALLEX*. Karolinum.
- Real Academia Española, & Asociación de Academias de la Lengua Española (Eds.). (2009). *Nueva gramática de la lengua española*. Espasa libros.

### Language resources

- BNC Consortium. (2007). *The British National Corpus, XML Edition*. Oxford Text Archive. <http://hdl.handle.net/20.500.12024/2554>.
- Čermák, F., Blatná, R., Hlaváčová, J., Klímová, J., Koček, J., Kopřivová, M., Křen, M., Petkevič, V., Schmiedtová, V., & Šulc, M. (2000). *SYN2000: Žánrově vyvážený korpus psané češtiny*. Ústav Českého národního korpusu FF UK.
- Geyken, A. (2007). The DWDS corpus: A reference corpus for the German language of the 20th century. In *Idioms and Collocations: Corpus-based Linguistic and Lexicographic Studies* (pp. 23–40). Continuum International Publishing Group.
- Real Academia Española. (2014). *El Corpus de referencia del español actual (CREA)*. <http://hdl.handle.net/11372/LRT-895>.

### Acknowledgements

This contribution was supported by the Charles University, project GA UK No. 246723, by the SVV project No. 260 698, Charles University Research Centre program No. 24/SSH/009, and by the Ministry of Education, Youth and Sports of the Czech Republic, Project No. LM2023062 LINDAT/CLARIAH-CZ, and by the Grant Agency of the Czech Republic under the EXPRO program as project “LUSyD” (project No. GX20-16819X).

## Syntactic behavior of Czech compounds

Veronika Kolářová (Charles University)

Compoundness is defined as a property of a complex word (compound) having at least two components with lexical meaning (Ološtiak – Vojteková 2021). Linguists' interest in syntactic behavior of compounds dates back to the works by Roeper and Siegel (1978) and Grimshaw (1990). In recent works (Borer 2013; Alexiadou 2017), morpho-syntactic properties of nominal synthetic compounds (*tomato-growing*, *tomato-growth*) are compared with those of nominals displaying argument structure (*the demolition of the house by the army in 2 hours*), including for example possibility to realize arguments (*the house demolition (\*by the army) (\*in two hours)*). Concerning the relation of the non-head (usually the first constituent of the compound) to the head (the compound's second constituent), some compounds with deverbal head are considered to be subject to restrictions on what role of the base verb the non-head of the compound can express (cf. *story-telling to children* vs. *\*children-telling of stories*; Levin – Rappaport Hovav 2005; Alexiadou 2017). With respect to the relations between valency of a whole compound and that of its head or non-head, Lapshinova-Koltunski (2008) differentiates several types of German nominal compounds: (i) valency of the whole compound is determined either by its head (*das Grundproblem, dass...* 'the basic problem, that...') or by its non-head (*das Beweismittel, dass...* 'the means of evidence, that...'), (ii) the compound has its own valency (*der Ehrgeiz, dass...* 'the ambition that...').

In this study, syntactic behavior of Czech noun and adjectival compounds will be examined (*blahopřání* 'congratulation', *ohleduplný* 'regardful'; Bozděchová 1994; Bozděchová – Wagner 2016). Compounds in which one of the complements of their head is saturated by the non-head, be it the head's agent (*krupobití* 'hail'), theme (*důvěryhodný* 'trustworthy'), "complement" (*živorodý* 'viviparous') or manner (*rovnoběžný* 'parallel'), deserve close attention. The aim is to explore how valency properties of compound's constituents contribute to valency of the compound as a whole.

To gather the language material for the research, Czech noun and adjectival compounds were searched for in DeriNet (Vidra et al. 2021) and in Czech corpora (using the PaReNT tool; Svoboda – Ševčíková 2024). For this study, compounds with the pronominal non-head (*samospráva* 'self-government') were excluded. In order to limit the list to compounds with constituents that are likely to be valent, the predicted components were set against the indexes of valency lexicons NomVallex (Kolářová – Vernerová – Klímová 2022) and PDT-Vallex (Urešová et al. 2021), and the list was sorted out manually. Compounds in the resulting list (about 300 items) pose the following research questions:

(i) In what forms can the non-head in Czech compounds be expressed? Cf. *ohledu<sub>GEN</sub>plný* 'regardful', *bohu<sub>DAT</sub>libý* 'meritorious', *srdce<sub>ACC</sub>rvoucí* 'heartbreaking'.

(ii) Working on the assumption that one of the complements of the head is saturated by the non-head, we explore the employment of the other prospective head's complements – are they likely to be expressed as complements of the whole compound, as specified in the valency frame in (1)?

(1) *blahopřání* 'congratulations': ACT<sub>Gen,poss,od+2,\*Ins</sub> ADDR<sub>Dat</sub> PAT<sub>k+Dat</sub>

(iii) What happens with the valency complements of the non-head when being a part of a compound? In some cases it is expressed as a complement of the whole compound (either in the same morphemic form (2–3) or in a different form (4)), but mostly it is not employed at all (5).

(2) *díkůvzdání světici za ochranu Prahy* 'thanksgiving to a saint for the protection of Prague'

(3) *láskyplný k lidem* 'full of love for people'

(4) *škodolibý vůči lidem* 'spiteful towards people' < *mající zálibu ve školení někomu* 'liking for doing harm to someone'

(5) *bojechtivý* 'combative' < *chtivý boje (s někým)* 'avid for fight (with someone)'

Syntactic behavior of Czech compounds is influenced by valency of their constituents, however, as the compounds take their own meaning, they – as a whole – also display their own valency.

## Bibliography

- ALEXIADOU, A. (2017). On the complex relationship between deverbal compounds and argument supporting nominals. In M. Bloch-Trojnar, A. Malicka-Kleparska (eds.), *Aspect and Valency in Nominals*, Berlin, Boston: De Gruyter Mouton, 53–82.
- BORER, H. (2013). *Structuring Sense: Volume III: Taking form*. Oxford University Press, Oxford.
- BOZDĚCHOVÁ, I. (1994). *Tvoření slov skládáním*. Praha: Institut sociálních vztahů.
- BOZDĚCHOVÁ, I. – WAGNER, R. (2016). Kompozitum. In P. Karlík, M. Nekula, J. Pleskalová (ed.): *Nový encyklopedický slovník češtiny*, Praha: Nakladatelství Lidové noviny, 845–851.
- GRIMSHAW, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.
- KOLÁŘOVÁ, V. – VERNEROVÁ, A. – KLÍMOVÁ, J. (2022). *NomVallex 2.0*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. Available at <http://hdl.handle.net/11234/1-4663>.
- LAPSHINOVA-KOLTUNSKI, E. (2008). Non-heads of Compounds as Valency Bearers: Extraction from Corpora, Classification and Implication for Dictionaries. In E. Bernal, J. DeCesaris (eds.), *Proceedings of the XIII EURALEX International Congress*, Barcelona: Universitat Pompeu Fabra, 449–456.
- LEVIN, B. C. – RAPPAPORT HOVAV, M. (2005). *Argument Realization*. Cambridge University Press, Cambridge, UK.
- OLOŠTIAK, M. – VOJTEKOVÁ, M. (2021). Kompozitnosť a kompozícia: príspevok k charakteristike zložených slov na materiáli západoslovanských jazykov. *Slovo a slovesnosť*, 82(2), 95–117.
- ROEPER, T. – SIEGEL, M. E. A. (1978). A Lexical Transformation for Verbal Compounds. *Linguistic Inquiry*, 9(2), 199–260.
- SVOBODA, E. – ŠEVČÍKOVÁ, M. (2024). PaReNT (Parent Retrieval Neural Tool): A Deep Dive into Word Formation across Languages. In N. Calzolari et al. (eds.), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, Torino, Italia: ELRA and ICCL, 12611–12621.
- UREŠOVÁ, Z. et al. (2021). *PDT-Vallex: Czech Valency lexicon linked to treebanks 4.0 (PDT-Vallex 4.0)*. LINDAT/CLARIAH-CZ repository, <http://hdl.handle.net/11234/1-3499>.
- VIDRA, J. et al. (2021). *DeriNet 2.1*. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11234/1-3765>.

## Diversity of schemas in English bahuvrihi compounds

Petr Kos (University of South Bohemia)

On an analysis of data drawn from the Araneum Anglicum Maius corpus, the paper aims to demonstrate that the formation of English bahuvrihi compounds with the right-hand constituent *head* arises from different and diversely specific schemas, as conceived within the Relational Morphology theory (Jackendoff & Audring, 2020), although they superficially share the same form. The paper will also provide the whole range of schemas instrumental in the formation of this type of compounds and their relatedness on both vertical and horizontal axes.

Within the Relational Morphology theory (Jackendoff & Audring, 2020), the lexicon comprises not only individual lexemes but also schemas. On the basis of the same-except principle, schemas express parallelism within a family of words that are the same in some respect, manifested as constants in the schema (e.g., a suffix or a fixed compounding base), and different in others, manifested as variables. Schemas have two different roles, the relational and the generative. Within the relational role, schemas provide a structure to the lexicon, and within the generative role they provide templates for the formation of new lexemes (word-formation) or forms of lexemes (inflection).

The search in the Araneum Anglicum Maius corpus has provided a sample of ca 400 English bahuvrihi compounds with the right-hand constituent *head*. By abstracting from semantic commonalities, the lexemes in the sample have been grouped into specific series representing, or instantiating, individual schemas.

The schemas are hierarchically structured, with higher-level schemas having the right-hand constituent as a variable (cf. the general schema for humans, instantiated, for instance, by *bighead* as a term for a conceited person, in which the right-hand constituent may be instantiated by a different lexeme representing a different part of a whole, as in *loudmouth*) and lower-level schemas having the right-hand constituent as the constant *head* (cf. the specific schema for drug addicts, which does not allow for a different right-hand constituent but *head*, as in *acidhead*, *methhead*, etc.).

The schemas also demonstrate relatedness on a horizontal axis since different specific sub-schemas are on the same level of hierarchy relative to the same superordinate schema, as exemplified by sub-schemas for drug addicts and different types of enthusiasts.

Some schemas demonstrate a partial overlap since individual lexemes may instantiate more specific schemas; for instance, *pizzahead* as a derogatory term for a person of Italian descent instantiates both a schema for derogatory terms denoting ethnic groups and a schema for enthusiasts and *dothead* as a derogatory term for a Hindu person instantiates both a schema for derogatory terms denoting ethnic groups and a schema for humans based on a physical feature. The formation of such lexemes is thus prompted by the two schemas simultaneously.

Since they are conceived of as categories, individual schemas may have both central and peripheral members. For instance, the left-hand variable in the schema for drug addicts is, in its vast majority, instantiated by a lexeme denoting a drug, these being the central members of the category, while we may find members of the same category that have the left-hand variable instantiated by a different concept, such as *baghead* and *tinhead*, *bag* and *tin* being devices that facilitate the drug consumption.

The presentation will also challenge the commonly accepted notion that bahuvrihi compounds denoting humans are necessarily derogatory since this derogatory connotation is absent in some specific schemas, e.g., in those denoting music enthusiasts, such as *jazzhead* and *rockhead*.

The presentation will first provide a brief introduction to the theoretical background of the research and then discuss and illustrate the individual issues with a rich variety of examples. It will also provide a full overview of both high-level and low-level schemas that emerge from the sample of lexemes provided by the corpus.



## **Bibliography**

Jackendoff, R. & Audring, J. (2020). *The Texture of the Lexicon: Relational Morphology and the Parallel Architecture*. Oxford: Oxford University Press.

**“What is go-ing on? Morphological sensitivity (and lack thereof)  
in the incorporation of -ing ending loanwords in French and Hebrew”**

Lior Laks (Bar-Ilan University), Fiammetta Namer (Université de Lorraine & ATILF)

Borrowings from English has diverse impacts on new word formation in the target languages [17] (see, e.g. [10] on the typology of borrowing). Our study examines on the adaptation of *Xing* nouns borrowed from English in two different linguistic systems, and how these nouns are used (or not) for verb formation. By means of a large corpus of *Xing* action nouns recently borrowed into French (405) and Hebrew (83), we compare these new formations in both languages. Data collection is essentially based on two corpora: Leipzig Corpora Collection for French<sup>9</sup> and HeTenTen for Hebrew<sup>10</sup>. We illustrate how the formation of neologisms diverges since French verb formation relies mostly on affixation, while Hebrew relies on non-concatenative morphology, i.e. the combination of a root and pattern. Examination of our data reveals the following generalizations. Both languages allow periphrastic verbal constructions, where the loanword follows the native verb ‘do’ (“*faire du shopping*” 1a, “*laasot šoping*” 2a). In addition, 50,6% of *Xing* nouns (*shopping.N*) are the base of a back-formed verb ([2], [16]) in French (*shopper*, 1b), and 12% of them are in a noun-to-verb conversion relation (*shoppinguer*, 1c). In contrast, the proportion of morphologically formed verbs based on *Xing* nouns in Hebrew is much lower (8%), and all are formed in the *CiCeC* pattern (*šipeng*, 2b). These observations are consistent with the properties of the two morphological systems. In French, verb neologisms surface in the first conjugation class (-*er* infinitive marking), and verbs are part of two types of derivational relations with their related action noun: conversion [18] or verb-to-noun suffixation. In Hebrew, all verbs are formed in patterns, where *CiCeC* is the most productive one ([3], [4], [5], [15], [13], [1], [19] among others). Moreover, both languages allow morphology-syntax competition, where periphrastic constructions are preferred, especially in Hebrew. In French, there is also competition between two verb formation strategies; the family of the borrowed noun contains a converted verb only when it also includes a back-formed verb. Finally, newly created verb(s)-noun relations in French expand into derivational paradigms [11] including agents (*shoppeur*, *shoppeuse*) locations (*shopperie*), etc. realizing action networks ([8],[14]).

- (1) a. Quand on **fait du shopping** il y a certaines choses à garder en tête<sup>11</sup>  
‘When one **does shopping\_N**, there are a few things to keep in mind’  
b. Bien **shopper** pendant les soldes, ça s’apprend.  
‘To **shop\_V** well during the sales can be learned’  
c. devenir copines...boire des verres, prendre l’air, **shoppinguer!**  
‘become friends...have a drink, get some fresh air, **shopping\_V!**’
- (2) a. im atem mexapsim efo **laasot šoping** be-prag - ze bišvilxem  
‘If you are looking for where to so shopping in Prague - this is for you’  
b. gam al i boded emca efo **lešapeng**  
‘also on an isolated island I will find where **to do shopping**’

The talk will rely on the above results to explain the way these two morphologically distinct languages form verbs based on the same borrowed nouns. First, they treat the internal structure of the loanword differently. French speakers recognize *-ing* as a morphological marker of action nouns, however foreign [9]. Conversely, Hebrew verb formation seems to be blocked due to the nominal *-ing* ending, while with other borrowed nouns, verb formation is more productive (*binjv* ‘binge’ / *binjevjv* ‘binge watch’). Second, French verb formation relies mainly on back-formation, although it is an irregular process in French, because it tends to preserve the optimal prosodic word [12], whereas *CiCeC* formation in Hebrew preserves the *ng* cluster and is faithful to the base.

<sup>9</sup> <https://wortschatz.uni-leipzig.de/en/download/French>

<sup>10</sup> <https://www.sketchengine.eu/hetenten-hebrew-corpus/>

<sup>11</sup> All examples are from the web.

## Bibliography

- [1] Bat-El, O. (1994). "Stem modification and cluster transfer in Modern Hebrew". *Natural Language and Linguistic Theory* 12, 572-596.
- [2] Becker, T. (1993). "Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology". *Yearbook of Morphology* 1992, 1-27.
- [3] Berman, R.A. (1978). *Modern Hebrew Structure*. University Publishing Projects, Tel-Aviv.
- [4] Bolozky, S. (1978). "Word formation strategies in Modern Hebrew verb system: denominative Verbs". *Afroasiatic Linguistics* 5, 1-26.
- [5] Bolozky, S. (1999). *Measuring productivity in word formation: The case of Israeli Hebrew*. Brill, Leiden.
- [6] Braselmann, P. (2015). "98-Word-formation and purism in French". *Word-Formation: An International Handbook of the Languages of Europe*, P. O. Müller, I. Ohnheiser, S. Olsen & F. Rainer (eds.). De Gruyter Mouton, Berlin / New York, 1708-1719.
- [7] Chesley, P. & Baayen H. (2010). "Predicting new words from newer words: Lexical borrowings in French". *Linguistics* 48-4, 1343-1374.
- [8] Fradin, B. (2020). "Characterizing derivational paradigms. Paradigmatic relations in derivational morphology". A. Bagasheva, J. Fernández-Domínguez & C. Lara-Clares (eds.). Koninklijke Brill NV, Leiden, 49-84.
- [9] Gardani, F. (2018). "On morphological borrowing". *Language and Linguistic Compass* 12(10), 1-17.
- [10] Haspelmath, M. & Tadmor, U. (eds.). (2009). *Loanwords in the World's Languages*. Mouton de Gruyter, Berlin.
- [11] Hathout, N. & Namer, F. (2019). "Paradigms in word formation: what are we up to?" *Morphology* 29(2), 153-165.
- [12] Plénat, M. (2000). "Quelques thèmes de recherche actuels en morphophonologie française". *Cahiers de lexicologie* 77, 27-62.
- [13] Ravid, D. (1990). "Internal structure constraints on new-word formation devices in Modern Hebrew". *Folia Linguistica* 24, 289-347.
- [14] Roché, M. (2023). "Les familles dérivationnelles : comment ça marche ?". *Lexique* 33: 166-209.
- [15] Schwarzwald, O.R. (1981). *Grammar and Reality in the Hebrew Verb*. Bar Ilan University Press, Ramat Gan. (in Hebrew)
- [16] Štekauer, P. (2015). "18-Backformation". *Word-Formation. An International Handbook of the Languages of Europe*. P. O. Müller, I. Ohnheiser, S. Olsen & F. Rainer eds. De Gruyter, Berlin / New York, 340-352.
- [17] Ten Hacken, P. & Panocová, R. (eds.). (2020). *The interaction of borrowing and word formation*. Edinburgh University Press, Edinburgh.
- [18] Tribout, D. (2020). "Nominalization, verbalization or both? Insights from the directionality of noun-verb conversion in French". *Zeitschrift für Wortbildung / Journal of Word Formation* 4(2), 187-207.
- [19] Ussishkin, A. (2005). "A fixed prosodic theory of nonconcatenative templatic morphology". *Natural Language and Linguistic Theory* 23, 169-218.

## Testing ChatGPT for the semantic annotation of (competing) doublets

Cristina Lara-Clares (University of Jaén), Cristina Fernández-Alcaina (Charles University)

Competition (or rivalry) in morphology has as of late received increasing attention, partly because of its role in the productivity of word-formation processes and also within research on paradigms (see, among others, Plag 1999, MacWhinney et al. 2014, Santana-Lario & Valera 2017, Rainer et al. 2019 and Huyghe & Varvara 2023). However, the definition of the concept has been subject to variability across studies, with some taking a broader view where any two (or more) processes expressing a similar meaning are considered as competitors, and a narrower view, where processes need to be used on the same base(s) and express the same meaning. This paper takes the more restricted view on competition.

Previous research on competition has demonstrated the need to analyze competition at the sense level, even more so when frequency analyses are performed (Fernández-Alcaina 2021, Lara-Clares 2023). Nevertheless, the manual semantic annotation of large quantities of data proves unfeasible. This paper seeks to explore alternatives for the semantic annotation of corpus concordances, making use of large language models (LLMs), which have potential applications to linguistic research (see, e.g. Grabowski 2023). Specifically, the analysis of concordance lines would allow to quantify the frequency of use of particular senses with the aim to explore the degree of semantic overlap between competitors (if any).

To this aim, tests will be performed on pairs of words derived from the same base where there is semantic overlap (e.g. *wrap*<sup>N</sup> and *wrapper*) and where there is not (e.g. *winter*<sup>V</sup> and *winterize*). They will be selected from a database on diachronic nominal and verbal competition extracted from both the *Oxford English Dictionary* (OED) and the *British National Corpus* (BNC, Davies 2004–). Doublets must fulfil the following conditions: i) being derived from the same base but by different suffixes, ii) overlapping in at least one sense based on the definitions of the OED, and iii) being attested in the BNC.

The preliminary results for the doublet *wrap*<sup>N</sup>/*wrapper* are presented here for illustration purposes. The two nouns overlap in the senses ‘a material that is used to cover or protect things’ and ‘a piece of clothing’. They also have similar frequencies in the BNC (*wrap* has a frequency of 290, and *wrapper*, 174). The test consists in the semantic manual annotation of their concordance lines in the BNC individually by the of authors this paper (in order to allow inter-rater reliability tests) and also by using ChatGPT-4 (OpenAI, 2024) to compare the results obtained with that of automatic annotation. In particular, for *wrap*<sup>N</sup>/*wrapper*, four different prompts are given to ChatGPT-4 using the definitions from four reference dictionaries (OED, *Cambridge dictionary*, *Collins dictionary*, and *Merriam-Webster dictionary*). The prompt used for the preliminary tests on this doublet is as follows:

You are an expert linguist working on semantics and your aim is to check in which of the following sentences, the [word-class] '[competitor]' reflects one of the meanings that are provided below. Specifically, the steps to follow are: First, read the sentence containing the [word-class] '[competitor]'. Second, annotate the sentence using the following labels (a-c) and justify your answer: a) '[label1]' if the meaning expressed by the [word-class] '[competitor]' refers to '[Dictionary definition]'; b) '[label2]' if the meaning expressed by the [word-class] '[competitor]' refers to '[Dictionary definition]'; c) 'NONE' if the [word-class] '[competitor]' does not reflect any of the senses above or if the word is a [different word-class]. First, answer if it's '[label1]', '[label2]' or 'NONE'. Second, explain the result.

Table 1 exemplifies the results obtained for four concordance lines of *wrap*<sup>N</sup> and *wrapper*. A preliminary analysis seems to show that the results obtained from the automatic annotation are highly dependent on the definitions used as prompt, but that the degree of variability among automatic and human annotations, pending statistical validation, does not differ so much as expected. Further tests will be performed to explore the feasibility of using LLMs such as ChatGPT as a support tool for the annotation of large quantities of data for the study of competition.

Table 4. Results obtained for the semantic annotation of concordance lines of *wrap*<sup>N</sup> and *wrapper*

Concordance line	Human 1	Human 2	Collins	Cambridge	Merriam-Webster	OED
The following assembler language "wrap around" provides the minimum acceptable error handling capability. The " <b>wrap</b> around" recognises and acts on the following events	none	none	material	none	none	none
Unbuckling the fastenings, he removes the <b>wraps</b> and uncovered a wickerwork cage.	material	material	material	material	material	material
They are so described in the offer. 'They,' the <b>wrappers</b> , 'will help you to get smash hit recordings.'	none	none	material	material	material	material
Dinah had her maid comb her hair, put a <b>wrapper</b> round her, and bring her coffee; still sleepy and cross, but aware that her glorious breasts showed to advantage	clothing	clothing	clothing	clothing	clothing	clothing

## Bibliography

- Cambridge Dictionary. 2022. (<https://dictionary.cambridge.org>) (Accessed 21-03-2024).
- Collins Online Dictionary. 2022. (<https://www.collinsdictionary.com>) (Accessed 21-03-2024).
- Davies, Mark. 2004. *British National Corpus* (from Oxford University Press) (<https://www.english-corpora.org/bnc/>) (Accessed 21-03-2024).
- Fernández-Alcaina, Cristina. 2021. *The competition of word-formation processes in the derivational paradigm of verbs*. Berlin: Peter Lang.
- Grabowski, Łukas. 2023. Statistician, programmer, data scientist? Who is, or should be, a corpus linguist in the 2020s? *Jazykovedný časopis* 74(1), 52-59. <https://doi.org/10.2478/jazcas-2023-0023>
- Huyghe, Richard & Varvara, Rosella (eds.). 2023. Quantitative approaches to affix rivalry. Themed issue in *Word Structure* 16(1). <https://doi.org/10.3366/word.2023.0217>
- Lara-Clares, Cristina. 2023. *Morphological competition in present-day English nominalisations*. Granada: Universidad de Granada. <https://hdl.handle.net/10481/85059>
- MacWhinney, Brian & Malchukov, Andrej & Moravcsik, Edith (eds.). 2014. *Competing motivations in grammar and usage*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198709848.001.0001>
- Merriam-Webster. 2022. (<https://www.merriam-webster.com/dictionary/>) (Accessed 21-03-2024).
- OED Online (Oxford English Dictionary online). Oxford University Press. (<https://www.oed.com>) (Accessed 21-03-2024).
- OpenAI. 2024. ChatGPT-4 [Large language model]. <https://chat.openai.com/chat>
- Plag, Ingo. 1999. *Morphological productivity: Structural constraints on English derivation*. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110802863>
- Rainer, Franz & Gardani, Francesco & Dressler, Wolfgang U. & Luschützky, Hans Christian (eds.). 2019. *Competition in inflection and word-formation*. Cham: Springer.
- Santana-Lario, Juan & Valera, Salvador (eds.). 2017. *Competing patterns in English affixation*. Bern: Peter Lang. (<https://www.peterlang.com/document/1055047>) (Accessed 21-03-2024).

**Measuring the degree of transparency of English derivational suffixes**

Verginica Barbu Mititelu<sup>1</sup>, Gianina Iordăchioaia<sup>2</sup>, Svetlozara Leseva<sup>3</sup>, Ivelina Stoyanova<sup>3</sup>  
(<sup>1</sup>Romanian Academy Research Institute for Artificial Intelligence, <sup>2</sup>University of Graz,  
<sup>3</sup>Institute of the Bulgarian Language, Bulgarian Academy of Sciences)

The relation between a derived word and its base is established at two levels: form and meaning. The form of the derived word is the concatenation of the base with the affix, while its meaning is the result of semantic compositionality: the meaning of the affix added to that of the base. In this paper, we compare overt and zero suffixes that form verbs out of nouns (e.g. *-ise*, *-ify*, zero) and nouns out of verbs (e.g. *-er*, *-ion*, *-ment*, zero, etc.) in English. Since these suffixes are known to be highly polysemous, we evaluate their degree of transparency in terms of the semantic relation they establish with their base and the semantic classes of nouns/verbs (primes) they typically involve. We aim to find out which overt suffixes may be more transparent than others and whether overt suffixes are more transparent than zeros as often claimed in the literature (Plag 1999, Lieber 2004).

The data we build on is a large collection of pairs of base - derived words extracted from the Princeton WordNet (PWN) (Miller, 1995), which are annotated with a set of 14 morpho-semantic relations (such as Agent, Instrument, Event, etc.) (Fellbaum et al., 2009). Given that the pairs involve senses instead of words, it is possible to find different morphosemantic relations between different meanings of two derivationally-related words. We have first extracted the frequencies associating a suffix with each of these 14 morpho-semantic relations. For example, the suffix *-er* is associated with 12 out of the 14 morpho-semantic relations, with prevalent ones being Agent (in 76.3% of cases) and Instrument (in 12% of cases). Exploiting the hierarchical organization of the nouns and verbs in PWN trees corresponding to semantic classes/primes, we have also associated such semantic information to the pairs in the dataset. For instance, with nouns derived with *-er* and having the prime noun.person, the relation is unambiguously determined as Agent, while for those with the semantic prime noun.artifact, the relation is most likely Instrument.

In order to evaluate the degree of transparency of the suffixed words, we propose a machine learning method for automatically assigning (predicting) a morphosemantic relation between derivationally-related noun-verb pairs. The predictability of the morpho-semantic relation is used as a proxy for the semantic transparency of derivations (and the suffixes they involve). We use two datasets in the machine learning experiment: the training is based on data already available (Fellbaum et al., 2009) or obtained in previous work; in the testing phase we use pairs of verbs and nouns extracted from PWN that are related through a derivational relation, such as *-er*, *-ing*, but do not have an assigned morpho-semantic relation. The experiment results are validated manually and their evaluation will offer new insights into the degree of transparency of each of the analyzed suffixes.

## **Bibliography**

Fellbaum, Christiane, Anne Osherson, and Peter E. Clark. 2009. Putting Semantics into WordNet's "Morphosemantic" Links. In *Proceedings of the 3rd Language and Technology Conference*, Springer LNCS, vol 5603, 350-358.

Lieber, Rochelle. 2004. *Morphology and lexical semantics*. Cambridge: Cambridge University Press.

Miller, George A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM* Vol. 38, No. 11: 39-41.

Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation*. Berlin and New York: De Gruyter.

Using the *Démonette* derivational database for research in French morphophonology

Fiammetta Namer<sup>1</sup>, Nabil Hathout<sup>2</sup>, Stéphanie Lignon<sup>1</sup>  
 (<sup>1</sup>Université de Lorraine & ATILF, <sup>2</sup>CLLE, Université de Toulouse)

Version 2.0 of the *Démonette* French database [7], [8] includes a table of 222,118 descriptions of derivational relations between pairs of lexemes and a table of 388,306 lexeme descriptions. It will soon include the semantic description of 15,000 names [2]. Another project currently underway aims to complete the database with a morpho-phonological characterization of the pairs that make up the relations [9]. The inclusion of lexeme forms and their phonological transcriptions in the table of lexemes makes this large-scale description possible. This effort is a continuation of an experiment carried out on an earlier version of the database [6].

To describe the morpho-phonological relations between the forms of lexeme pairs, we need to identify their stems and phonological exponents (Figure 1). For example, for a W1/W2 pair like (a), we identify a stem /lav/ shared by both words and the affixes /de-/ and /abl/. This is an “ideal” configuration as both words have the same stem, and the exponents are directly concatenated to the stem to obtain the complete forms. This is not always the case, especially when the derivation involves a suppletive stem as in (b), a consonant variation (/k/s/ in c) or a vowel variation (/œ/ɔ/ in d), a nasal vowel (e), the realization of a latent consonant (/d/ in f), an epenthesis (/t/ in g), or other more complex alternations (h).

As part of this project, a program that compares the stems of 53,198 relations has been implemented, enabling us to (1) isolate the phonetic transcriptions of the affixes (columns **aff1** and **aff2** in Figure 1), (2) compare the stems resulting from this segmentation (columns **stem1** and **stem2**), calculating their editing distance and displaying similarities and differences in the form of patterns (column **Var.pattern**), where identical sequences are represented by the same uppercase variable.

Identifying phonological stems and exponents for all *Démonette* pairs is a challenge, given the size of the database and the diversity of the pairs it contains. This work also raises the question of a working definition of stems and exponents. In many cases, the boundaries of these sequences are difficult to delimit. For example, the exponent /-ɛl/ in *éternel* ‘eternal’ (i) can appear as /-yɛl/ in *rituel* ‘ritual’ (j), *individuel* ‘individual’ (k) or *conflictuel* ‘conflicting’ (l). However, the analysis of this variation is not uniform, because it depends among other things on etymological and lexical factors [4], [5]: in (j, l) /y/ is part of the affix variant, in (k) it is part of the stem. Case (l) raises the question of the delimitation of the suppletive stem ([1], [3]). If the answer is obvious in (b), the question arises in (l) and (m): in (l), because /kt/ is not a natural epenthesis in French, and in (m), because 3 sequences (X, Y, Z) are shared by 2 stems, but are not consecutive. Finally, another problem arises for pairs like in (k) and (n), in which there is an overlap between stem and exponent, resulting in a haplology to avoid the repetition of /y/ in (k) and of /iz/ in (n).

As we will show, the coding of morphophonological properties of relations leads us to reformulate theoretical questions, because of the quantity of data taken into account and the variety of annotations: in particular, what is the boundary between stem and exponent? between regular variation and suppletion? In turn, the answers to these questions are likely to call into question some of the current annotations in what we assumed to be a homogeneous data sets, which could lead to a revision of the constructions included in the database.



**Figure 1 - Excerpt from the table of morphophonological variations**

	W1	W2	stem1	stem2	aff1	aff2	Var. pattern
a	délaver wash out.v	lavable washable.a	/lav/	/lav/	/de-/	/-abl/	X/X
b	emprisonner imprison.v	incarcérer incarcerate.v	/prizɔn/	/karsɛr/	/ã-/	/ɛ̃-/	X/Y
c	clinique clinic.n	clinicien clinician.n	/klinik/	/klinis/	-	/-jɛ̃/	Xk/Xs
d	fleurette small flower.n	floral floral.a	/flœr/	/flɔr/	/-ɛt/	/-al/	XœC/XɔC
e	maison house.n	maisonnette small house.n	/mezɔ̃/	/mezɔn/	-	/-ɛt/	XṼ/XVn
f	lézard lizard.n	lézarder crack.v	/lezɑ̃/	/lezɑ̃d/	-	-	X/Xd
g	texto text message.n	textoter text message.v	/tɛksto/	/tɛkstot/	-	-	X/Xt
h	marteau hammer.n	marteler hammer.v	/mɑ̃to/	/mɑ̃tɔl/	-	-	Xo/Xəl
i	éternité eternity.n	éternel eternal.a	/etɛ̃n/	/etɛ̃n/	/-ite/	/-ɛl/	X/X
j	rite rite.n	rituel ritual.a	/ʁit/	/ʁit/	-	/-ye/	X/X
k	individu individual.n	individuel individual.a	/ẽdividy/	/ẽdividy/	-	/-ɛl/	X/X
l	conflit conflict.n	conflictuel conflincting.a	/kɔ̃fli/	/kɔ̃flikt/	-	/-ye/	X/Xkt
m	lettre letter.n	littéraire literary.a	/lɛtʁ/	/litɛʁ/	-	/-ɛʁ/	XɛYZ/XiYeZ
n	analyse analysis.n	analyste analyst.n	/analiz/	/anal/	-	/-ist/	Xiz/X

## Bibliography

- [1] Corbett, G. (2007). "Canonical typology, suppletion and possible words". *Language* 83, 8-42.
- [2] Huguin, M., Barque, L., Haas, P. & Tribout, D. (2023) "Typage sémantique des noms dans la ressource morphologique Démonette", *Lexique* 33, 41-56.
- [3] Maiden, M. (2004). "When lexemes become allomorphs - On the genesis of suppletion". *Folia Linguistica* 38 (3-4), 227-256.
- [4] Montermini, F. (2018). "Les affixes dérivationnels ont-ils des allomorphes ? Pour une modélisation de la variation des exposants dans une morphologie à contraintes". Bonami, O., Boyé, G., Dal, G., Giraud, H. & Namer F. (Eds.) *The lexeme in descriptive and theoretical morphology*, LSP.
- [5] Moskal B. (2015). "Limits on Allomorphy: A Case Study in Nominal Suppletion". *Linguistic Inquiry* 46 (2), 363-376.
- [6] Namer, F., Hathout, N. & Lignon, S. (2017). "Adding morpho-phonological features into a French morpho-semantic resource: the Démonette derivational database". Litta, E. & Passarotti, M. (Eds.) *First International DeriMo Workshop*. Milan, Italy, 49-61.
- [7] Namer, F. & Hathout, N. (2020). "ParaDis and Démonette – From Theory to Resources for Derivational Paradigms." *The Prague Bulletin of Mathematical Linguistics* 114, 5-33.
- [8] Namer, F., Hathout, N. *et alii*. (2023). "Démonette-2, a derivational database for French with broad lexical coverage and fine-grained morphological descriptions." *Lexique* 33, 6-40
- [9] Shrebko, D. & Herbin, Q. (2023). *Automatic detection of phonological variations between morphologically related words*. Tutored Project Final Report. University of Lorraine, IDMC.

**How the measures of derivational productivity depend on lemmatization quality  
(the case of Lithuanian deverbal nouns)**

Jurgis Pakerys (Vilnius University), Virginijus Dadurkevičius (Vytautas Magnus University),  
Agnė Navickaitė-Klišauskienė (Vilnius University)

To estimate derivational productivity in corpora, realized, expanding, and potential productivity measures are widely applied (Baayen 2009), (Gaeta & Ricca 2015: 844–849), (Dal & Namer 2016: 73–76). Our study aims to demonstrate that these measures may be heavily dependent on the quality of lemmatization, cf. (Dal et al. 2008: 1528). We believe that Lithuanian (Indo-European, Balto-Slavic), having elaborate inflectional morphology, suits well for this purpose.

In our study, we employed the Joint Corpus of Lithuanian of 1.3 billion words lemmatized with the help of *Hunspell*-based lemmatizer (Dadurkevičius 2017), (Dadurkevičius & Petrauskaitė 2020: 123–124), (Dadurkevičius 2020a), (Dadurkevičius 2020b). For the illustration of our results, we discuss only the data of deverbal agent nouns, while for the conference talk, we will add the data of action nominals, instrument, result, and location nouns. We also limited our study to type and hapax counts that represent realized and expanding productivity respectively.

Table 1 below (see next page) lists the data of three most productive Lithuanian deverbal agent noun suffixes *-toj-*, *-ėj-*, and *-ik-* which derive masculine and feminine formations distinguished by corresponding case and number inflections, e.g., *rašy-ti* ‘write’ → *rašy-toj-as* ‘writer’ (nom. sg. masc.), *rašy-toj-a* ‘idem’ (nom. sg. fem.), *kep-ti* ‘bake’ → *kep-ėj-as*, *-a* ‘baker’, *plauk-ti* ‘swim’ → *plauk-ik-as*, *-ė* ‘swimmer’.

The initial fully automatic lemmatization produced suspiciously low type and hapax counts, see the first two number columns in Table 1. We still reviewed these data to determine how many derivationally non-transparent formations would have to be excluded, see results of the manual clean-up in the 3<sup>rd</sup> and the 4<sup>th</sup> column.

In our next step, we proceeded to additional semi-manual lemmatization. The tokens were automatically filtered according to pattern SUFFIX + (all possible) ENDINGS, morphologically annotated, and grouped into lemmas. The resulting lemma lists were manually reviewed and derivationally non-transparent nouns were excluded. As can be seen from columns 5 through 8, additional lemmatization significantly increased type and hapax counts.

On a closer inspection, however, we noticed that due to homographic forms of masculine and feminine nouns, type and hapax counts were inflated to some degree. The reason for this was the lack of morphological disambiguation module in our lemmatizer: when a token that could be interpreted as a form of a masculine or a feminine noun was analyzed, both interpretations were accepted. As manual review of all tokens would be too time-consuming, we decided to disambiguate only the hapaxes. As seen in the last two columns, manual disambiguation mostly reduced hapax counts of feminine formations in *-toj-a* and *-ėj-a*, and this can be explained by the fact that masculine and feminine formations in *-toj-* and *-ėj-* have more homographic forms than the formations in *-ik-*.

We conclude that the quality of lemmatization may significantly influence the estimates of derivational productivity. Additional semi-automatic lemmatization leads to improvement, but one also needs to consider how the lemmatizer solves the cases of morphological ambiguity. We demonstrated that in the case of Lithuanian agent nouns, homographic forms significantly influenced the hapax counts. We were unable to estimate, however, to what extent the type counts and the total frequencies (needed for measuring the potential productivity) were affected.

Suffix	Initial automatic lemmatization				Additional semi-automatic lemmatization				
	Before manual review		After manual review		Before manual review		After manual review		
	V	V <sub>1</sub>	V	V <sub>1</sub>	V	V <sub>1</sub>	V	V <sub>1</sub> (before manual disambiguation of m/f)	V <sub>1</sub> (after manual disambiguation of m/f)
-toj-as (m)	632	11	627	9	3,305	822	2,456	532	464
-toj-a (f)	547	11	543	9	2,590	637	2,299	525	65
-ėj-as (m)	251	2	235	2	2,351	642	686	130	112
-ėj-a (f)	258	3	201	2	1,576	384	640	128	24
-ik-as (m)	293	2	88	0	911	189	254	65	61
-ik-ė (f)	159	6	50	4	857	274	92	27	20

**Table 1.** Derivational productivity of deverbal agent nouns in the Joint Corpus of Lithuanian (V – types, V<sub>1</sub> – hapaxes; m – masculine, f – feminine)

## Bibliography

Baayen, Harald R. 2009. Corpus linguistics in morphology: Morphological productivity. In Lüdeling, Anke & Kytö, Merja (eds.), *Corpus Linguistics: An International Handbook*, vol. 2, 899–919. Berlin, New York: Mouton de Gruyter. (doi:10.1515/9783110213881.2.899)

Dadurkevičius, Virginijus. 2017. Lietuvių kalbos morfologija atvirojo kodo „Hunspell“ platformoje [Lithuanian morphology in the Hunspell framework]. *Bendrinė kalba* 90. [1-17] (<https://journals.lki.lt/bendrinekalba/article/view/156>).

Dadurkevičius, Virginijus. 2020a. Wordlist of Lemmas from the Joint Corpus of Lithuanian. CLARIN-LT digital library in the Republic of Lithuania. (<http://hdl.handle.net/20.500.11821/41>)

Dadurkevičius, Virginijus. 2020b. Assessment Data of the Dictionary of Modern Lithuanian versus Joint Corpora. CLARIN-LT digital library in the Republic of Lithuania (<https://clarin.vdu.lt/xmlui/handle/20.500.11821/36>).

Dadurkevičius, Virginijus & Petrauskaitė, Rūta. 2020. Corpus-based methods for assessment of traditional dictionaries. In Utkā, Andrius & Vaičėnienė, Jurgita & Kovalevskaitė, Jolanta & Kalinauskaitė, Danguolė (eds.), *Human Language Technologies – The Baltic Perspective (Frontiers in Artificial Intelligence and Applications 328)*, 123–126. IOS Press. (doi:10.3233/FAIA200613)

Dal, Georgette & Fradin, Bernard & Grabar, Natalia & Namer, Fiammetta & Lignon, Stéphanie & Plancq, Clément & Zweigenbaum, Pierre & Yvon, François. 2008. Quelques préalables au calcul de la productivité des règles constructionnelles et premiers résultats. In Durand, Jacques & Habert, Benoît & Laks, Bernard (eds.), *Actes du premier Congrès mondial de linguistique française, Paris, 9–12 juillet 2008*, 1587–1599. Paris: Institut de Linguistique Française. (doi:10.1051/cmlf08184)

Dal, Georgette & Namer, Fiammetta. 2016. Productivity. In Hippisley, Andrew & Stump, Gregory (eds.), *The Cambridge Handbook of Morphology*, 70–90. Cambridge: Cambridge University Press. (doi:10.1017/9781139814720.004)

Gaeta, Livio & Ricca, Davide. 2015. Productivity. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-Formation: An International Handbook of the Languages of Europe*, vol. 2, 842–858. Berlin, Boston: De Gruyter. (doi:10.1515/9783110246278-003)

## Potential and actual word-formation lacunas in the system of Slavic Verbal Nouns

Olena Pchelintseva (Johannes Gutenberg University)

The focus of study is on Slavic verbal nouns (VNs) (such as Ukr. *čytannja*, Czech *čtení*– ‘reading’). Derivation of these nouns in Slavic has a common basis: the most productive derivational model «verb stem + –ije» is of Proto-Slavic origin. This is evidenced by its presence in all Slavic languages, including dialects. Despite the significant similarity of the developmental features of the word-formative systems of the Slavic languages, VNs constitute heterogeneous classes within one and the same language, but also in a comparative perspective. Studies of Slavic verbal nouns show that the number of VNs, as well as their retention of verb properties, increases considerably "from east to west" [1,4].

1) Russian VNs are formed only from every 6th verb (33,000 verbs stems correspond to 5,500 VNs), they practically lose aspectual and diathetical characteristics, and are semantically unpredictable; 2) Bulgarian, Serbian, Croatian and Ukrainian VNs are more productive and regular, for example, Ukrainian VNs are formed from every 4th verb, about 98% of units retain aspect, but don't retain diathetic properties; 3) 99% of Polish and Czech verbs have correlative VNs, nouns with *-nie (-tie)* have aspectual pairs, partially retain diathetical properties. See a typical example: one Russian VN which simultaneously relates to verbs of either aspect (e.g., *povtorenie* ‘repetition’) corresponds to two Ukrainian, Bulgarian, Serbian, Croatian VNs (Ukr. *povtorennja*/pfv. – *povtorjuvannja*/ipfv.; Bg. *povtarjane*/ipfv. – *povtorenie*/pfv. or ipfv.), to two or four Polish or Czech VNs (*powtarzanie*/ipfv. – *powtórzenie*/pfv. +/- reflexive marker *się*).

At the same time, despite the usual interpretation of Polish and Czech VNs as absolutely regular forms, some data indicate that this is not entirely true. For example, although the Czech language is characterised by the potentiality of VN (there are many forms in the texts that are not fixed in dictionaries; speakers form and use such VNs in suitable contexts quite freely, and their perception does not present any difficulties [2,6]), nevertheless, Czech VNs cannot be formed from causative, frequency and state verbs. In addition, they record the lexicalization of grammatical aspectual semantics and the violation of strict aspectual correlation of VN in speech [3].

Similarly, despite the usual treatment of Polish VNs as absolutely regular forms, VNs are not formed from some verbs. This may be influenced by the relational or modal nature of the verb meaning (for example, Po *mieć* ‘to have’ → *\*mienie*) or the impersonal meaning of the verb (Po *dnieć* ‘to dawn’ → *\*dnienie*) [5]. Much more interesting, however, is the other probable cause: some corpus data indicate that Polish VNs formed from verbs of certain aspectual classes exist only as potential (vocabulary), not actual units; they are absent in speech or lose some of their verb semantics. These aspectual classes of verbs, in particular, are semantically linked to an explicit temporal limit – delimitative, quantum limiting, inceptive (as e.g. Pol. *pobadać* ‘to study (some time)’ → *pobadanie?*). This may mean that the range of possible aspectual functions of the derived VN differs from the range of the main verb. Certain semantic components of the verb aspect are compatible with nominal derivations (duration, repetition, etc.), while other components are not compatible at all: apparently, “no-go” for VN in Polish is a temporal limitation of action.

Such semantic constraints need to be clarified. We also hypothesise that they may be peculiar to Czech VNs as well. We will test this hypothesis by means of a continuous comparison of verbs of certain aspectual classes (delimitative, quantum limiting, durative-limiting) and their derived VNs in Czech and Polish. We then compare the dictionary data with corpus data. We see five possible scenarios for the influence of supposed semantic restrictions on the word formation of VNs from verbs of certain classes: 1) absence of VNs in the dictionary, 2) presence in the dictionary, but the absence of such VNs in the corpus, 3) in the presence of such VNs in the corpora – reduction of verb meanings, and also 4) semantic drift towards substantiveness and 5) using forms VNs as a language game.

## Bibliography

- [1] Dickey, Stephen M. 2000. *Parameters of Slavic Aspect*. Stanford, CA.
- [2] Kocková, Jana. 2017. Substantiva mezi slovesem a jménem. Substantiva na -ní (-tí)/ -nie (-tie) v češtině a ruštině ve světle paralelního korpusu. *Časopis pro moderní filologii* 99/1, 55–64.
- [3] Panevová, Jarmila. 1980. *Formy a funkce ve stavbě české věty*. Praha.
- [4] Pčelinceva, Elena. 2016. *Ot glagola k imeni: aspektual'nost' v russkich, ukrainskich i pol'skich imenach dejstvija*. Sankt-Peterburg: Nauka
- [5] Puzynina, Jadwiga. 1969. *Nazwy czynności we współczesnym języku polskim (słowotwórstwo, semantyka, składnia)*. Warszawa: PWN SA.
- [6] Stehlíková, Lucie. 2010. *Morfosyntaktické vlastnosti deverbálních jmen na -ní/-tí*. Unpublished PhD dissertation, Masaryk University, Brno.

## Inflectional and derivational predictions in morphological families. A case study on Latin

Matteo Pellegrini, Federica Iurescia, Eleonora Litta, Marco Passarotti  
(CIRCSE Research Centre, Università Cattolica del Sacro Cuore, Milan)

In the last decades, a large amount of work has been conducted on the mutual predictability of different inflected forms of lexemes. While this issue can be tackled with different methodologies (see Stump & Finkel 2013 for a set-theoretic approach, and Sims 2020 for a graph-theoretic one), a strategy to investigate this issue in a quantitative and data-based fashion that has been increasingly gaining popularity is to use conditional entropy as a measure of the uncertainty in predicting an unknown form from a known one, performing such a measurement on all combinations of forms for lexemes documented in large inflected lexicons (cf. Ackerman et al. 2009 for a first proposal, Bonami & Boyé 2014 for suggesting a major revision of the procedure to take into account type frequency and avoid requiring a segmentation of exponents, Beniamine 2018 for implementing such procedure so as to make it more widely applicable to diverse languages).

In Bonami & Strnadová (2019), this methodology is applied to forms that are derivationally – rather than inflectionally – related. Such a move is motivated by the increasing interest for derivational paradigms (cf., among else, Štekauer 2014, Bauer 2019, Hathout and Namer 2019, Fernández-Domínguez et al. 2020, Ruz et al. 2022): if the notion of paradigm is extended from inflection to derivation, then also techniques used to investigate inflectional predictions can be extended to derivational ones.

Bonami & Strnadová (2019) only consider the citation form of derivationally related lexemes. However, as acknowledged by the authors themselves, a more thorough study would also check if there is variation in the predictability of derivatives from different inflected forms. This is what we explore in this work, taking as a case study deverbal action and agent nouns in Latin. These are especially promising because the two most used morphological processes involved usually require the addition of suffixes (namely, *-or* for agent nouns and *-io* for action nouns) to the stem that is displayed in verbal forms different than the citation form (namely, the first-person singular of the present indicative). However, other, less frequent processes (like *-ium* and *-ido* for action nouns) tend to display the present stem, on which the citation form is based, as shown in Table 1. This investigation is made easier by the availability of resources documenting both inflected forms of lexemes (PrinParLat, cf. Pellegrini 2023) and derivational relations (WFL, cf. Litta & Passarotti 2019), and by the ease with which such resources can be used in combination thanks to their inclusion into the LiLa Knowledge Base (Passarotti et al. 2020).

The paradigmatic system under investigation (see Table 1) consists on the one hand of the present infinitive, future indicative (in the third-person singular), perfect indicative (in the first-person singular) and perfect participle (in the neuter nominative singular) extracted from PrinParLat for verbs that are used as bases for the creation of agent nouns in *-or* and of action nouns in *-io*, *-ura*, *-ium* and *-ido* (other suffixes, like *-men* and *-mentum*, are excluded because they often denote something more specific, e.g. the means or place of an action) according to WFL, on the other hand of the nominative and genitive singular of such nouns, for a total of about 2,500 morphological families.

Preliminary results (reported in Table 2) confirm our expectation that the perfect participle is a more reliable predictor of agent nouns if compared to forms based on other stems, and the same holds – although to a lesser extent – for action nouns, despite the presence of suffixes normally attaching to the stem of the citation form, given their quantitative marginality. Additionally, it seems that forms based on the present stem are better predictors than those based on the perfect stem for action nouns, while for agent nouns there is no big difference. Among forms based on the present system, the infinitive is a better predictor than the future for action nouns, but the opposite holds for agent nouns, due to the different theme vowel oppositions across conjugations displayed by the two forms – although in both cases entropy values are close. These findings suggest that evaluating the predictability of derivational processes on the basis of the citation form alone might mean missing crucial information.

**Table 1 – Some partial morphological families from our dataset**

V_prs.act. inf	V_fut.act. ind.3.sg	V_prf.act. ind.1.sg	V_prf.pass. ptcp.nom.n.sg	ActionN_ nom.sg	ActionN_ gen.sg	AgentN_ nom.sg	AgentN_ gen.sg
<i>abdicare</i> 'indicate'	<i>abdicabit</i>	<i>abdicauī</i>	<i>abdicatum</i>	<i>abdicatio</i>	<i>abdicationis</i>	<i>abdicator</i>	<i>abdicatoris</i>
<i>profundere</i> 'pour'	<i>profundet</i>	<i>profudi</i>	<i>profusum</i>	<i>profusio</i>	<i>profusionis</i>	<i>profusor</i>	<i>profusoris</i>
<i>improperare</i> 'hasten into'	<i>improperabit</i>	<i>improperauī</i>	<i>improperatum</i>	<i>improperium</i>	<i>improperi</i>	<i>improperator</i>	<i>improperatoris</i>
<i>cupere</i> 'desire'	<i>cupiet</i>	<i>cupiui/cupii</i>	<i>cupitum</i>	<i>cupido</i>	<i>cupidinis</i>	<i>cupitor</i>	<i>cupitoris</i>

**Table 2 – Results: matrix of conditional entropy of predicting form in column given form in row**

	V_prs. act.inf	V_fut.act. ind.3.sg	V_prf.act. ind.1.sg	V_prf.pass. ptcp.nom.n. sg	ActionN_ nom.sg	ActionN_ gen.sg	AgentN_ nom.sg	AgentN_ gen.sg
V_prs.act.inf		0.38	0.36	0.21	0.28	0.29	0.26	0.26
V_fut.act.ind.3.sg	0.1		0.3	0.19	0.3	0.3	0.19	0.19
V_prf.act.ind.1.sg	0.29	0.3		0.32	0.36	0.37	0.27	0.26
V_prf.pass.ptcp.nom.n.sg	0.26	0.31	0.41		0.23	0.23	0.05	0.05
ActionN_nom.sg	0.22	0.26	0.39	0.04		0	0.07	0.07
ActionN_gen.sg	0.22	0.26	0.39	0.04	0		0.07	0.07
AgentN_nom.sg	0.24	0.27	0.39	0.05	0.19	0.19		0
AgentN_gen.sg	0.24	0.27	0.37	0.04	0.19	0.19	0	

## Bibliography

- Ackerman, F., Blevins, J.P. & Malouf, R. 2009. Parts and wholes: Implicative patterns in inflectional paradigms. In Blevins, J. & Blevins, J.P. *Analogy in grammar: Form and acquisition*. Oxford: OUP.
- Bauer, L. 2019. Notions of paradigm and their value in word-formation. *Word Structure* 12(2), 153-175.
- Beniamine, S. 2018. *Classifications flexionnelles. Étude quantitative des structures de paradigmes*. PhD Thesis. Université Sorbonne Paris Cité-Université Paris Diderot (Paris 7).
- Bonami, O. & Boyé, G. 2014. De formes en thèmes. In Villoing, F., Leroy, S. & David, S., *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*. Paris: Presses Universitaires de Paris-Ouest, 17-45.
- Bonami, O. & Strnadová, J. 2019. Paradigm structure and predictability in derivational morphology. *Morphology* 29(2), 167-197.
- Fernández-Domínguez, J., Bagasheva, A. & Lara-Clares, C. (eds.). 2020. *Paradigmatic Relations in Word Formation*. Leiden / Boston: Brill.
- Hathout, N. & Namer, F. 2019. Paradigms in word formation: what are we up to? *Morphology* 29(2), 153-165.
- Litta E. & Passarotti M. 2019. (When) inflection needs derivation: a word formation lexicon for Latin. In: Holmes, N., Ottink, M., Schrickx J. & Selig, M. (eds.) *Lemmata Linguistica Latina. Volume I: Words and Sounds*. Berlin / Boston: De Gruyter, 224-239.
- Passarotti, M., Mambrini, F., Franzini, G., Cecchini, F.M., Litta, E., Moretti, G., Ruffolo, P. & Sprugnoli, R. 2020. Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin. *Studi e Saggi Linguistici* LVIII(1), 177-212.
- Pellegrini, M. 2023. Flexemes in theory and in practice: Modelling overabundance in Latin verb paradigms, *Morphology*, 33(3), 361-395.
- Ruz, A.E., Lara-Clares, C. & Fernández-Alcaina, C. (eds.). 2022. *Paradigms in Word Formation*. Amsterdam / Philadelphia: John Benjamins.
- Sims, A.D. 2020. Inflectional networks: Graph-theoretic tools for inflectional typology. *Proceedings of the Society for Computation in Linguistics 2020*, 302-312.
- Štekauer, P. 2014. Derivational paradigms. In Lieber, R. & Štekauer, P. (eds.), *The Oxford Handbook of Derivational Morphology*. Oxford: OUP, 354-369.
- Stump, G. & Finkel R.A. 2013. *Morphological typology: From word to paradigm*. Cambridge: CUP.

**Exploring the diachrony of Italian word formation through the Google Books corpus:  
The case of NN compounds and their rival syntagmatic NPN counterparts**

Jan Radimský (University of South Bohemia in České Budějovice),  
Silvia Micheli, Pavel Štichauer (Charles University, Prague)

Rivalry between morphological strategies has recently been the subject of renewed interest. Nevertheless, while the prevailing focus has been on synchronic data, studies that have approached the subject from a diachronic perspective are fewer in number, partly due to the lesser availability of large diachronic corpora.

In this paper, we explore the diachronic development of two Italian constructions based on Google n-grams frequency lists (2020), which represents the most extensive diachronic linguistic data currently available. We will focus on two constructions that are currently regarded as competitors, as discussed in Masini (2019), namely, subordinate NN compounds (e.g., *sala stampa* ‘press room’) and Noun-Preposition-Noun phrasal lexemes (e.g., *sala da tè* ‘tea room’; see Masini 2012 for a comprehensive overview). In previous literature (see, e.g., Terreni 2005, Lombardi Vallauri 2008), the prevailing hypothesis suggests that Italian NNs are primarily derived from multi-word NPNs, where the preposition has been omitted, though, in some instances, analogy from similar NN forms in languages like English or French is also considered. We will consider both verbal-nexus NN compounds (e.g., *trattamento rifiuti* ‘waste treatment’), where the head is a deverbal noun, and grounding NN compounds, where the head is not verb-derived (e.g., *sala stampa* ‘press room’; see Rainer 2016, Baroni, Guevara, & Zamparelli 2009, Radimský 2023). Concerning NPNs, we examine phrasal nouns incorporating both a simple preposition (e.g., *centro di ricerca* ‘research center’) and a preposition combined with a definite article (e.g., *addetto alla stampa* ‘press officer’).

The status of verbal-nexus NNs and grounding NNs is expected to vary in their competition with NPNs. For verbal-nexus NNs, a NPN paraphrasis with the preposition *di* should always be available and prevalent in actual data (e.g., *trattamento ‘di’ rifiuti*, meaning ‘waste treatment’). In the case of grounding NNs, a wider variety of prepositions is expected to appear in NPNs, and the NPN counterpart may be absent in the data (either entirely or over significant diachronic periods), or at times, even deemed ungrammatical (see Baroni, Guevara, & Pirrelli 2009).

Our goal is to examine a sample of these subordinate NN compounds and their potential or existing NPN equivalents. This selection comprises roughly 3,000 NNs, thoroughly curated through a combination of automated and manual procedures using Google n-gram frequency lists covering the period from 1850 to the present day. We intend to describe in detail the process of the extraction of the data focussing on some major obstacles in handling such amounts of data (noise elimination, false positives, etc.), and we will concentrate on the construction of the final sample, in which the principal step was to test, for each identified NN, the actual presence of a potential NPN counterpart within the same dataset and analyze the diachronic relative frequency curves of these competing constructions.

The outcomes of our analysis will show that only a subset of NNs display a similar diachronic frequency distribution compared to their NPN counterparts. We will strive to not only identify the various scenarios of diachronic competition between NNs and NPNs but also specific “niches” (as described by Aronoff in 2019) in which NNs gradually tend to dominate over NPNs. Our findings also challenge the hypothesis that subordinate NN compounds can be regarded as mere abbreviations of full NPN constructions, primarily motivated by the need for conciseness, potentially influenced by distinct linguistic registers.

On the theoretical side, by using the tools provided by Diachronic Construction Grammar (Hilpert 2021) and Relational Morphology (Jackendoff & Audring 2020), namely certain key hypotheses, such as *structural intersection*, *coverage*, *upward-strengthening* or *statistical pre-emption*, we endeavour to demonstrate the gradual development of lower-level constructions, lacking significant NPN variants, plays a pivotal role in shaping the overarching pattern of Italian subordinate NN compounds.



## Bibliography

- Aronoff, M. (2019). Competitors and Alternants in Linguistic Morphology. In F. Rainer et al., F. Gardani, W. U. Dressler, & Luschützky, H. C. (Eds.), *Competition in Inflection and Word-Formation* (Vol. 5, pp. 39–66).
- Baroni, M., Guevara, E., & Pirrelli, V. (2009). Sulla tipologia dei composti N+N in italiano: Principi categoriali ed evidenza distribuzionale a confronto. In G. Ferrari, R. Benatti, & M. Mosca (Eds.), *Linguistica e modelli tecnologici di ricerca*. (pp. 73–96). Roma: Bulzoni Editore.
- Baroni, M., Guevara, E., & Zamparelli, R. (2009). The dual nature of Deverbal Nominal Constructions: Evidence from acceptability ratings and corpus analysis. *Corpus Linguistics and Linguistic Theory* 5(1), 27-60.
- Hilpert, M. (2021). *Ten lectures on diachronic construction grammar*. Leiden: Brill.
- Jackendoff, R., & Audring, J. (2020). *The texture of the lexicon: relational morphology and the parallel architecture*. Oxford: Oxford University Press.
- Lombardi Vallauri, E. (2008). Composti intitolativi in italiano: Un'oscillazione. In *Prospettive nello studio del lessico italiano*. Firenze: Firenze University Press.
- Masini, F. (2012). *Parole sintagmatiche in italiano*. Roma: Caissa.
- Masini, F. (2019). Competition between morphological words and multiword expressions. Competition in inflection and word-formation. In F. Rainer et al. (Eds.), *Competition in Inflection and Word-Formation* (pp. 281–305).
- Radimský, J. (2023). Where did the Italian Verbal-Nexus N+N compounds come from? In J. Audring et al., *Mediterranean Morphology Meetings – Comparing typologies*, 13, Patras, University of Patras (pp. 71–82).
- Rainer, F. (2016). Italian. In P. O. Müller, I. Ohnheiser, & S. Olsen (Eds.), *Word-formation. An International Handbook of the Languages of Europe* (pp. 2712-2731). Berlin/Boston: Walter de Gruyter.
- Terreni, R. (2005). Composti N + N e sintassi: I tipi economici lista nozze e notizia-curiosità. In A. M. Thornton, & M. Grossmann (Eds.), *La formazione delle parole*. Roma: Bulzoni Editore.

## Creating and exploiting a lexical database of deverbal nouns in French

Rossella Varvara, Justine Salvadori, Richard Huyghe (University of Fribourg)

Interest in the semantics of derivational processes has increased in recent years (see e.g. Bauer et al. 2015, Lieber 2016, Kotowski and Plag 2023). Common topics of investigation have been, for example, the ambiguity of nominalizations (e.g. Melloni 2011, Plag et al. 2018), the transfer of properties from the base word to the noun, e.g. with respect to lexical aspect (Haas et al. 2008, Fábregas & Marín 2012), or semantic differences between competing morphological processes (e.g. Naccarato 2019, Nagano 2023). However, attempts of quantifying the diffusion of these phenomena are missing or are very rare, especially taking into account the differences between specific affixes. Moreover, the data used to analyze these phenomena are limited in the number of lexemes or affixes considered, the annotation performed is specific to the analysis conducted and does not allow comparison or generalizations in a wider perspective, or the properties investigated are obtained automatically (and thus not totally accurate).

In this contribution, we first present a large-scale semantic annotation of 5,225 French deverbal nouns that has been specifically designed to allow: the comparison among lexemes and morphological processes, as well as between base and derived terms; a detailed treatment of polysemy with the annotation of word senses; a generalization based on a considerable amount of data. The annotation has covered (i) the semantic type of deverbal nouns and their different senses; (ii) the lexical aspect and argument structure of base verbs and deverbal nouns; (iii) the semantic roles assigned by bases and derivatives.

In this presentation, we will first provide details on the sample selection, the annotation scheme, and on the inter-annotator agreement. Secondly, we will exploit the database to answer three research questions on central topics in the semantics of derivation:

- Is there a correlation between the number of semantic types realized by a nominalizing suffix (i.e. affix polyfunctionality) and the number of senses of the derivatives they form (i.e. lexical ambiguity)?
- Do eventive nominalizations preserve the lexical aspect and the argument structure of their base verbs?
- How are semantic types organized in morphological families, and which semantic paradigms can be observed for the different base verbs?

With regard to the first question, we found a moderate correlation between lexical ambiguity and affix polyfunctionality (Kendall's  $\tau = .41$ ,  $p = .004$ ). This result confirms that, even if they are related, affix polyfunctionality and lexical ambiguity need to be distinguished in theoretical and empirical studies on derivation. As far as aspectual preservation is concerned, our data show that eventive nominalizations inherit the properties of the base verbs (durativity, dynamicity, telicity, and post-phase) in most cases (2,449 nouns out of 2,645). At least one argument from the base verb is inherited in most cases (93%), usually with the semantic role assigned to the subject of the verb. This contrasts with non-eventive nominalizations, for which preservation is observed in 14% of the cases only. Lastly, while inspecting paradigmatic organization of morphological families, we observed that most verbs have only one derivative (which is most frequently an eventive nominalization). The second most common pattern involves an event and an agent noun (for 276 verbs), followed by a paradigm of three nouns (event, agent, instrument, for 89 verbs). Four derivatives (denoting events, agents, instruments, and results) are observed for 53 verbs, whereas other families of at least four members are attested for a smaller number of verbs (less than 10). Further details and a discussion of the theoretical implications of these findings will be given in the presentation.

## Bibliography

- Bauer, L., Körtvélyessy, L., Štekauer, P. (Eds.) (2015). *Semantics of complex words*. Studies in Morphology, vol 3. Springer, Cham.
- Fábregas, A. & Rafael M. (2012). The role of Aktionsart in deverbal nouns: State nominalizations across languages. *Journal of Linguistics* 48, 35-70.
- Haas, P., Huyghe, R., & Marín, R. (2008). Du verbe au nom : calques et décalages aspectuels. In *Congrès Mondial de Linguistique Française 2008*, J. Durand, B. Habert, B. Laks (éds), Paris : Institut de Linguistique Française, 2051-2065.
- Kotowski, S., & Plag, I. (Eds.). (2023). *The semantics of derivational morphology: Theory, methods, evidence* (Vol. 586). Walter de Gruyter GmbH & Co KG.
- Lieber, R. (2016). *English nouns: The ecology of nominalization*. Cambridge: Cambridge University Press.
- Melloni, C. (2011). *Event and result nominals*. Bern: Peter Lang.
- Naccarato, Chiara. (2019). Agentive (para)synthetic compounds in Russian: a quantitative study of rival constructions. *Morphology* 29(1). 1–30.
- Nagano, Akiko. (2023). Affixal rivalry and its purely semantic resolution among English derived adjectives. *Journal of Linguistics* 59(3), 499-530.
- Plag, I., Andreou, M., & Kawaletz, L. (2018). A frame-semantic approach to polysemy in affixation. In O. Bonami, G. Boyé, G. Dal, H. Giraud, & F. Namer (Eds.), *The lexeme in descriptive and theoretical morphology* (pp. 467–486). Berlin: Language Science Press.

16

*Polar (yes/no) questions:  
form, meaning and more*

organizers

Maria Razguliaeva

Maria Onoeva

(ENG)

# Workshop description

Questions pose a challenge for the traditional truth-conditional semantics because their meaning cannot be reduced to the truth or falsity of a proposition. Among the prominent approaches to this problem are the analyses of Hamblin (1973) and Groenendijk and Stokhof (1984) according to which a question denotes a set of propositions corresponding to its possible or true answers. In the case of polar questions (PQs), the sets contain two answers corresponding to ‘yes’ and ‘no’. However, these approaches alone do not seem to capture some question functions in language and discourse.

For instance, the impact of negation on PQs is beyond the scope of such approaches since the ultimate semantic meaning would be independent of the presence or absence of negation (Groenendijk and Stokhof 1984). Positive (‘Is it raining?’) and negative PQs (‘Is it not raining?’ / ‘Isn’t it raining?’) differ in the contexts where they can be used (Sudo 2013, Kopecký 2010, AnderBois 2019) and in the possible reactions to them (Farkas and Bruce 2009). Negation itself can have a different effect depending on the nature of the negation operator and its place in the derivational structure (Ladd 1981, Romero and Han 2004, Abels 2005).

Many different strategies of indicating PQs are observed across languages, ranging from question particles in Mandarin and verb morphology in Korean over interrogative word order in German to intonation in Modern Standard Arabic (Dryer 2013). Along with the neutral expressions of question meaning, pragmatically marked variants of PQs have also been studied. These can contain certain particles like *vajon* in Hungarian (Gärtner and Gyuris 2023), *-no* and *-desho* in Japanese (Sudo 2013), and *razve* in Russian (Repp 2013, Korotkova 2023), as well as diverge in word order (Gunlogson 2002, Staňková 2023) or prosodic patterns (e.g., Grice and Savino 2003, Goodhue 2022, Esipova and Romero 2023). Exploring the conditions in which these types of questions can occur has led to multiple notions of an additional layer of meaning that captures the discourse functions of an utterance (e.g., Gutzmann 2015, Ciardelli et al. 2018).

This panel aims to bring together researchers who work on forms, meanings, and functions of PQs in their intra- and cross-linguistic variability. We invite a diverse range of submissions that draw upon theoretical observations, experimental results, corpus research, and other relevant perspectives. The goal is to establish a productive dialog between different approaches and gain new insights into the nature of PQs. It is planned to address the following research questions (the list is not exhaustive):

- How do forms of PQs map to various facets of their meanings across languages?
- Which contexts can be considered natural for certain PQ types?
- How can the meaning of PQs be formalized and what predictions emerge from certain formalizations?
- How are PQs processed by speakers, as well as during language acquisition and/or learning?

## References

- Abels, K. (2005). "Expletive" negation in Russian: A conspiracy theory. *Journal of Slavic Linguistics* 1(13), 5–74.
- AnderBois, S. (2019). Negation, alternatives, and negative polar questions in American English. In K. von Heusinger, E. Onea, & M. Zimmerman (Eds.), *Questions in Discourse – volume 1: Semantics* (pp. 118–171). CRISPI.
- Ciardelli, I., Groenendijk, J., & Roelofsen, F. (2018). *Inquisitive semantics*. Oxford University Press eBooks. <https://doi.org/10.1093/oso/9780198814788.001.0001>
- Dryer, M. S. (2013). Polar Questions (v2020.3). In M. S. Dryer & M. Haspelmath (eds.), *The World Atlas of Language Structures Online*. Zenodo. doi:10.5281/zenodo.7385533.
- Farkas, D. F., & Bruce, K. B. (2009). On reacting to assertions and polar questions. *Journal of Semantics*, 27(1), 81–118. <https://doi.org/10.1093/jos/ffp010>
- Esipova, M. & Romero, M. (2023). Prejacent truth in rhetorical questions. Ms., University of Konstanz. In prep.
- Gärtner, H. M. & Gyuris B. (2023). What Does Vajon Contribute? In Onoeva M., Staňková A., & Šimík R. (eds.): *Proceedings of Sinn und Bedeutung 27* (pp. 228–239). Charles University.
- Goodhue, D. (2022). Isn't there more than one way to bias a polar question? *Natural Language Semantics*, 30(4), 379–413. Retrieved from <https://link.springer.com/article/10.1007/s11050-022-09198-2>
- Grice, M., & Savino, M. (2003). Map Tasks in Italian: Asking Questions about Given, Accessible and New Information. *Catalan Journal of Linguistics*, 2, 153–180. <https://doi.org/10.5565/rev/catjl.48>
- Groenendijk, J. A. G., & Stokhof, M. J. B. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. Univ. Amsterdam.
- Gunlogson, C. (2002). Declarative Questions. *Semantics and Linguistic Theory*, 12, 124–143. doi:10.3765/salt.v12i0.2860
- Gutzmann, D. (2015). *Use-Conditional meaning: Studies in Multidimensional Semantics*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198723820.001.0001>
- Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language*, 10(1), 41–53. Retrieved from <https://www.jstor.org/stable/25000703>
- Kopecký, J. (2010). Komunikační funkce záporných otázek zjišťovacích. *Naše řeč*, 93(1), 1–15.
- Korotkova, N. (2023). Conversational dynamics of Russian questions with *razve*. In Onoeva M., Staňková A., & Šimík R. (eds.): *Proceedings of Sinn und Bedeutung 27* (pp. 328–346). Charles University.
- Ladd, D. R. (1981). A First Look at the Semantics and Pragmatics of Negative Questions and Tag Questions. In R. Hendrick, C. Masek, & M. F. Miller (Eds.), *Papers from the Seventeenth Regional Meeting of the Chicago Linguistic Society* (pp. 164–171). Retrieved from <http://www.lel.ed.ac.uk/~bob/PAPERS/CLS1981.pdf>
- Repp, S. (2013). Common ground management: Modal particles, illocutionary negation and *verum*. In D. Gutzmann & H.-M. Gärtner (Eds.), *Beyond expressives: Explorations in use-conditional meaning* (pp. 231–274). BRILL eBooks. [https://doi.org/10.1163/9789004183988\\_008](https://doi.org/10.1163/9789004183988_008)
- Romero, M., & Han, C.-H. (2004). On Negative Yes/No Questions. *Linguistics and Philosophy*, 27(5), 609–658. doi:10.1023/b:ling.0000033850.15705.94
- Šimík, R. (to appear). Polar question semantics and bias: Lessons from Czech/Slavic. In B. Gehrke & R. Šimík (Eds.), *Topics in the semantics of Slavic languages*. Language Science Press.
- Staňková, A. (2023). *The expression of speaker's bias in Czech polar questions*. Master Thesis, Charles University.
- Sudo, Y. (2013). Biased polar questions in English and Japanese. *Current Research in the Semantics/Pragmatics Interface*, 28, 275–295. doi:10.1163/9789004183988\_009

# Table of Contents

*Marking speaker authority in responding acts: A study of Hungarian*

Beáta Gyuris (HUN-REN Hungarian Research Centre for Linguistics & ELTE Eötvös Loránd University, Budapest)

*Partition by exhaustification and polar questions*

Tue Trinh (Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin)

*Bias in Polish czy- and bare polar questions: a corpus-based study*

Roland Meyer (Humboldt-Universität zu Berlin)

*Negative polar questions in Hungarian Bible translations*

Cecília Sarolta Molnár (HUN-REN Hungarian Research Centre for Linguistics)

*Cross-linguistic variation in the meaning of high negation in polar questions*

Radek Šimík, Anna Staňková, Michaela Chodounská (Charles University)

*Word order and information structure in Czech polar questions*

Anna Staňková (Charles University)

*Negation is present: Subjunctive questions in Serbian*

Neda Todorović (Reed College)

*Preposed negation vs. non-Preposed negation in Greek polar questions*

Evripidis Tsiakmakis (Universitat Pompeu Fabra)

## Marking speaker authority in responding acts: A study of Hungarian

Beáta Gyuris

(HUN-REN Hungarian Research Centre for Linguistics &amp; ELTE Eötvös Loránd University, Budapest)

The paper discusses the interpretation of the particle *tényleg* 'really'/'indeed' in Hungarian, considered analogous to those of *csakugyan* and *valóban* (Egedi, 2009), the former's stylistic variants. *Tényleg* can appear syntactically and prosodically integrated or non-integrated in the sentence, cf. (1B') for the latter. Integrated *tényleg* appears in declaratives used as responses to assertions or to polar questions and in (responding) polar interrogatives, non-integrated *tényleg* can also appear in *wh*-questions.

We propose that *tényleg* only contributes to the use-conditions (Gutzmann 2015), and its integrated use is felicitous in a sentence encoding a speech act  $\mathfrak{U}_{S_1, S_2}$  provided that *AnaC* and *AuthC* hold:

- (1) *Anaphoricity Condition (AnaC)*:  $\mathfrak{U}_{S_1, S_2}$  must be preceded by an utterance in *c* where  $S_i$  makes an (evidenced or categorial, cf. (Farkas, 2023)) commitment to the sentence radical *p*.
- (2) *Authority Condition (AuthC)*: The *epistemic authority* of the perspective center of  $\mathfrak{U}$  regarding the truth of *p* is higher than that of any  $S_i$ .

We assume that epistemic authority is either status-based, (3B), or source-based (i.e., having more information), as in (4B) (cf. Enfield 2011, Heritage & Raymond 2005), and reference to the perspective center (Bylinina, McCready, & Sudo, 2014) accounts for the fact that *tényleg* undergoes *interrogative flip*. It follows from these assumptions that *tényleg* cannot mark a contradiction to a previous utterance, or answer a(n indirect) question, cf. (5), (6) and (8), but it is felicitous in an assertion that aims to decide a debate, cf. (7). In polar interrogatives, (9), it indicates that confirmation is asked for, cf. Kiefer (1988).

- |                                                                                                                                                                                                                                                                                                                                                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>(3) A: Esett éjjel az eső.<br/>fell at.night the rain<br/>'It rained last night.'</p> <p>B: (B heard the sound of rain last night,<br/>so he knows that it rained.)<br/># <i>Tényleg</i> esett.<br/>TÉNYLEG fell<br/>'It rained indeed.' (OK if A is a child!)</p> <p>B': (B did not hear anything, but looks<br/>out and sees that the ground is wet.)<br/><i>Tényleg</i>(, esett).<br/>'It rained indeed.'</p> | <p>(4) A: Pali lehet, hogy külföldre ment.<br/>Pali maybe that abroad went<br/>'Maybe Pali went abroad.'</p> <p>B: <i>Tényleg</i> külföldre ment.<br/>'He went abroad indeed.'</p> <p>(5) A: Pali nem ment külföldre.<br/>Pali not went abroad<br/>'Pali didn't go abroad.'</p> <p>B: (#<i>Tényleg</i>) külföldre ment.</p> <p>(6) A: Pali külföldre ment?<br/>'Did Pali go abroad?'</p> <p>B: (#<i>Tényleg</i>) külföldre ment.</p> <p>(8) A: Kérdés, hogy P. külföldre ment-e.<br/>question that P. abroad went-Q<br/>'It's a question whether P. went abroad.'</p> <p>B: (#<i>Tényleg</i>) külföldre ment.</p> |
| <p>(7) A: Pali külföldre ment.<br/>'Pali went abroad.'</p> <p>B: Nem ment.<br/>'He didn't.'</p> <p>C: <i>Tényleg</i> külföldre ment.</p>                                                                                                                                                                                                                                                                            | <p>(9) A: (Állítólag) Pali külföldre ment.<br/>allegedly Pali abroad went<br/>'(Allegedly,) Pali went abroad.'</p> <p>B: <i>Tényleg</i> (külföldre ment)?<br/>'Did he really go abroad?'</p>                                                                                                                                                                                                                                                                                                                                                                                                                      |

Besides formalizing the contribution of *tényleg* to assertions and questions in the Table model (Farkas & Bruce 2010), the talk will compare its distribution and interpretation to that of VERUM (cf. Gutzmann et al. 2020, Romero & Han 2004), Japanese *yo* (Northrup 2014), and Mandarin *dique* and *zhende* (Yuan & Hara 2019). The account will be extended to non-integrated uses of *tényleg* by transferring *AnaC* and *AuthC* to the speech act level.



## Bibliography

- Bylina, L., McCready, E., & Sudo, Y. (2014). *The landscape of perspective shifting*. Paper presented at the Pronouns in Embedded Contexts at the Syntax-Semantics Interface, Tübingen.
- Egedi, B. (2009). Adverbial (dis)ambiguities. Syntactic and prosodic features of ambiguous predicational adverbs. In K. É. Kiss (ed.), *Adverbs and adverbial adjuncts at the interfaces* (pp. 103-132). Berlin: Mouton de Gruyter.
- Enfield, N. J. (2011). Sources of asymmetry in human interaction: enchrony, status, knowledge and agency. In T. Stivers, L. Mondada, & J. Steensig (eds.), *The morality of knowledge in conversation* (pp. 285-312). Cambridge: Cambridge University Press.
- Farkas, D. (2023). Bias and anti-bias: Two case studies from Hungarian. *Journal of Uralic Linguistics*, 2, 96-126.
- Farkas, D., & Bruce, K. (2010). On reacting to assertions and polar questions. *Journal of Semantics*, 27, 81-118.
- Gutzmann, D. (2015). *Use-conditional meaning. Studies in multidimensional semantics*. Oxford: Oxford University Press.
- Gutzmann, D., Hartmann, K., Matthewson, L., & evidence., C.-l. (2020). Verum focus is verum, not focus. *Cross-linguistic evidence.*, 5(1), 1-48.
- Heritage, J., & Raymond, G. (2005). The terms of agreement: Indexing epistemic authority and subordination in talk-in-interaction. *Social Psychology Quarterly*, 68(1), 15-38.
- Kiefer, F. (1988). Modal particles as discourse markers in questions. *Acta Linguistica Hungarica*, 38, 107-125.
- Northrup, O. B. (2014). *Grounds for Commitment*. PhD dissertation, UC Santa Cruz.
- Romero, M., & Han, C.-H. (2004). On negative yes/no questions. *Linguistics and Philosophy*, 27, 609-658.
- Yuan, M., & Hara, Y. (2019). Guiding assertions and questions in discourse: Mandarin *dique* and *zhende*. *Natural Language and Linguistic Theory*, 37, 1545–1583.

**Partition by exhaustification and polar questions**

Tue Trinh (Leibniz-Zentrum Allgemeine Sprachwissenschaft, Berlin)

Fox (2019, 2020) proposes a condition on questions which he calls "Partition by Exhaustification" (PbE): A question is only felicitous if its answers, once exhaustified, partitions the context set. The case for PbE was made using constituent questions. Polar questions were not discussed. In this talk, I argue that PbE derives a number of facts about the interpretation of polar questions which have been taken for granted but which, given familiar assumptions about the syntax of polar questions, should be puzzling. I then show that given PbE, many cases of "ill-formedness" with respect to polar questions in Vietnamese can be given the same analysis as cases of "unavailable readings" with respect to polar questions in English.

**Bibliography**

Fox, Danny. 2019. Partition by exhaustification: Comments on Dayal 1996. *Proceedings of Sinn und Bedeutung* 22:403–434.

Fox, Danny. 2020. Partition by exhaustification: Towards a solution to Gentile and Scharwarz's puzzle. Manuscript, MIT.

Bias in Polish *czy*- and bare polar questions: a corpus-based study

Roland Meyer (Humboldt-Universität zu Berlin)

Slavic languages mark Polar Questions (PQs) in many different ways, including verb movement, intonation and interrogative particles (see e.g. Logvinova 2022 for a recent overview). Polish (Pol) makes use of the marker *czy* (cf. the Ukrainian cognate *čy* and Belarusian *ci*), which is the obligatory interrogative complementizer in embedded interrogatives, but occurs only optionally in main clause PQs. The status of *czy* in main clause PQs calls for clarification in several ways: First, it seems to be related to register and occur rarely in the spoken language, according to native speakers' intuitions; Logvinova (2022:30) reports that only about half of all the PQs in her corpus sample (48%) contained *czy*. Second, its semantic or pragmatic function, especially its bias profile in comparison to *czy*-less, "bare" PQs, remains to be clarified. Šimík (*to appear*:25ff) found that *czy*-PQs, but not bare PQs, are acceptable in quiz contexts, which count as maximally unbiased; thus, it seems that *czy*-PQs are actually the most neutral, unbiased way of asking a question in Pol. Third, *czy*-PQs must be viewed in the context of other PQ structures in Pol, and in comparison with the PQ marking devices of closely related (West Slavic) languages. The present paper addresses these issues, departing from a detailed corpus-based investigation of *czy*- and bare PQs in Pol.

Our main empirical findings are the following: (i) Verb fronting is not used as a question-marking device in Pol, other than in Cz (cf. also Šimík *to appear*; Staňková 2023). If we exclude a large amount of uninformative evidence involving, e.g., null subjects or postverbal forms of address like *pan(i)* 'you-honorific' or *ojciec* 'father (priest)', we are left with virtually only presentational sentences and postverbal subjects under minimal focus – i.e., Verb (V)–Subject (S) word order in PQs occurs under similar information-structural conditions as in declarative clauses. In our sample of PQs with pre- vs. postverbal names in nominative (to avoid tagging errors with common nouns) drawn from the 300 Mio. tokens balanced subcorpus of the Polish National Corpus (Przepiórkowski et al. 2012), SV outnumbered VS by 733:84 in *czy*-PQs, and by 231:60 in bare PQs. (ii) *Czy* is equally compatible with outer and inner negation, as shown by its cooccurrence with positive polarity items (PPIs) (outer negation) and negative concord items (NCIs) (inner negation). In the NKJP 300 Mio. sample, we found 222 negated *czy*-PQs with PPI *jakiś* 'some' and 135 with NCI *żaden* 'any', cf. (1)-(2). (iii) *Czy*-PQs with outer negation (as evidenced by presence of a PPI) mostly occur in contexts which suggest a positive epistemic bias towards the prejacent (cf. Sudo 2013 for terminology); i.e., while asking whether the proposition *p* does not hold, the speaker concomitantly conveys a background expectation that *p* holds. What actual evidence in the situation suggests about the validity of *p*, does not seem to matter. (iv) *Czy*-PQs with inner negation are very different: Here the situational evidence strongly suggests that non-*p* holds, in line with observations about inner negation-PQs in other languages (cf. Repp & Geist 2022). This creates a strong flavour of incredulity, which is often made explicit by adverbs like *naprawdę* 'really', cf. (3). (v) A preliminary investigation of bare PQs with inner negation in the corpus sample indicates that the conflict between epistemic bias and evidence to the negative is absent or at least much weaker here.

Note that the conflict of biases found in (iv), but not in (v), can only obtain if *czy*-PQs differ from bare PQs semantically or pragmatically, rather than only by register. We interpret this as an argument for the distinction put forward in Šimík (*to appear*) in the spirit of Biezma & Rawlins (2012), that *czy*-PQs denote a balanced set of alternatives  $\{p, \text{non-}p\}$ , whereas bare PQs may denote only a single proposition, putting the alternatives at issue. This idea will be further spelled out to derive the difference in (iv)/(v), and complemented with evidence from a recent acceptability study.

## Examples

- (1) *Czy nie powinniśmy mieć jakiegoś podarunku dla króla?*  
PTCL NEG obliged.1PL have.INF some.GEN.SG present.GEN.SG for king.GEN  
'Shouldn't we have some present for the king?'
- (2) *Czy Ciebie nie trapią żadne wątpliwości?*  
PTCL you.ACC NEG haunt.3PL any.NOM.PL doubts.NOM.PL  
'Are you not haunted by any doubts?'
- (3) [*To znaczy, że biznes – w swoim własnym długofalowym interesie – również odczuwa potrzebę etyki i sprawnej sprawiedliwości. Tylko co z młodymi ludźmi?* – 'This means that business, in its own long-term interest, equally feels the need for ethics and for real justice. But what about young people?']  
*Czy naprawdę nie wierzą już nikomu i liczą tylko na własną, indywidualną zaradność?*  
PTCL really NEG believe.3PL already nobody.DAT and count.3PL only on own.ACC individual.ACC competence.ACC  
'Do they really not believe anybody any more, and count only on their own individual competence?'

## Bibliography

- Biezma, M. & Rawlins, K. (2012). Responding to alternative and polar questions. *Linguistics and Philosophy* 35(5), 361–406.
- Logvinova, N. (2022). *Aspects of the syntax of root and complement polar questions in Slavic*. MA thesis, Higher School of Economy, Moscow.
- Przepiórkowski, A., Bańko, M., Górski, R. & Lewandowska-Tomaszczyk, B. (2012). *Narodowy Korpus Języka Polskiego*. Wydawnictwo Naukowe PWN, Warsaw.
- Repp, S. & Geist, L. (2022). *Negative polar questions in Russian: Question bias and question concern*. ms., University of Cologne and University of Stuttgart.
- Staňková, A. (2023). *The expression of speaker's bias in Czech polar questions*. MA thesis, Charles University, Prague.
- Sudo, Y. (2013). Biased polar questions in English and Japanese. In: Gutzmann, D. & H.-M. Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning*, 275–295. Brill, Leiden.
- Šimík, Radek (to appear). *Polar question semantics and bias: Lessons from Slavic/Czech*. ms., Charles University, Prague.

## Negative polar questions in Hungarian Bible translations

Cecília Sarolta Molnár (HUN-REN Hungarian Research Centre for Linguistics)

Interrogatives realizing simple polar questions can be formulated either in positive (1a) or in negative form (1b).

- |     |    |                    |           |       |
|-----|----|--------------------|-----------|-------|
| (1) | a. | Is it raining?     | $p?$      | PPINT |
|     | b. | Is it not raining? | $\neg p?$ | NPINT |

The semantic value of (1a) and (1b) is the same: both denote a set of propositions (Hamblin 1973; Groenendijk–Stokhof 1997). In spite of their semantic equivalence, the possible contexts in which they can be used felicitously are different (Ladd 1981; Büring–Gunlogson 2000; Romero–Han 2004; Sudo 2013, among others). The choice between PPINT and NPINT depend on several factors: the availability of contextual evidence; the speaker’s beliefs, expectations, desires; and the overall goal of the conversation.

In Hungarian, interrogativity is either marked by intonation (2a,  $\wedge$ PPINT), that is, a rising-falling tone which has its peak on the penultimate syllable (marked by  $\wedge$ ), or by question particle *-e* attached to the finite verb (2a’, PPINT-E). Negative forms (2b–b’,  $\wedge$ NPINT, NPINT-E) are ambiguous between the “inside” and the “outside” negation readings (Gyuris 2017).

- |     |    |                      |     |      |    |                      |     |      |
|-----|----|----------------------|-----|------|----|----------------------|-----|------|
| (2) | a. | $\wedge$ Esik        | az  | eső? | a’ | Esik-e               | az  | eső? |
|     |    | fall.PRES.3SG        | the | rain |    | fall.PRES.3SG-E      | the | rain |
|     |    | ‘Is it raining?’     |     |      |    | ‘Is it raining?’     |     |      |
|     | b. | $\wedge$ Nem esik    | az  | eső? | b’ | Nem esik-e           | az  | eső? |
|     |    | not all.PRES.3SG     | the | rain |    | not all.PRES.3SG-E   | the | rain |
|     |    | ‘Is it not raining?’ |     |      |    | ‘Is it not raining?’ |     |      |

Experimental evidence showed that in Hungarian  $\wedge$ PPINT forms are clearly preferred over  $\wedge$ NPINT forms both in neutral and in biased contexts. Although contextual evidence for the negative answer increases the preference of  $\wedge$ NPINT forms,  $\wedge$ PPINT seems to be the default polar interrogative form (Gyuris–Molnár–Mády 2017).

Narrative biblical texts frequently establish biased contexts, thus one can expect that the use of biased question forms is frequent, and has text-organizing or persuasive functions (Elbert 2004; Estes 2013). The present talk aims at contributing to the description and understanding of the use conditions NPINT forms in Hungarian with the help of a corpus-based investigation of such forms in different translations of the gospels. The study takes into account 8 different translations, the “macro” and the “micro” contexts of the individual NPINT-occurrences, e. g. the speaker’s status, background knowledge, expectations, intentions; the relation between the discourse participants, the common ground of the discourse, and the “compelling contextual evidence” present in the situation.

(3f) shows the ESV translation of John 11:40b, whose current Hungarian translations are listed in (3a–e). Jesus addresses Martha just before going to the tomb of her brother, Lazarus, to bring him back to life. In this situation (3) can be interpreted as a reminder, as a reproach, as an argument against what Martha has previously said, or as a request to let him go to the tomb, and believe. (3d) only allows the outside negation reading (Gyuris 2016), while (3a, c) remain ambiguous between the inside and the outside negation reading. (3b, e) use tag-like particles (*nemde, ugye*); (3e) completely avoids negation, while (3b) contains the negative element *nem* incorporated in *nemde*, which can only have outside negation reading. In (3) different forms lead to different interpretive possibilities with respect to the degree of commitment of the speaker. (3a, c) express weaker commitment than (3d), while (3b, e) express the strongest commitment.

NPINT forms like those in (3a-d), are relatively frequent in the gospel of John: out of 74 polar interrogatives 30 are NPINT forms. The talk gives an overview of the NPINT forms and functions used in the Hungarian Gospel-translations, and presents case-studies of certain loci of the biblical corpus, especially those having different interpretative possibilities.

- (3) a. **Nem** azt mondtam neked, hogy ha hiszel, meglátod Isten dicsőségét?<sup>12</sup>  
not it.ACC say.PST.1SG you.DAT
- b. **Nemde** azt mondtam: ha hiszel, meglátod Isten dicsőségét?  
not-yes it.ACC say.PST.1SG
- c. **Nem** mondtam: ha hiszel, meglátod Isten dicsőségét?  
not say.PST.1SG
- d. **Nem** mondtam-e neked, hogy ha hiszel, meglátod az Isten dicsőségét?  
not say.PST.1SG-E you.DAT
- e. **Ugye** megmondtam neked, hogy ha hiszel, meglátod Isten dicsőségét?  
TAG VM-say.PST.1SG-E you.DAT
- f. Did I not tell you that if you believed you would see the glory of God?

### Bibliography

- Büring, Daniel – Gunlogson, Christine (2000). Aren't Positive and Negative Polar Questions the Same? (ms.)
- Elbert, Paul (2004). An Observation on Luke's Composition and Narrative Style of Questions. *The Catholic Biblical Quarterly*, January 2004, Vol. 66, No. 1, pp. 98–109
- Estes, Douglas (2013). *The Questions of Jesus in John. Logic, Rhetoric and Persuasive Discourse.* Brill, Leiden, Boston
- Farkas, Donka – Bruce, Kim (2010). On Reacting to Assertions and Polar Questions. *Journal of Semantics* 27: pp. 81–118.
- Farkas, Donka – Roelofsen, Floris (2017). Division of Labor in the Interpretation of Declaratives and Interrogatives. *Journal of Semantics*, 34, pp. 237–289.
- Farkas, Donka (2022). Non-Intrusive Questions as a Special Type of Non-Canonical Questions. *Journal of Semantics* 39:295-337.
- Groenendijk, Jeroen – Stokhof, Martin (1997). Questions. In: *Handbook of Logic and Language*, edited by Johan van Benthem and Alice Ter Meulen, Cambridge: MIT.
- Gyuris Beáta (2016). A magyar nyelv tagadószós eldöntendő kérdő mondatainak jelentéséhez. *Jelentés és Nyelvhasználat* 3: 169–190.
- Gyuris, Beáta (2017). New perspectives on bias in polar questions: a study of Hungarian -e. *International Review of Pragmatics*, 9(1), pp. 1–50.
- Gyuris, Beáta, Molnár, Cecília Sarolta and Mády, Katalin (2017). A magyar eldöntendő kérdő mondatok használatának tanulmányozása kísérletes módszerekkel. In Zoltán Bánréti (ed.), *Általános Nyelvészeti Tanulmányok XXIX. Kísérletes nyelvészet.* Budapest: Akadémiai Kiadó, pp. 53-75.
- Hamblin, Charles L. (1973). Questions in Montague English. *Foundations of Language* 10:1, pp. 41–53
- Kohnen, Thomas (2010). Religious discourse. In: *Historical Pragmatics*, edited by Jucker, Andreas H. and Taavitsainen, Irma. Berlin, New York: De Gruyter Mouton, 2010. pp. 523–549.
- Ladd, D. Robert (1981). A first look at the semantics and pragmatics of negative questions and tag questions. In: *Papers from the 17th Regional Meeting of the Chicago Linguistic Society* 17. University of Chicago, 164–171.
- Romero, M., Han, CH. (2004). On Negative Yes/No Questions. *Linguistics and Philosophy* 27, 609–658 (2004). <https://doi.org/10.1023/B:LING.0000033850.15705.94>
- Sudo, Yasutada (2013). Biased polar questions in English and Japanese. In Daniel Gutzmann and Hans-Martin Gärtner (eds.), *Beyond Expressives. Explorations in Conventional Non-truth-conditional Meaning.* Leiden: Brill, pp. 277-297
- Van Rooy, Robert, Šafářová, Maria (2003). On polar Questions. In: *Proceedings from SALT 13*, edited by Robert B. Young and Yuping Zhou, pp. 292–309.

<sup>12</sup> Hungarian translations are cited from <https://ebo.kre.hu/>: (3a): KNB-1997, (3b): SZIT-1973, (3c): BD-1951, (3d): RUF-2014, (3e): STL-2017. The English version (3f) is cited from English Standard Version (ESV) from <https://www.biblestudytools.com/parallel-bible>.

## Cross-linguistic variation in the meaning of high negation in polar questions

Radek Šimík, Anna Staňková, Michaela Chodounská (Charles University)

**Background** So-called high negation in polar questions/PQs (*Aren't you...* = high neg vs. *Are you not...* = low neg) has been observed to trigger the speaker's commitment/epistemic bias to the positive prejacent [1,5,9,12 *inter alia*]. The question *Aren't you going to Berlin?* (cf. the affirmative counterpart *Are you going to Berlin?*) implies that the speaker had held the belief (at least until shortly prior to the utterance) that the addressee was going to go to Berlin. It is felicitously uttered in a situation where the speaker is exposed to evidence that appears to contradict this belief (e.g. the addressee turning an unexpected direction). The implication has been documented cross-linguistically [1,8,12].

**Contribution** We argue that high negation in Czech has a weaker meaning and, correspondingly, a broader distribution. The Czech translation of the PQ above, (1), can convey prior speaker's belief, as in English, but is equally felicitous in a situation where the speaker only deems it *possible* that the addressee is going to Berlin. The PQ could thus be uttered by a hitchhiker trying to find somebody who would take her to Berlin. Moreover, Czech high neg PQ are not restricted to epistemic (or factual) queries but can also be used as offers and requests. The PQ in (2) can be interpreted as an offer to the addressee to join a trip to Berlin. Czech high neg PQs are thus **weaker** in their **force** (possibility rather than necessity) and **underspecified** in their **flavor** (not just epistemic, but also root); (3).

- (1) Nejedete do Berlína?      (2) Nejedete s námi do Berlína?  
 NEG.go to Berlin            NEG.go with us to Berlin  
 'Are(n't) you going to Berlin?'    'Would you like to go to Berlin with us?'

- (3)  $[[\text{high neg } p]]^w$  conventionally implicates  $\exists w' [MB_w(w') \ \& \ [[p]](w') = 1]$   
 (where MB is a suitable modal base; epistemic/stereotypical in ordinary/factual PQs)  
 $[[\text{high neg } p]]^w = \{[[p]], \sim[[p]]\}$

**Empirical arguments 1.** The obligatory strong positive epistemic bias in English high neg PQs is accompanied by the incompatibility with positive contextual evidence [1,2,12]. Czech high neg PQs are compatible with positive evidence and convey a suggested explanation for a currently observed effect (cf. [3]); (4). **2.** The adverbial *by (any) chance* has been claimed to be incompatible with strong epistemic bias (*Are(\*n't) you going to Berlin by any chance?*) [6,11]. Its Czech translation *náhodou*, on the other hand, is even *conditioned* by the presence of high negation (example (4) would be ungrammatical without negation), which is supported by corpus evidence (syn v12; [4]): 98 out of 100 randomly chosen PQs with *náhodou* involve high neg. We argue that *by chance/náhodou* relaxes the stereotypical ordering source of an epistemic possibility modal (present in English positive PQs and Czech high neg PQs). This predicts that *náhodou* can only be used in ordinary (factual) PQs (e.g. (1)) and not in offers or requests (e.g. (2)); this prediction will be shown to be confirmed by experimental (rating) evidence. **3.** A parallel corpus search (InterCorp v16; [10]) reveals that translations of English high neg PQs to Czech yield high neg significantly more often than translations from Czech to English (88 % vs. 69 %) ( $\chi^2(1) = 10.7$ ,  $p = .001$ ). This is consistent with the weaker lexical specification of Czech high neg as compared to the English one: Czech high neg is expected to appear in a proper superset of situations in which English high neg is used.

- (4) *Scenario*: Sounds coming from the flat above. [evidence for positive prejacent of the ex below]  
 Nestěhují sousedi (náhodou) piano?  
 NEG.move neighbors by.chance piano  
 'Are the neighbors moving a piano, perhaps?'

**Implications** The facts presented shed doubt on accounts which attempt to derive the semantic contribution of high neg in PQs by standard negation applied in a high syntactic position [2,9] and suggest an approach in terms of a specialized operator (like FALSUM; [7]), which in turn lends itself more easily to cross-linguistic parametrization.

## Bibliography

- [1] Gärtner, Hans-Martin and Beáta Gyuris. 2017. On delimiting the space of bias profiles for polar interrogatives. *Linguistische Berichte* 251, 293–316.
- [2] Goodhue, Daniel. 2022. [Isn't there more than one way to bias a polar question?](#) *Natural Language Semantics* 30(4), 379–413.
- [3] Korotkova, Nataša. 2023. [Conversational dynamics of Russian questions with \*razve\*](#). In *Proceedings of Sinn und Bedeutung 27*, 328–346. Praha: Charles University.
- [4] Křen, Michal et al. 2022. Corpus SYN, version 11. Praha: Charles University.
- [5] Ladd, Robert. 1981. A first look at the semantics and pragmatics of negative questions and tag questions. In *Papers from the Seventeenth Regional Meeting of the Chicago Linguistics Society*, 164–171. Chicago, IL: CLS.
- [6] Reese, Brian and Nicholas Asher. 2010. Biased questions, intonation and discourse. In Malte Zimmermann and Caroline Féry (eds.), *Information structure: Theoretical, typological, and experimental perspectives*, 139–173. Oxford: Oxford University Press.
- [7] Repp, Sophie. 2013. Common ground management: Modal particles, illocutionary negation and VERUM. In Daniel Gutzmann and Hans-Martin Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning*, 231–274. Leiden: Brill.
- [8] Repp, Sophie and Ljudmila Geist. 2023. Negative polar questions in Russian: Question bias and question concern. Manuscript, University of Cologne and University of Stuttgart.
- [9] Romero, Maribel and Chung-Hye Han. 2004. [On negative yes/no questions](#). *Linguistics and Philosophy* 27(5), 609–658.
- [10] Rosen, Alexandr, Martin Vavřín, and Adrian Zasina. 2023. Korpus InterCorp – čeština, verze 16. Praha: Charles University.
- [11] Sadock, Jerry. 1971. Queclaratives. In *Papers from the Seventh Regional Meeting of the Chicago Linguistics Society*, 223–231. Chicago, IL: CLS.
- [12] Sudo, Yasutada. 2013. Biased polar questions in English and Japanese. In Daniel Gutzmann and Hans-Martin Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning*, 275–295. Leiden: Brill.



## Word order and information structure in Czech polar questions

Anna Staňková (Charles University)

**Background** Czech polar questions (PQs) are signaled by rising intonation and interrogative (verb-first) word order (Štícha 1995). Recent research shows that negation in interrogative PQs is interpreted as OUTER NEGATION ( $\rightarrow$  ON-PQs), which can be analyzed as the operator FALSUM. FALSUM conveys the speaker's EPISTEMIC BIAS, namely their beliefs about the possible answer (Repp 2013, Staňková 2023, Šimík to appear). Negation in declarative PQs with the canonical SVO order is often interpreted as INNER, i.e. semantic negation, which conveys EVIDENTIAL BIAS of the speaker (Sudo 2013). Staňková & Šimík (to appear) further claim that outer negation is compatible both with the interrogative and declarative PQs.

**Contribution** I zoom in on the difference between interrogative and declarative ON-PQs, esp. how the choice between these two strategies is motivated. I claim that it is motivated by information structure. The interrogative strategy is suitable if the question contains an exhaustive focus (EXH) on some XP; the declarative strategy is suitable if the question contains a contrastive topic (CT). I propose explicit analyses of these two types, utilizing Rooth's (1995) F-marking system and Constant's (2014) CT- $\lambda$  operator.

**Exhaustive focus vs. contrastive topic in ON-PQs** The subject in the interrogative ON-PQ in 1B allows for an exhaustive reading. 1B asks about B(p) (here B(p) = 'Petr bought a car') and implicates  $\forall x.x \neq p \rightarrow \neg B(x)$  (i.e. it is not the case that Eva bought a car, Maria bought a car etc.). It is an instance of pragmatic exhaustivity, since it is cancelable by the additive particle *také* 'too'. An additional inference arises, namely that B corrects A's statement (corrective focus). If the negated verb stays in situ, as in 1B', the implicature does not appear, resulting into an anti-exhaustive reading (the speaker considers the proposition 'Eva bought a car' as possibly true). Crucially, the choice between an interrogative vs. declarative PQ depends on the presence of a CT or the exhaustive focus.

- (1) A: Eva si prý koupila auto. B: Nekoupil si (také) [Petr]<sub>EXH</sub> nějaké auto? VSO  
 Eva REFL allegedly bought car NEG.bought REFL (too) Petr PPI car  
 'Allegedly, Eva bought a car.' B': A [Petr]<sub>CT</sub> si nekoupil nějaké auto? SVO  
 and Petr REFL NEG.bought PPI car  
 'Didn't Petr buy a car?'

**Proposal** In ON-PQs, the main verb indicates standard polarity focus – the speaker wishes to resolve whether p or  $\neg p$  holds. Once there is an XP with exhaustive focus in the PQ, the negated verb moves into the initial position, see (2). The verb fronting is a purely formal movement on the surface level. It could be triggered by a clitic in the structure, which seeks phonological support, or to signal further epistemic states of the speaker.

- (2) [FALSUM [NEG-V [[SUBJECT]<sub>F</sub> tV ...]]] interrogative ON-PQ

The declarative word order is only possible when the ON-PQ introduces a CT to the discourse. In line with Wagner (2012), I suggest that on the LF, the CT is F-marked (since it evokes a set of alternatives), and it always precedes the focused verb; thus their mutual distribution is restricted. This is enabled by the CT- $\lambda$  operator as proposed by Constant (2014). The CT raises into the position above CT- $\lambda$ , which binds the CT with its trace, see (3) with a CT subject and covert movement. Importantly, CT raising can likewise be overt (in the case of CT objects, for instance). The verb obligatorily carries prosodic prominence (details of prosodic realization to be discussed at the talk).

- (3) [FALSUM [[SUBJECT]<sub>F</sub> [CT- $\lambda$  [<sub>t</sub>1 [[NEG-V]<sub>F</sub> ...]]]]] declarative ON-PQ

## **Bibliography:**

**Constant 2014** Contrastive topic: Meanings and realizations. PhD thesis. UMass.

**Repp 2013** Common ground management: Modal particles, illocutionary negation and verum. In *Beyond Expressives: Explorations in Use-Conditional Meaning*, eds. Daniel Gutzmann and Hans-Martin Gärtner. Brill. 231–274. Brill.

**Rooth 1995** A theory of focus interpretation. In *Natural Language Semantics* 1: 75–116.

**Staňková 2023** The expression of speaker's bias in Czech polar questions. Master's thesis. CU.

**Staňková & Šimík to appear** Negation in Czech polar questions. In *Journal of Slavic Linguistics – FASL 32 special issue*.

**Sudo 2013** Biased polar questions in English and Japanese. In *Beyond Expressives: Explorations in Use-Conditional Meaning*, eds. Daniel Gutzmann and Hans-Martin Gärtner. Brill.

**Šimík to appear** Polar question semantics and bias: Lessons from Slavic/Czech. In *Topics in the semantics of Slavic languages*, eds. Berit Gehrke and Radek Šimík. Language Science Press.

**Štícha 1995** Otázky predikátové: inference, implicitnost a explicitní výrazy ilokučních funkcí II. In *Naše řeč* 56(3), 204–217.

**Wagner 2012** Contrastive topics decomposed. In *Semantics and Pragmatics* 5(8): 1–54.

## Negation is present: Subjunctive questions in Serbian

Neda Todorović (Reed College)

This paper discusses negated subjunctive polar questions in Serbian, in which negation seems vacuous. This appears supported by their lack of NPI-licensing. I argue that negation is not vacuous, but it is too high to license an NPI. Plus, only if contentful, can the negation interact with modals the way it does.

**Data** In Serbian, an indicative polar question (IQ) contains *da* and a question clitic *li* (1). A subjunctive polar question (SQ) contain *da*SUBJV and present tense (2). (2) denotes modality, despite the absence of an overt modal (Oikonomou & Ilić to appear (O&I)).

- (1) **Da** **li** deca večeraju? (2) **Da** deca večeraju?  
*da*IND Q kids dine.3PL.PRES *da*SUBJV kids dine.3PL.PRES  
 ‘Are the kids having dinner?’ ‘Should the kids have dinner?’ (adapted from O&I 4,5)

IQ and SQ combine with negation. In IQ, negation is in different slots (3a vs. 3b), but it is interpretable. In SQ, negation is in the same slot, but with different meanings. In NegSQ<sub>1</sub> (4a), the negation is interpretable – given the questionee’s authority, should the kids not have dinner. In NegSQ<sub>2</sub> (4b), it is not – given the questionee’s knowledge, are the kids having dinner. So, is the negation in (4b) vacuous?

- (3) a. **Da** **li** deca **ne** večeraju? b. **Ne** večeraju **li** deca?  
*da*IND Q kids NEG dine.3PL.PRES NEG dine.3PL.PRES Q kids  
 ‘Are the kids not having dinner?’ (low NegIQ) ‘Aren’t the kids having dinner?’ (high NegIQ)
- (4) a. **Da** deca **ne** večeraju? b. **Da** deca (možda) **ne** večeraju?  
*da*SUBJV kids NEG dine.3PL.PRS *da*SUBJV kids maybe NEG dine.3PL.PRS  
 ‘Should the kids not have dinner?’ (NegSQ<sub>1</sub>) ‘Could it be that the kids are having dinner?’ (NegSQ<sub>2</sub>)

**NPIs** Serbian has *ni*-NPIs and *i*-NPIs. *Ni*-NPIs occur with a clausemate negation, *i*-NPIs do not tolerate it, as in low NegIQ (5a) and NegSQ<sub>1</sub> (5b). High NegIQ (6a) and NegSQ<sub>2</sub> (6b) do not license *ni*-NPIs.

- (5) a. **Da** **li** deca **ne** večeraju \*išta / **ništa**? b. **Da** deca **ne** večeraju \*išta / **ništa**?  
*da*IND Q kids NEG dine anything/ nothing *da*SUBJV kids NEG dine anything/nothing  
 ‘Are the kids not eating anything for dinner?’ ‘Should the kids not eat anything for dinner?’
- (6) a. **Ne** večeraju **li** deca išta / \***ništa**? b. **Da** deca. **ne** večeraju išta / \***ništa**?  
 NEG dine Q kids anything/ nothing *da*SUBJV kids NEG dine anything/ nothing  
 ‘Aren’t the kids eating something for dinner?’ ‘Could it be the kids are eating something for dinner?’

Does the lack of *ni*-NPI in (6b) equal vacuous negation? No. Assume that (a) NPI-licensing is in the TP and (b) there are two polarity phrases, one below and one above the TP (Progovac 2005). If the negation is below the TP, it licenses *ni*-NPIs (5a-5b). If it is above the TP, it does not (6a-6b). Crucially, this is due to syntax; high negation can still be contentful.

**Modal + negation** Consider (4a). For O&I, (i) *da*SUBJV is a weak necessity prioritizing modal above T. Assume that (i) holds, (ii) negation is contentful and below TP, (iii) question is the set of its possible answers. Combining it (7), we get the set of answers in which, under the questionee’s authority, the kids should not run or it is possible for them to run. Consider (4b). Assume that (i) negation is contentful, above TP, but below the modal, and assume (iii). Combining it (8), the interaction of modal and negation is like in (7). However, in (4b), *da*SUBJV is a weak necessity epistemic modal. So, we get the set of answers in which, given the questionee’s knowledge, it should be the case that the kids are not running or it is possible that they are running. Crucially, if the negation is contentful, we can capture the meaning of (4b) in a predictable way. Finally – why is the negation absent in the translation of (4b)? This is due to the questioner’s positive bias, which I will show is captured via the semantics-pragmatics interaction.

- (7) [Qop [Mod [T [Neg...]]]] ?¬p = {¬p, ¬¬p} = {□(¬p), ¬(□(¬p))} = { □(¬p), ◇p } =  
 {It should be [that the kids not run]; It is possible [that the kids run]}
- (8) [Qop [Mod [Neg [T...]]]] ?¬p = {¬p, ¬¬p} = {□(¬p), ¬(□(¬p))} = { □(¬p), ◇p } =  
 {It should be [that the kids not run]; It is possible [that the kids run]}

## **Bibliography**

Oikonomou, Despina & Ivona Ilić. To appear. Subjunctive Questions in Serbian. Special issue of *Journal of Slavic Linguistics: Proceedings of FASL* 30.

Progovac, Ljiljana. 2005. *A Syntax of Serbian*. Bloomington, IN: Slavica Publishers.

## Proposed negation vs. non-Proposed negation in Greek polar questions

Evripidis Tsiakmakis (Universitat Pompeu Fabra)

Romero and Han (2004), building on insights mainly from Ladd (1981) and Büring and Gunlogson (2000), introduce a distinction between Preposed Negation English polar questions (1a), that convey a positive speaker bias, and non-Preposed Negation English polar questions (1b), that may but need not convey positive speaker bias.

- (1) a. Didn't John drink coffee?                      Preposed Negation polar question  
       b. Did John not drink coffee?                    non-Preposed Negation polar question

The authors proceed to tentatively extend their generalization to Greek, citing the example below (Romero & Han 2004: 614, ex. (14)):

- (2) a. *Dhen* ipie o Yanis kafe?    b. O Yanis *dhen* ipie kafe?  
       NEG drank the Yanis coffee    the Yanis NEG drank coffee  
       'Didn't Yanis drink coffee?'    'Did Yanis not drink coffee?'

Importantly, Greek is a pro-drop language (Holton et al. 1997); this means that the distinction in (2) is observable only in those cases where the subject is overtly expressed. Moreover, the most natural way to form a polar question in Greek is by preposing the verb (Holton et al. 1997) –or the negation + verb complex; this suggests that the preposed status of the negative marker *dhen* in (2a) is not necessarily equivalent to the status of its English counterpart in (1a).

Given the background described above, I sought additional empirical support for a preposed vs. non-preposed negation distinction in Greek polar questions by applying the tests already established in the literature (Ladd 1981; Sailor 2013; Goodhue 2022, a.o.). Unfortunately, all of them failed to provide conclusive evidence in favor of this distinction. Specifically, the alleged Preposed Negation variant in (2a) is compatible with the licensing of both the Positive Polarity Item *idhi* 'already' and the Negative Polarity Item *akoma* 'yet' (3). This means that, at least under certain circumstances, *dhen* in (2a) is interpreted similarly to *dhen* in (2b).

- (3) *Dhen* ipie *idhi/ akoma* o Yanis kafe?  
       NEG drank already yet the Yanis coffee  
       'Didn't Yanis drink coffee already?'/ 'Did Yanis not drink coffee yet?'

Moreover, the preposed negation in (2a) is compatible with alternative questions (4). This would not be the case if Greek Preposed Negation questions were necessarily biased questions, as Romero and Han (2004) suggest.

- (4) *Ipie* i *dhen* ipie o Yanis kafe?  
       drank or NEG drank the Yanis coffee  
       'Did Yanis drink coffee or not?'

Last but not least, modification by *again* (5) or an *as*-parenthetical (6) shows that, in Greek Preposed Negation questions like (2a), the positive proposition corresponding to *Yanis drank coffee* is as available as its negative counterpart *Yanis didn't drink coffee*:

- (5) *Dhen* ipie o Yanis kafe *pali*?  
       NEG drank the Yanis coffee again  
       'Didn't Yanis drink coffee again?'/ 'Did Yanis not drink coffee again?'
- (6) *Dhen* ipie o Yanis kafe, opos proevlepses?  
       NEG drank the Yanis coffee as you.predicted  
       'Didn't Yanis drink coffee, as you predicted?'/ 'Did Yanis not drink coffee, as you predicted?'

Greek negative polar questions can receive biased interpretations, along the lines of Romero and Han (2004). At the same time, the data presented here suggest that the surface position of Greek negation in questions cannot be straightforwardly associated with the presence of speaker bias. This invites a more meticulous comparison of Greek and English Preposed Negation questions, as well as a closer examination of the form of biased polar questions in Greek.

## Bibliography

Büring, Daniel & Cristine Gunlogson. 2000. Aren't positive and negative polar questions the same? Ms.

Goodhue, Daniel. 2022. Isn't there more than one way to bias a polar question? *Natural Language Semantics* 30. 379–413.

Holton, David, Peter Mackridge & Irene Philippaki-Warbuton. 1997. *Greek: A Comprehensive Grammar of the Modern Language*. London: Routledge.

Ladd, D. Robert. 1981. A first look at the semantics and pragmatics of negative questions and tag questions. In Robert A. Hendrick, Carrie S. Masek & Mary Frances Miller (eds.), *Proceedings of Chicago Linguistic Society* 17, 164–171.

Romero, Maribel & Chung-hye Han. 2004. On negative yes/no questions. *Linguistics and Philosophy* 27(5). 609–658.

Sailor, Craig. 2013. Questionable negation. Handout from the LSA annual meeting. Boston, MA.

17

*Mezi jazyky: metody kontrastivního  
výzkumu s využitím korpusů*

*Between languages: methods of  
contrastive research  
using corpora*

organizují / organizers

Alexandr Rosen

Michaela Martinková

Olga Nádvorníková

(CZE, ENG)

Bienále české lingvistiky 2024  
Biennial of Czech Linguistics 2024

## Popis panelu / Workshop description

(CZE) Srovnávání jazyků má dlouhou tradici a zahrnuje řadu přístupů – od historických, přes typologické až po kontrastivně-lingvistické. Jejich cíle, šíře záběru i metody zkoumání se liší. Zájem o kontrastivní analýzu navíc přicházel ve vlnách: nadšení v 60. letech, kdy byla kontrastivní analýza vnímána jako cesta k zefektivnění jazykové výuky, je vystřídáno zklamáním a nakonec realistickým pohledem na roli transferu. S rozvojem funkčních přístupů k jazyku, příchodem kognitivní lingvistiky a zejména nástupem jazykových korpusů však dochází k renesanci kontrastivní lingvistiky a také k jejímu sbližování s translatologií (Grangerová 2003). Pro další vývoj kontrastivní lingvistiky jsou zásadní jak korpusy srovnatelné, tak korpusy paralelní (překladové), protože představují cenný bilingvní output, který ospravedlňuje předpoklad srovnatelnosti (Gast 2012). Zároveň se rozvíjí metodologie kontrastivního výzkumu na paralelních korpusech, která mimo jiné bere v úvahu potenciální rozdíly mezi překladovým a nepřekladovým jazykem (Johansson 2007), a v deskriptivní translatologii nastoluje Bakerová (1993) jako novou agendu studium tzv. překladových univerzálií, tedy specifických rysů překladu.

Náš panel je otevřen příspěvkům na nejrůznější témata spojená s kontrastivním zkoumáním jazyků na základě korpusů (zejména paralelních, ale i srovnatelných a jednojazyčných). I přes tématickou pestrost analyzovaných jevů očekáváme, že se v příspěvcích projeví sdílený zájem o metodologii korpusově založené kontrastivní lingvistiky (Enghels a kol. 2020), přičemž uvítáme také příspěvky zaměřené na korpusově založenou translatologii a na průniky mezi translatologií a kontrastivní lingvistikou. Účastníky by tak spojoval právě tento akcent na metodologické problémy a jejich možná řešení, a to napříč různými jazyky i teoretickými přístupy. Mezi možné metodologické otázky patří např.:

- tertium comparationis (Connor & Moreno 2005; Gast 2012, Egan 2013)
- vizualizace dat (Atayan & Gast, in prep)
- jazyková triangulace (Malamatidou 2018; Martinková, v tisku)
- reprezentativnost (Biber 1994)
- statistické zpracování dat (Granger & Lefer 2022)
- rozdíly v objemu dat mezi různými jazyky a žánry, srovnatelnost dat
- srovnávání typologicky odlišných jazyků
- výhody a nevýhody srovnatelných a paralelních korpusů
- nové typy dat v kontrastivní lingvistice, např. mluvená data (tlumočení, filmové titulky, Enghels a kol. 2020)
- možnosti kompenzace omezeného množství dat pro některé jazyky
- jazykové modely a jejich využití v kontrastivním zkoumání jazyků (Boas 2010)
- paralelní korpusy, specifika překladového jazyka a kontrastivní lingvistika
- využití jednotné morfologické a syntaktické anotace podle standardu Universal Dependencies (<https://universaldependencies.org/>)
- umělá inteligence ve srovnání s výsledky analýz korpusových dat

Panel by měl být jedinečnou příležitostí k setkáním s kolegy i k objevování nových přístupů a koncepcí. I proto na závěr zařadíme diskusi u kulatého stolu.



**(ENG)** Comparing languages has a long tradition and encompasses a variety of approaches – ranging from historical to typological and contrastive. Their goals, scopes, and research methods differ. Moreover, the interest in contrastive analysis has come in cyclical patterns (Granger 2004, Enghels et al. 2020): the enthusiasm in the 1960s, when contrastive analysis was seen as a way to more effective foreign language teaching, was replaced by disappointment and eventually a realistic view of the role of transfer. With the development of functional approaches to language, the advent of cognitive linguistics, and especially the rise of language corpora, contrastive linguistics has undergone a renaissance and has also come closer to translation studies (Granger 2003). For the further development of contrastive linguistics, both comparable and parallel (translation) corpora are crucial, as they represent valuable bilingual output, justifying the assumption of comparability (Gast 2012). At the same time, the methodology of contrastive research on parallel corpora is developing, which, among other things, takes into account potential differences between the translated and non-translated language (Johansson 2007). In descriptive translation studies, Baker (1993) sets a new agenda for the study of so-called translation universals, i.e., specific features of translation.

Our panel is open to contributions on various topics related to the contrastive examination of languages based on corpora (especially parallel, but also comparable and monolingual). Despite the thematic diversity of the analyzed phenomena, we expect the contributions to reflect a shared interest in the methodology of corpus-based contrastive linguistics (Enghels et al. 2020). We also welcome contributions focused on corpus-based translation studies and intersections between translation studies and contrastive linguistics. Participants should thus be united by this emphasis on methodological problems and their possible solutions, across various languages and theoretical approaches. Possible methodological questions include:

- Tertium comparationis (Connor & Moreno 2005; Gast 2012, Egan 2013)
- Data visualization (Atayan & Gast, in prep)
- Language triangulation (Malamatidou 2018; Martinková, in print)
- Representativeness (Biber 1994)
- Statistical data processing (Granger & Lefer 2022)
- Differences in data volume between different languages and genres, data comparability
- Comparison of typologically different languages
- Advantages and disadvantages of comparable and parallel corpora
- New types of data in contrastive linguistics, e.g., spoken data (interpretation, movie subtitles, Enghels et al. 2020)
- Options of compensating for limited data availability for some languages
- Language models and their use in contrastive language studies (Boas 2010)
- Parallel corpora, specifics of translation language and contrastive linguistics
- Use of uniform morphological and syntactic annotation according to the Universal Dependencies standard (<https://universaldependencies.org/>)
- Artificial intelligence in comparison with corpus data analysis results

The panel should be a unique opportunity for meetings with colleagues and for discovering new approaches and concepts. We will conclude with a round-table discussion to open this window for all participants.

## Literatura / References

- Atayan, Vahram & Gast, Volker (in prep). Visualizing multi-dimensional count data in translation studies and contrastive linguistics. In: Kloudová, V., Nádvorníková, O. & Svoboda, T. (eds.), *Advancing Interdisciplinarity in Empirical Translation Research*. Berlin: LangSci.
- Baker, Mona (1993). Corpus linguistics and Translation studies. In Baker, M., Francis, G. & Tognini-Bonelli, E. (eds.) *Text and Technology*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Biber, Douglas (1994). Representativeness in Corpus Design. In: Zampolli, A., Calzolari, N. & Palmer, M. (eds) *Current Issues in Computational Linguistics: In Honour of Don Walker*. *Linguistica Computazionale*, vol 9. Springer, Dordrecht. [https://doi.org/10.1007/978-0-585-35958-8\\_20](https://doi.org/10.1007/978-0-585-35958-8_20)
- Boas, Hans C. (2010). *Contrastive studies in Construction Grammar*. John Benjamins Publishing Company.
- Connor, Ulla M. & Moreno, Ana I. (2005). Tertium Comparationis: A vital component in contrastive research methodology. In Bruthiaux P., Atkinson, D., Eggington, W. G., Grabe, W. & Ramanathan, V. (eds) *Directions in Applied Linguistics: Essays in Honor of Robert B. Kaplan*. Clevedon, 153–164.
- Egan, Thomas (2013). *Tertia comparationis* in multilingual corpora. In Aijmer K. & Altenberg B. (eds) *Advances in Corpus-based Contrastive Linguistics: Studies in honour of Stig Johansson*. John Benjamins Publishing Company, 7–24.
- Engiels, Renata & Defrancq, Bart & Jansegers, Marlies (eds.) (2020). *New Approaches to Contrastive Linguistics: empirical and methodological challenges*. Ghent: Universiteit Gent.
- Gast, Volker (2012). Contrastive analysis: Theories and methods. In Kortmann B. (ed.) *Dictionaries of Linguistics and Communication Science: Linguistic theory and methodology*. Berlin: Mouton de Gruyter.
- Granger, Sylviane (2003). The Corpus Approach: A Common Way Forward for Contrastive Linguistics and Translation Studies? In Granger, S., Lerot, J. & Petch-Tyson, S. (eds.), *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam: Rodopi.
- Granger, Sylviane & Lefer, Marie-Aude (eds) (2022). *Extending the Scope of Corpus-based Translation Studies*. Bloomsbury Advances in Translation series. London: Bloomsbury.
- Johansson, Stig (2007). Seeing through Multilingual Corpora. In Facchinetti R. (ed.) *Corpus linguistics 25 Years on*. Amsterdam – New York: Rodopi.
- Malamatidou, Sofia (2018). *Corpus Triangulation: Combining Data and Methods in Corpus-Based Translation Studies*. London and New York: Routledge.
- Martinková, Michaela (v tisku). *English and Czech Motion Verbs, Constructions, and ‘Satellite-framing’*. Olomouc: UPOL.

# Obsah / Table of Contents

*Kauzativní konstrukce v italštině a španělštině a jejich české protějšky v paralelním korpusu: statistický model predikce překladové ekvivalence*

Petr Čermák, Pavel Štichauer (Univerzita Karlova)

*Je viitsiä jedinečné? Lexikálně-syntaktický popis jednoho finského modálního slovesa*

Lenka Fárová (Univerzita Karlova)

*Slovesný vid v rusko-českém kontrastivním korpusovém výzkumu*

Markus Giger (Ústav východoevropských studií Filozofické fakulty Univerzity Karlovy)

*Rozdíly v užívání pasivních konstrukcí v chorvatštině a češtině*

Karel Jirásek (Filozofická fakulta Univerzity Karlovy)

*Polish modal expressions and their Czech translation equivalents based on the parallel corpus InterCorp*

Elżbieta Kaczmarska-Zglejszewska (University of Warsaw)

*InterCorp & co. – Co se všechno (ne)stihlo*

Tomáš Káňa (Masarykova univerzita)

*Srovnatelnost dat v rámci paralelních korpusů. Pasti dat*

Jana Kocková (Slovanský ústav AV ČR)

*České otázky s podmětem vyjádřeným zájmeny 1. a 2. osoby v zrcadle anglických překladových protějšků*

Markéta Malá, Zuzana Nádraská (Univerzita Karlova)

*České a polské modální výrazy – výzkum na materiálu InterCorpu*

Karolína Skwarska (Slovanský ústav AV ČR)

*Specifika překladového jazyka v souvislosti se slovesy pohybu na příkladu češtiny a španělštiny*

Petr Šlechta (Ostravská univerzita)

*Sémantická a gramatická životnost v češtině a polštině*

Karolina Wieserová (University of Salzburg)

*Are film subtitles more spoken or written data? Using multidimensional model to quasi-spoken texts*

Adrian Jan Zasina (Charles University)

*Kontrastivní lingvistika jako průsečík různých lingvistických oborů: aktuální výzkumné a teoreticko-metodologické trendy a výzvy (kulatý stůl)*

Viktor Elšík, Mirjam Fried, Olga Nádvorníková, Alexandr Rosen, Pavel Štichauer (Univerzita Karlova)

## Kauzativní konstrukce v italštině a španělštině a jejich české protějšky v paralelním korpusu: statistický model predikce překladové ekvivalence

Petr Čermák, Pavel Štichauer (Univerzita Karlova)

V tomto příspěvku chceme zcela nově nahlédnout v kontrastivní perspektivě jev, kterému v minulosti již byla věnována řada velmi detailních studií, totiž kauzativní konstrukce v italštině a španělštině typu *far arrivare / hacer llegar* „přivést, dovést, nechat přijet, přimět k příjezdu aj.“, v nichž vystupuje kauzativní sloveso *fare/hacer* v kombinaci s infinitivem báze slovesa (v tomto případě *arrivare/llegar* „přijet“). Dosavadní výzkum (srov. Čermák & Štichauer 2010; Čermák & Nádvorníková et al. 2015, kap. 2; Štichauer & Čermák 2016; Čermák & Nádvorníková et al. 2020, kap. 4) se totiž zatím omezil na podrobnou popisnou statistiku, založenou na kvantifikaci různě definovaných překladových protějšků, a jejich interpretaci. Tento výzkum vyšel nejprve z definice šesti překladových respondentů ilustrovaných na paralelních datech z italštiny a španělštiny. Později spolu s ambiciózním rozšířením též na francouzštinu a portugalštinu přistoupil k daleko jemnějšímu vymezení jedenácti překladových respondentů a jejich popisu v této panrománské kontrastivní perspektivě. Ve všech zmíněných studiích však chybí pokus o pokročilejší statistický model, jako jsou např. různé regresní techniky, které by volbu těchto respondentů na základě studovaných faktorů dokázaly nějak predikovat.

O takový model se chceme pokusit v tomto příspěvku, v němž se vzhledem k nesmírné pracnosti s každým získaným paralelním segmentem omezíme na italštinu a španělštinu a přidržíme se zároveň původní šestičlenné typologie, jež je pro potřeby takového modelování dostatečně robustní. Po vzoru některých podobných studií (např. Rosemeyer & Enrique-Arias 2016) použijeme multinomiální logistickou regresi k predikci jednoho ze šesti překladových protějšků v závislosti na faktorech, které nyní popíšeme.

Jak už jsme naznačili, považujeme za vhodné vrátit se k původní typologii šesti protějšků. První (A) odpovídá zjevné kauzativní prefixaci *plakat > rozplakat, sedět – usadit*, často odpovídající konstrukcím jako *far piangere, fare sedere / hacer llorar, hacer sentar*. Druhý typ (B) překladatelsky řeší tuto konstrukci pomocí prostého slovesa, jako je např. *přesvědčit* pro *far capire / hacer comprender* nebo *seznámit* pro *far conoscere / hacer conocer*. Třetí (C) odpovídá syntakticky románské konstrukci, jako je tomu např. u *dát si ostříhat vlasy* pro it. *farsi tagliare i capelli*. Čtvrtý typ (D) je podobně analytický jako (C), ale místo báze slovesa tu je jiný slovní druh a celá konstrukce je často nějakou kolokací, jako je např. *dát na vědomí, dát na srozuměnou*. Pátý typ (E) spočívá v souvětěné realizaci, v níž predikátem hlavní věty je nějaké kauzativní sloveso (*zařídít, přimět, požádat*) řídicí vedlejší větu typu *zařídít to tak, aby...* Šestý typ (F) kauzativní konstrukci v překladu eliminuje a realizuje báze sloveso jako prostý predikát s vlastní valencí a kauzativitu kóduje (pokud vůbec) jinými prostředky, jako je tomu např. v it. *ci hanno fatto sentire male – cítili jsme se (kvůli nim) špatně*. I zde počítáme – podobně jako zmíněné studie – s určitým reziduem v podobě těžko určitelných či neexistujících překladů (kódováno jako X).

Přestože je ze zmíněného výzkumu – díky velmi detailní popisné kvantifikaci – patrné, že rozhodně neexistují přímočaré faktory, které by tak či onak volbu jednoho z respondentů určovaly, lze se přece jen pokusit na základě následujících faktorů vytvořit model jejich predikce, jenž by zároveň spolehlivě vyhodnotil vliv těchto faktorů. Na prvním místě to bude samozřejmě konkrétní sloveso, jeho syntaktická povaha (tranzitivní vs. intranzitivní), povaha komplementace (abstraktní vs. konkrétní jméno, zájmeno, vedlejší věta), životnost vs. neživotnost a konečně i konkrétní překladatel (otázka, zda je volba takových protějšků ovlivněna osobním stylem překladatele, není bez zajímavosti). Samozřejmou součástí našeho příspěvku bude i zveřejnění celého datového souboru ve formátu .csv pro jakékoli další užití či modelování.

## Bibliografie

- Čermák, P., & Nádvořníková, O., et al. (2015). *Románské jazyky a čeština ve světle paralelních korpusů*. Praha: Karolinum.
- Čermák, P., & Štichauer, P. (2010). Španělské a italské kauzativní konstrukce *hacer/fare* + sloveso a jejich české ekvivalenty. In: F. Čermák, J. Koček et al. (Eds.), (pp. 70–90). Praha: Lidové noviny.
- Čermák, P., Nádvořníková, O., Kratochvílová, D., Štichauer, P. (Eds.) (2020). *Complex Words, Causatives, Verbal Periphrases and the Gerund. Romance Languages versus Czech (A Parallel Corpus-Based Study)*. Praha: Karolinum.
- Rosemeyer, M., & Enrique-Arias, A. (2016). A match made in heaven: Using parallel corpora and multinomial logistic regression to analyze the expression of possession in Old Spanish. *Language Variation and Change*, 28(3), 307–334.
- Štichauer, P., & Čermák, P. (2016). Causative constructions of the *hacer/fare*+ verb type in Spanish and Italian, and their Czech counterparts: a parallel corpus-based study. *Linguistica Pragensia*, 26(2), 7–20.

Je *viitsiä* jedinečné? Lexikálně-syntaktický popis jednoho finského modálního slovesa

Lenka Fárová (Univerzita Karlova)

Většina morfosyntaktických či lexikálních jevů je detailně popsána v deskriptivních gramatikách, učebnicích či slovnících. Existují však i takové, které se popisují méně, hůře nebo dokonce vůbec, což mj. platí o tzv. unique items, tedy lexikálních jednotkách či frazémeh, které jsou pro určitý jazyk jedinečné a v jiných jazycích postrádají jednoslovný a jasný překladový ekvivalent (viz Tirkkonen-Condit 2002, 2004). Tento příspěvek se zaměřuje na to, jak k popisu a výkladu takovýchto jednotek mohou přispět paralelní korpusy v případě typologicky odlišných jazyků (srov. Johansson S. 2004).

Zkoumání je podrobeno finské modální sloveso *viitsiä*, které bývá uváděno jako příklad jedinečného prvku (mj. Tirkkonen-Condit 2002:209), a široká škála jeho překladových ekvivalentů v češtině a angličtině. Toto sloveso reprezentuje dynamickou modalitu s významem dostatku (či spíše nedostatku) mentální energie k vykonání určité činnosti (viz např. Flint 1980, Kangasniemi 1992, Johansson M. & Nordrum L. 2021). Důrazem na mentální energii se odlišuje od frekventovanějšího modálního slovesa *jaksaa*, které akcentuje dostatek či nedostatek fyzické energie – srov. *En viitsi lähteä kurssille*. ‘Nemám dost chuti jít na kurz.’ a *En jaksaa lähteä kurssille*. ‘Nemám dost sil jít na kurz.’.

Výkladový slovník současné finštiny *Kielitoimiston sanakirja* (2022) uvádí sloveso *viitsiä* ve třech významech: 1. obtěžovat se (častěji v negativním kontextu), např. *En viitsinyt edes sängystä nousta*. ‘Neobtěžoval/a jsem se ani vstát z postele.’; 2. jako výraz znechucení, např. *Älä viitsi nalkuttaa*. ‘Neopovažuj se brblat.’ a 3. v žádostech ve formě otázky, kde je synonymem slovesa *voida* (mocht). Překladové slovníky heslo *viitsiä* často neuvádějí vůbec, a pokud jej obsahují, volí jen jeden z významů (např. Lindroos 1984 uvádí výrazy ‘(ne)dat si práci, (ne)chtít’ odpovídající významu 1.), nebo do hesla vyberou pouze izolovanou příkladovou větu (viz Särkkä 1992).

Data z paralelního korpusu InterCorp v16 však naznačují, že použití slovesa *viitsiä* je v nepřekladových textech širší, než dokládají slovníky. Jako většina modálních sloves je vázáno na A-infinitiv plnovýznamového slovesa, přesto se v daném materiálu v téměř 10 % vyskytuje také elipticky. Preferenci negativní formy slovesa potvrzuje subkorpus Core InterCorpu v16, v němž v nepřekladových finských textech tvoří kladné tvary slovesa *viitsiä* pouze 23 % výskytů.

Lexikálně-syntaktický obraz ovlivňuje i typ textu či žánr, a to zejména s ohledem na značné rozdíly mezi psanou a mluvenou finštinou. Zastoupení v jednotlivých částech InterCorpu není rovnoměrné. Zatímco v subkorpusu Bible se toto sloveso objevilo pouze jednou a v subkorpusu Europarl devětkrát, subkorpus Core obsahuje 204 výskytů (z toho 85 v nepřekladové beletrii) s i.p.m. 20,68 a v subkorpusu Subtitles je 18 000 výskytů s i.p.m. 143,85. To dokazuje, že sloveso *viitsiä* je významně frekventovanější v mluvené finštině, jejíž zastoupení je v korpusech omezené.

Vzhledem k neexistenci jednoznačného a jednoslovného překladového ekvivalentu se význam zkoumaného slovesa do češtiny i angličtiny překládá spíše na syntaktické než na lexikální úrovni. Rozostřenost jeho lexikálního obrazu dokládají mj. stovky českých i anglických ekvivalentů navrhovaných databází Treq, z nichž většina s významem nekoresponduje. Některé ekvivalenty ovšem naznačují, že se toto sloveso zejména v subkorpusu Subtitles, do češtiny ani angličtiny nepřekládá slovesně, ale partikulemi (čes. ‘no’, ‘no tak’, či ‘prosím’; angl. ‘please’, ‘come on’. Zároveň některé frekventované tvary dosáhly již určitého stupně lexikalizace, jak dokládá poměrně jednotný způsob jejich překladu, což platí např. pro eliptický záporný imperativ 2. os. sg. *Älä viitsi/viitti*. překládaný do češtiny jako ‘Nech toho.’ a do angličtiny jako ‘Give me a break.’.

Komplexní lexikálně-syntaktický popis slovesa *viitsiä* ukazuje možnosti využití paralelních korpusů a překladových ekvivalentů pro zpřesnění lexikografického popisu jedinečných prvků jak v jednojazyčných, tak i v překladových slovnících. Zároveň poukazuje i na některá metodologická úskalí při srovnávání typologicky odlišných jazyků dané rozsahem i obsahem dostupných korpusů.

## Bibliografie

- Benko, Vladimír. 2015. Araneum Finnicum Minus, verze 15.04. Praha: Ústav Českého národního korpusu FF UK. Dostupný z WWW: <http://www.korpus.cz>
- Fárová, Lenka, Vavřín, Martin, Zasina, Adrian Jan. 2023. Korpus InterCorp – finština, verze 16 z 12. 10. 2023. Praha: Ústav Českého národního korpusu FF UK. Dostupný z WWW <http://www.korpus.cz>
- Flint, Aili. 1980. Semantic Structure in the Finnish Lexicon: Verbs of Possibility and Sufficiency. Helsinki: Suomen Kirjallisuuden Seura.
- Johansson, Mats – Nordrum, Lene. 2021. Tracing processes in auxiliarization. Time-sufficiency verbs from a Norwegian-Swedish-English contrastive perspective. In Čermáková, Anna, Egan, Thomas, Hasselgård, Hilde and Sylvi Rørvik (Eds). Time in Languages, Languages in Time. Amsterdam: John Benjamins. 67-94.
- Johansson, Stig. 2007. Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies. Amsterdam: John Benjamins.
- Kangasniemi, Heikki. 1992. Modal Expressions in Finnish. Helsinki: Suomen Kirjallisuuden Seura.
- Kielitoimiston sanakirja. 2022. Kotimaisten kielten keskus ja Kielikone Oy. Dostupný z: <https://www.kielitoimistonsanakirja.fi>
- Lindroos, Hilka Annikki. 1984. Suomi-Tšekki-Suomi taskusanakirja. Porvoo: Werner Söderström Osakeyhtiö.
- Särkkä, Heikki. 1992. Englanti-suomi-englanti sanakirja. Helsinki: Otava.
- Tirkkonen-Condit, Sonja. 2002. Translationese – a myth or an empirical fact? A study into the linguistic identifiability of translated language. Target, 14(2), 207-220.
- Tirkkonen-Condit, Sonja. 2004. Unique items-over-or under-represented in translated language? In Mauranen, Anna and Pekka Kujamäki (Eds.). Translation universals: do they exist? Amsterdam: Benjamins. 177-184.

## Slovesný vid v rusko-českém kontrastivním korpusovém výzkumu

Markus Giger (Ústav východoevropských studií Filozofické fakulty Univerzity Karlovy)

Gramatická kategorie slovesného vidu je v ruštině a češtině obvykle chápána jako binární, obsahuje tedy dva gramémy, vid nedokonavý a dokonavý. Existuje řada prací, která se zabývá paralelami a rozdíly na rovině funkční (např. Stunová 1993, Petruchina 2012). Formální prostředky pro vyjadřování této kategorie jsou v obou jazycích ve velké míře (ovšem nikoliv v úplnosti) analogické a původem slovo tvorné, jde zejm. o řadu prefixů a sufixů a výjimečně o supletivní kmene.

Přes tyto paralely existuje řada rozdílů jak v inventáři vidových tvarů, tak v jejich tradiční domácí filologické interpretaci v domácí rusistice a bohemistice. Tyto rozdíly se odrážejí v korpusech včetně paralelních, často však nestojí ve fokusu zájmu (srov. Bláha 2020). Tak např. nejen NKRJa, ale fakticky ani ruská část InterCorp v16 ČNK nemá (resp. nepoužívá) značku pro biaspektuální slovesa, což vede k tomu, že v české části uvedeného korpusu je téměř 125 000 biaspektuálních sloves, zatímco v ruské části – podle taggingu – není ani jedno (viz tabulku). Jinak řečeno, v ruském korpusu je každé lexikograficky biaspektuální sloveso aktuálně přiřazeno nedok., resp. dok. vidu, kdežto v českém korpusu tvoří třetí skupinu. Lze ale připomenout i různý vývoj některých typů přejatých a původně biaspektuálních sloves: v češtině bývají přejatá verba simplicia časem interpretována jako nedokonavá a může k nim vzniknout nové sloveso dok. pomocí prefigace, srov. *organizovat<sup>1</sup>/zorganizovat<sup>P</sup>, realizovat<sup>1</sup>/zrealizovat<sup>P</sup>*. V ruštině existuje tento typ také (např. *operirovat<sup>1</sup>/prooperirovat<sup>P</sup>*, srov. Čertkova/Čang 1998), vedle toho je však tendence některá přejatá simplicia interpretovat jako dok. a vytvořit k nim pomocí sufigace nové sloveso nedok. (*organizovyvat<sup>1</sup>/organizovat<sup>P</sup>, realizovyvat<sup>1</sup>/realizovat<sup>P</sup>*); jsou to slovesa s přízvukem na sufixu -ová- (srov. Jászay 1999). Tím stojí někdy v češtině prefigované a příznakově dokonavé sloveso v opozici k jednoduchému původně biaspektuálnímu slovesu, zatímco v sémanticky analogických případech v ruštině stojí v opozici k původně biaspektuálnímu jednoduchému slovesu sufigované sloveso příznakově nedokonavé.

Známé jsou rozdíly v oblasti participiálních tvarů: v systému je asymetrie v oblasti participií minulých činných, kde si v češtině konkurují dokonavé tvary na -lý a -(v)ší (*přišlý<sup>P</sup>, přišedší<sup>P</sup>*), zatímco v ruštině je pouze tvar na -(v)šij, ovšem v obou videch (*pisavšij<sup>1</sup>, napisavšij<sup>P</sup>*); a také v oblasti participií trpných, kde je v češtině v obou videch produktivní přičestí *n-/t-ové* (*otevirán<sup>1</sup>, otevirany<sup>1</sup>//otevřen/-y<sup>P</sup>*), zatímco v ruštině je produktivní pouze *n-/t-ové* přičestí dok. (*otkryt(yj)<sup>P</sup>*, neexistuje *\*otkryvan(nyj)<sup>1</sup>*), což nevyrovnají omezené tvary nedok. vidu *n-/t-ové* (*delan(nyj)<sup>1</sup>*) a *m-ové* (*otkryvaem(yj)<sup>1</sup>*). Kromě toho jsou však zásadní rozdíly v interpretaci – tradiční bohemistika považuje všechny participiální tvary se složenou koncovkou za adjektiva a neanotuje v korpusech jejich vid, při hledání slovesných tvarů podle vidu se tudíž neobjevují, zatímco ruská participia jsou součástí odpovídajících konkordancí. Důležité rozdíly jsou i v úzu: obecně jsou participiální konstrukce v psané ruštině častější. To nemusí mít dopad na frekvenci vidově anotovaných tvarů v textu, když sklonnému přičestí odpovídá vedlejší věta (r. *pišuščij<sup>1</sup>* – č.  *který píše<sup>1</sup>/psal<sup>1</sup>/bude psát<sup>1</sup>*), někdy to ovšem dopad má (r. *pišuščij<sup>1</sup> eti stroki* – č.  *autor těchto řádků*) (srov. Kocková 2022: 46). Velmi výrazný je rozdíl u přechodníků, které v češtině jsou značně okrajové. V paralelním korpusu jako analoga ruských přechodníků nevystupují jen finitní slovesné tvary (r. *delaja<sup>1</sup>* – č.  *když dělá<sup>1</sup>/dělal<sup>1</sup>/bude dělat<sup>1</sup>* atd.), nýbrž i verbální nomina (r. *prostivšis<sup>P</sup>* – č.  *po rozloučení<sup>P</sup>*) (srov. Kocková 2022: 111, 117). Tato verbální nomina mohou v češtině vystupovat v obou videch (na rozdíl od ruštiny) a vyjádřit řadu vidových funkcí (Giger/Kocková 2024). V korpusech však ohl. vidu anotována nejsou., což má dopad na frekvenci nedok. a dok. slovesných tvarů v češtině.

V navrženém příspěvku bude podán podrobný popis dopadů uvedených rozdílů na formální kontrastivní popis vidového systému ruštiny a češtiny při jeho užívání v psaných textech. Jako materiál budou sloužit aktuální rusko-české paralelní korpusy ČNK a NKRJa.



Tvary v InterCorp v16	ruština	čeština
nedok. slovesa	12 553 890	21 549 762 (podstatnou úlohu pro frekvenci hrají mj. slovesa modální a pomocná)
<i>z toho participia se složeným tvarem</i>	133 759	-
„adjektivum odvozené od slovesného tvaru přítomného přechodníku“ (bez tagu vidu)	-	139 584
dok. slovesa	9 853 353	8 826 745
<i>z toho participia se složeným tvarem</i>	166 856	-
„adjektivum odvozené od slovesného tvaru minulého přechodníku“ (bez tagu vidu)	-	1 320
biaspektuální slovesa	-	124 686
příklady:	<i>organizovat</i> <sup>1/P</sup> (4911) <i>organizovyvat</i> <sup>1</sup> (398)	<i>organizovat</i> <sup>B</sup> (806) <i>zorganizovat</i> <sup>P</sup> (886)
	<i>realizovat</i> <sup>1/P</sup> (668) <i>realizovyvat</i> <sup>1</sup> (50)	<i>realizovat</i> <sup>B</sup> (612) <i>zrealizovat</i> <sup>P</sup> (82)

## Bibliografie

Bláha, O. 2020. Typologické aspekty ruského a českého vidového systému. *Slavia* 79, 245-255.

Čertkova, M. Ju., Čang, P.-Č. 1998. Évolucija dvuvidovych glagolov v sovremennom ruskom jazyke. *Russian Linguistics* 22, 13-34.

ČNK: Český národní korpus. [www.korpus.cz](http://www.korpus.cz)

Giger, M., Kocková, J. 2024. Grenzüberschreitungen an der Peripherie: Aspektuelle Funktionen von Aktivpartizipien und Verbalsubstantiven im Tschechischen. *Zeitschrift für Slawistik* 69, č. 1. 18 s.

Jászay, L. 1999. Vidovye korreljaty pri dvuvidovych glagolach. *Studia Russica* 17, 169-177.

Kocková, J. 2022. *Neurčité tvary slovesné v češtině, ruštině a němčině a jejich vzájemná ekvivalence*. Praha.

NKRJa: Nacional'nyj korpus ruskogo jazyka. [www.ruscorpora.ru](http://www.ruscorpora.ru)

Petruchina, E. V. 2012. *Aspektual'nye kategorii glagola v ruskom jazyke v sopostavlenii s češkim, slovackim, pol'skim i bolgarskim jazykami*. Moskva.

Stunová, A. 1993. *A contrastive study of Russian and Czech aspect. Invariance vs. discourse*. Amsterdam.

**Rozdíly v užívání pasivních konstrukcí v chorvatštině a češtině**

Karel Jirásek (Filozofická fakulta Univerzity Karlovy)

Popis gramatické struktury češtiny i chorvatštiny obsahuje na první pohled obdobné tvary pasiva a způsoby tvoření pasivních konstrukcí. Přesto zde existují jisté rozdíly ve dvou rovinách. Zaprvé je to rovina stylistická, kdy můžeme použít stejné pasivní konstrukce v jednom i druhém jazyce, ale zatímco některé jsou v jednom jazyce téměř neutrální, ve druhém jazyce jsou značně příznakové, buď archaické a omezené jen na určitou stylovou rovinu, či velmi málo používané a vyznívající nepřirozeně. Projevuje se to ve frekvenci výskytu těchto pasivních konstrukcí, která se v obou jazycích značně liší. Např. u vět typu *To je považováno...* / *To se považuje...* chorvatština výrazně preferuje formu tzv. zvrátého pasiva, respektive reflexivní formy slovesné (ch. *To se smatra...*). Ve svém příspěvku chci porovnat výskyt takovýchto konstrukcí v korpusech českého a chorvatského jazyka. Zadruhé existuje několik typů pasivních konstrukcí, které ve druhém jazyce nelze v žádném případě použít a jsou proto při překladu kompenzovány jiným způsobem. Opět jde např. o typ reflexivní formy slovesné (neosobního reflexivního pasiva) v chorvatštině (ch. *Ako vas se uhvati... Kad ga se pita... Ljude se tjera...*), který nelze v češtině použít a kterému nejlépe odpovídá překlad osobním pasivem (*Pokud budete přistiženi... Když je tázán... Lidé jsou nuceni...*). Na paralelních korpusech chci demonstrovat způsoby této kompenzace. Tam, kde to rozsah dostupných dat umožňuje, bych rovněž rád poukázal na vývoj četnosti jednotlivých typů pasivních konstrukcí v čase, tj. zda lze vysledovat nějaký trend v poklesu či nárůstu jejich užívání a zda z toho lze vyvodit nějaká doporučení pro jejich překládání mezi chorvatštinou a češtinou.

## Bibliografie

- Jirásek, K. – Vavříň, M. – Zasina, A. J.: Korpus InterCorp – chorvatština, verze 16 z 12. 10. 2023. Ústav Českého národního korpusu FF UK, Praha 2023. Dostupný z WWW <http://www.korpus.cz>
- Jirásek, K.: Dynamika pronikání adjektivních tvarů v dějovém a stavovém pasivu na počátku 21. století. (2018) In: Spisovná čeština a jazyková kultura 2018, 89-99.
- Karlík, P. – Nekula, M. – Rusínová, Z. (eds.) (1995): Příruční mluvnice češtiny. Praha: Nakladatelství Lidové noviny.
- Karlík, P. (2017): PASIVUM. In: Karlík, P. – Nekula, M. – Pleskalová, J. (eds.), CzechEncy - Nový encyklopedický slovník češtiny. URL: <https://www.czechency.org/slovník/PASIVUM> (poslední přístup: 30. 3. 2024)
- Křen, M. – Cvrček, V. – Čapka, T. – Hnátková, M. – Jelínek, T. – Kocěk, J. – Kovářiková, D. – Křivan, J. – Milička, J. – Petkevič, V. – Skoumalová, H. – Šindlerová, J. – Škrabal, M.: Korpus SYN, verze 12 z 29. 12. 2023. Ústav Českého národního korpusu FF UK, Praha 2023. Dostupný z WWW: <https://www.korpus.cz>
- Mrázková, K. (2017): Užívání pasivního participia v mluvené češtině: srovnání institucionální komunikace a běžného dorozumívání. Časopis pro moderní filologii, 99, s. 181–192.
- Štícha, F. (1990): K užívání opisného pasíva v současné češtině. NŘ 73, 1990, 63–73.
- Štícha, F. (2013): Akademická gramatika spisovné češtiny. Praha: Academia.

**Polish modal expressions and their Czech translation equivalents  
based on the parallel corpus InterCorp**

Elżbieta Kaczmarska-Zglejszewska (University of Warsaw)

The presentation focuses on selected types of modalities: epistemic modality, alethic modality, deontic modality, ability modality, volitive modality (Boniecka 1976; Rytel 1982; Palmer 1986; Palmer 2001; von Fintel 2006; Labocha 2020). Impersonal constructions in Polish and their Czech counterparts will be analyzed; specifically, it concerns expressions with units: *mieć* ‘to have’, *musieć* ‘must’, *powinien* ‘should’, *trzeba* ‘must / have to’, *wolno* ‘may’, *należy* ‘ought to’, *można* ‘can’, *chcieć* ‘to want’, *woleć* ‘prefer’, *móc* ‘to be able to’. These units are part of both lexical and grammatical means of expressing various types of modalities (Dokulil 1954). In this article, we will attempt to determine whether texts from different fields “choose” specific lexical and grammatical means to express the mentioned types of modalities. Pilot studies reveal that, for example, texts in the fields of law, marketing, and psychology utilize different linguistic means to express modality. This is associated with the problem of translation and the danger of the translator choosing inappropriate equivalents that may be considered universal (Nowak-Michalska 2011; Rybicki 2013). In the era of frequent use of artificial intelligence, this problem may be even more relevant. That's why we will also conduct an experiment using aligned pairs from the InterCorp corpus and material translated by an automatic translator. The conclusion of the study will involve determining: 1) which of the gathered means are utilized by texts representing specific types (legal, psychological, marketing, and others), and 2) whether the examined units have any specific equivalents in different types of texts. The research results will also be presented in tables.

## **Bibliography**

Boniecka, B. 1976. O pojęciu modalności (przegląd problemów badawczych). *Język Polski* LVI (2): 99-110.

Dokulil, M. K modální výstavbě věty. In *Studie a práce lingvistické* 1, 1954, 255–262.

Labocha, J. 2020. Modalność w opisie wypowiedzi językowych. In M. Szczyrbak & A. Tereszkiewicz (Eds.), *Kontakty językowe* (pp. 205–210). Wydawnictwo Uniwersytetu Jagiellońskiego.

Nowak-Michalska J. 2012. Modalność deontyczna w języku prawnym na przykładzie polskiego i hiszpańskiego kodeksu cywilnego. Poznań.

Palmer, F. R. *Mood and Modality*, 1986.

Palmer, F. R. *Mood and Modality*, 2001.

Rybicki, J. 2013. Stylometryczna niewidzialność tłumacza. *Przekładaniec* 27/2013: 61–87.

Rytel, D. 1982. *Leksykalne środki wyrażania modalności w języku czeskim i polskim*. Wrocław.

von Wright, G. H. *An Essay in Modal Logic*, 1951.

## InterCorp &amp; co. – Co se všechno (ne)stihlo

Tomáš Káňa (Masarykova univerzita)

**Východisko:** Paralelní korpus InterCorp funguje od roku 2008, a jak je známo práce v současnosti na něm nepokračují v takové míře a tempu, jak dřív, což snad neznamená jeho konec nebo ztrátu důležitosti. Dva méně známé paralelní korpusy vznikly v Česku však už o pár let před ním: Kačenka (1997) a Kačenka 2 (2002/2003) na FF MU a ČNPK (2002) na PdF MU. Všechny jsou stále funkční a daly vzniknout spoustě pracím studentským, kvalifikačním, habilitačním. Přesto zůstává práce s korpusy obecně ve výuce spíš hobby než standard (Káňa, 2023). Tento deficit je východiskem problému a otevírá prostor k závěrečné diskusi.

**Metody:** Metodou dotazování (strukturované i nestrukturované dotazníky) a vyhodnocování údajů z vysokoškolských kurikulů přicházíme na to, že spolehlivá jazyková data z korpusů nehrají v oblasti výuky (především cizích) jazyků primární roli, často ani marginální, a to i navzdory faktu, že během posledních let vznikla nejen celá řada odborných monografií (např. v řadě Studie z korpusové lingvistiky NLN) i článků relevantních pro vysokoškolskou výuku jazyků, ale i prací studentských, kvalifikačních. Kombinací kvalitativní a kvantitativní metody analýzy korpusových dat s corpus based a corpus driven (McEnery & Hardie, 2012) je možné ukázat jednotlivé jazykové jevy, a to i v kontrastu.

**Cíle:** Příspěvek shrnuje možnosti práce s paralelními (především s InterCorpem), ale i jinými korpusy ve výuce cizích jazyků v terciárním vzdělávání a ukazuje limity, které plynou jak z vlastností dat a funkčností manažerů, tak i z nedostatečné publicity a průpravy uživatelů. Úkoly v oblasti (kontrastivní) morfologie, syntaxe i lexikologie se dají zpracovávat efektivně a elegantně, navíc lze práce s korpusy pro tyto účely i celkem jednoduše a rychle vysvětlit. Naopak oblasti jazyka, kde je potřeba širší kontext (analýza diskursu / textová lingvistika) nebo roviny méně spjaté s formální/“povrchovou“ stránkou (sémantika a především pragmatika) běžné korpusové manažery nalézt nedovedou. Avšak i zde je možné data tomuto účelu anotovat (srov. Kaiser, 2023).

Cílem je ukázat roli (paralelních a srovnávacích) korpusů v rámci data driven learning (Corino & Onesti, 2019), především ve výuce (cizích) jazyků, demonstrovat na jednotlivých příkladech indukčních i dedukčních úkolů smysluplnost jejich nasazení v běžné (jazykové) výuce na různých typech škol a multiplikovat povědomí o korpusech jako demokratických nástrojích výzkumu a výuky jazyků.

Jedním ze zkoumaných jevů, které zůstávaly dosud (možná i chtěně) skryta, je problematika textotvorných prvků, tedy prostředků koheze a koherence textu (rekurence, proformy a textové deixe [deBeaugrande & Dressler, 1981]), protože většina běžně užívaných korpusů neumožňuje uživatelům (z důvodu copyrightu) získat celý text. Příklady se tedy soustředí především na tyto jevy a jejich analýzu v běžně užívaných korpusech (InterCorp, 2023 a DeReKo, 2024 jako zástupce národního korpusu jiného jazyka).

**Diskuse:** Vzhledem k tomu, že korpusy vznikaly, vznikají a existují na akademických institucích vzdělávajících k výuce jazyků, a především tam se také používají, je nasnadě se podívat, jak mohou v současné době konkurovat užívání běžně dostupných nástrojů umělé inteligence. V závěru by se měla rozvinout diskuse k těmto otázkám: Jsou data jazykových korpusů dostatečná, aby mohla podat objektivní obrázek o daném jevu? Jakou roli hraje dnes AI? Doveden nahradit výzkum na korpusových datech, a proč (ne)?

## Bibliografie

- Corino, E. & Onesti, C. (2019). Data-driven learning: A scaffolding methodology for CLIL and LSP teaching and learning. *Frontiers in Education*, 4 (7), 1–12.
- Český národní korpus: Wiki. (2023) *InterCorp*. <https://wiki.korpus.cz/doku.php/cnk:intercorp>
- Department of English and American Studies, Faculty of Arts, Masaryk University (1997). *Parallel Corpus of English and Czech Texts (kačenka)*.  
<https://www.phil.muni.cz/angl/kacenska/kachna.html>
- Department of English and American Studies, Faculty of Arts, Masaryk University (2002/2003). *The Kačenka 2 project*. <https://www.phil.muni.cz/angl/kacenska2/>
- De Beaugrande, R., & Dressler, W. U. (1981). *Einführung in die Textlinguistik*. Max Niemeyer Verlag.
- DeReKo (2024). *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache*. Leibniz-Institut für Deutsche Sprache. [www.ids-mannheim.de/dereko](http://www.ids-mannheim.de/dereko).
- Kaiser, J. (2003) ZuHand: Zugang zu Handlungssequenzen und Handlungsbezogenen Themenausschnitten in einem qualitativ annotierten FOLK-Subkorpus. *Zeitschrift Korpora Deutsch als Fremdsprache* 3 (1), 92–111.
- Káňa, T. (2023). Instrumente für ein demokratisches Sprachenlernen. Dvorecký, M., Reitbrecht, S., Sorger, B. & Schweiger, H. (eds.). *IDT 2022: \*mit.sprache.teil.haben Band 3: Sprachliche Teilhabe fördern Innovative Ansätze und Technologien in Sprachunterricht und Hochschulbildung*. Erich Schmidt Verlag GmbH & Co., 57–69.
- Katedra NJ PdF MU. (2002). *Česko-německý paralelní korpus (ČNPK)*.  
<https://knjl.ped.muni.cz/aktivity/cesko-nemecky-paralelni-korpus>
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: method, theory and practice*. Cambridge University Press.

## Srovnatelnost dat v rámci paralelních korpusů. Pasti dat

Jana Kocková (Slovanský ústav AV ČR)

Korpusy a především paralelní korpusy jsou dnes stabilní součástí kontrastivního jazykového výzkumu, kromě cíleného získávání většího množství jazykového materiálu umožňují využívat i řadu statistických údajů. Cílem tohoto příspěvku je upozornit na problematiku srovnatelnosti dat získaných z paralelních korpusů a doložit ji na příkladech z češtiny, ruštiny a němčiny.

Při kontrastivním jazykovém výzkumu, ale i v korpusové translologii (Cvrček/Chlumská 2015; Kaczmarek/Rosen 2015; Sitchinava 2018; Ten Hacken/Panocová 2020; Wiemer/Wrzesień-Kwiatkowska/Laziński 2020; Nádvorníková 2021) se z dat poskytovaných paralelními korpusy využívají především údaje o absolutní frekvenci a relativní frekvenci, tj. frekvenci vzhledem k celkové velikosti korpusu: např. i.p.m. i ARF v ČNK, i.p.m. v Ruském národním korpusu, resp. „na milion slov“ v COSMAS IDS Mannheim, „na milion“ Araneum apod. (The Russian National Corpus 2003—2023; Aranea Corpora 2020; IDS: Deutsches Referenzkorpus Mannheim 2022; Rosen/Vavřín/Zasina InterCorp ČNK 2023). Údaje o frekvenci jsou významnou komponentou kontrastivního jazykového výzkumu a umožňují zejména doložit mezijazykové rozdíly v situacích, kdy se daný jev (na různé úrovni jazykové struktury – od kombinace hlásek, přes frekvenci lemmat, gramatické kategorie, po frekvenci syntagmat) vyskytuje v různých jazycích, ale jeho četnost se zásadně liší. Typickým příkladem může být frekvence přechodníků v češtině a ruštině (srov. *вы улыбаетесь, читая это* [vy ulybajetes', čítaja eto]; *usmíváte se čta to*, viz také Kocková 2022), nebo kondicionálu minulého v češtině a konjunktivu II. v němčině (*er hätte es gesagt/ byl by to řekl*).

I přes zdánlivou jednoznačnost dat, které korpusy poskytují, se v těchto údajích odráží rozdílné lingvistické tradice, rozdílný pravopis a další faktory (srov. Stefanowitsch 2020). Těmto faktorům nebývá věnována větší pozornost i přesto, že často zásadním způsobem ovlivňují statistické údaje korpusů. Proč má české *učit* ve stejném paralelním korpusu dvakrát vyšší frekvenci než ruské *учить* [učít]? Proč má ruské *участвовать* [účastvovat'] třikrát vyšší frekvenci než české *účastnit*? Proč je dvakrát vyšší frekvence českého *vstát* než německého *aufstehen* i *aufgerstehen* dohromady? (srov. tabulka 1).

Problematika (ne)srovnatelnosti dat v mnoha případech souvisí s nestejnou lematizací a tokenizací v jednotlivých jazycích. Různá tokenizace se promítá i do celkové velikosti korpusu a ovlivňuje tak nepřímou i srovnatelnost relativních frekvencí. Například pokud v česko-německo-ruském subkorpusu InterCorp v16, tedy korpusu zahrnujícím stejné texty, srovnáme frekvenci lemmatu *Švédsko/Schweden/Швеция* [Švečija] získáme při podobné absolutní frekvenci velmi rozdílné relativní frekvence (i. p. m. v rozsahu 1,23 až 4,67, srov. tabulka 2). Odlišnosti lematizace často vycházejí z různého pojetí hranic slov (1) a různého rozsahu paradigmatu (2):

(1) Stanovení hranic slov je často ovlivněno pravopisem, například u psaní kompozit (*safari park*, *Safaripark*, *сафари-парк* [safari-park]), psaní negace (*не говорить* [ne govorit'], *nemluvit*, *nicht sprechen*) a dalších jevů. Stanovení hranic slov v korpusech také naráží na možnosti automatické anotace/tokenizace, která je problematická především u částí slov, které nemají zcela pevné místo ve větě, např. odlučitelné předpony v němčině nebo reflexivní *se/si* v češtině.

(2) Rozsah lemmatu, tedy jaké tvary jsou zahrnuty do jednoho paradigmatu slova, vychází se struktury daného jazyka, ale velmi často je i ovlivněno různými lingvistickými tradicemi, nebo teoriemi. Velmi rozdílné bývá lemma slovesa (srov. Giger 2020), ale rozdíly najdeme i u jiných slovních druhů, např. lemma adjektiv se liší zahrnutím/nezahrnutím negace, nebo různých tvarů stupňování.

Cílem příspěvku je shrnout základní faktory, které ovlivňují srovnatelnost frekvence slovních tvarů, lemmat a gramatických kategorií, a upozornit, jakým způsobem mohou ovlivnit validnost komparativního výzkumu jazyků.



<b>Jazyk: lemma</b>	<b>absolutní frekvence</b>	<b>i.p.m.</b>	<b>Jazyk: lemma</b>	<b>absolutní frekvence</b>	<b>i.p.m.</b>
CZ: <i>učít</i>	15757	30,41	RU: <i>учить</i> [učítʹ]	8060	49,43
CZ: <i>účastnit</i>	2217	4,28	RU: <i>участвовать</i> [učastvo- vatʹ]	7818	47,91
CZ: <i>vstát</i>	12794	24,69	DE: <i>aufstehen/auferstehen</i>	5533/771	24,14/3,36
CZ: <i>safari</i>	183	0,35	DE: <i>Safari</i>	108	0,47

Frekvence vybraných lemmat v česko-ruském, resp. česko-německém korpusu InterCorp v16

<b>Jazyk: lemma</b>	<b>absolutní frekvence</b>	<b>i.p.m.</b>	<b>ARF</b>
CZ: Švédsko	636	1,23	109,46
DE: Schweden	687	3	140,19
RU: Швеция [Švecija]	762	4,67	137,35

Frekvence lemmat v česko-německo-ruském subkorpusu InterCorp v16

## Bibliografie

- Aranea Corpora (2020): L. Štúr Institute of Linguistics, Slovak Academy of Sciences.  
<http://unesco.uniba.sk/>.
- Cvrček, Václav/Chlumská, Lucie (2015): Simplification in translated Czech. In: *Russian Linguistics*, 39(3), pp. 309–325.
- Giger, Markus (2020): Několik poznámek k přičestí přítomnému činnému v ruštině a češtině. In: *Lingvistika - korpus - empirie*. Vydání první. Praha: Ústav pro jazyk český, pp. 9–16.
- IDS: Deutsches Referenzkorpus Mannheim (2022): : Leibniz-Institut für Deutsche Sprache:  
<https://cosmas2.ids-mannheim.de>.
- Kaczmarška, Elžbieta/Rosen, Alexandr (2015): Jak najít optimální překlad polysémních sloves: porovnání metod formální analýzy paralelních textů. In: *ČASOPIS PRO MODERNÍ FILOLOGII*, 97(2), pp. 157–168.
- Kocková, Jana (2022): Neurčité tvary slovesné v češtině, ruštině a němčině a jejich vzájemná ekvivalence. Vydání první. Praha: Academia Slovanský ústav AV ČR (Práce Slovanského ústavu AV ČR, svazek 57).
- Nádvorníková, Olga (2021): Stylistic normalisation, convergence and cross-linguistic interference in translation: The case of the Czech transgressive. In: *Empirical studies in translation and discourse*. Berlin: Language Science Press, pp. 53–91.
- Rosen, Alexandr/Vavřín, Martin/Zasina, Adrian (2023): *Korpus InterCorp – czech, version 16 z 12*. 10. 2023. <http://www.korpus.cz>.
- Sitchinava, Dmitri (2018): Pluperfect and Pluperfect-related forms in Slavic: typological and areal properties. In: *Slavic Linguistics. XVI international Congress of Slavists*. Belgrade, August 20-27, 2018. Reports of the Russian Delegation. *XVI international Congress of Slavists*, Institute of Slavic Studies of the Russian Academy of Sciences, pp. 244–266.
- Stefanowitsch, Anatol (2020): *Corpus linguistics: A guide to the methodology*. Berlin: Language Science Press (Language Sciences 7).
- Ten Hacken, Pius/Panocová, Renáta (2020): *The Interaction of Borrowing and Word Formation*. Edinburgh University Press.
- The Russian National Corpus ([ruscorpora.ru](http://ruscorpora.ru)). 2003—2023. (2003): [ruscorpora.ru](http://ruscorpora.ru).
- Wiemer, Björn/Wrzesień-Kwiatkowska, Joanna/Łaziński, Marek (2020): Badania aspektu w językach polskim, czeskim i rosyjskim za pomocą korpusów i baz danych (pierwsze podsumowanie tematu). In: *Forum Lingwistyczne*, (7), pp. 45–58.

## České otázky s podmětem vyjádřeným zájmeny 1. a 2. osoby v zrcadle anglických překladových protějšků

Markéta Malá, Zuzana Nádraská (Univerzita Karlova)

Jak píše už Johansson (2007, s. 138), díky paralelním korpusům můžeme nejen mapovat korespondence mezi jazyky, ale také jasněji vidět vlastnosti jednotlivých jazyků. V našem příspěvku využíváme paralelní korpus *InterCorp* (verze 16, cz-en/en-cz subkorpus textů ‚Jádra‘) právě k osvětlení jazykových funkcí, které nemusí být zjevné jen při pohledu na jediný jazyk, češtinu. Problém, na který se zaměříme, jsou důvody vyjadřování podmětu zájmeny 1. a 2. osoby v českých zjišťovacích otázkách.

Pokud má přísudek v české větě formu 1. nebo 2. osoby, je informace o tom, kdo je participantem ve funkci podmětu, vyvoditelná z koncovky verba finita a absence podmětu je proto vnímána jako základní a bezpříznaková, kdežto jeho slovní vyjádření je sekundární, příznakové a zpravidla bývá gramaticky, informačně, stylisticky, rytmicky či emocionálně motivováno. Mezi nejsilnější faktory ovlivňující přítomnost pronominálního podmětu 1. a 2. osoby se řadí gramatická struktura věty a aktuální členění větné, zejména v kontextech vyjadřujících zdůraznění, protiklad či selekci. Velmi důležitou roli hraje vyjadřování postojů, emocí a emfáze (Grepl a Karlík, 1986, s. 239; Daneš a kol., 1987, s. 26; Zimová, 1994, s. 26, 67; Grepl a kol., 1995, s. 411-412; Daneš, 1999, s. 375-377; Štícha a kol., 2013, s. 392-393; Karlík, 2017; Panevová a Škrabalová, 2017). Zatímco ve větě oznamovací bývají zájmena 1. a 2. osoby „prostředkem důrazu vytykávého“ a „expresivní složka tu bývá většinou jen průvodní“, v otázkách má jejich užití povahu převážně citovou (Grepl, 1967, s. 79). Bývá jich užito v počáteční pozici v otázkách nepravých (podivových, dubitativních), kde je záměr mluvčího získat informaci upozaděn a jejich výskyt je spojován s vyjadřováním emocí a postojů, zvláště údivu a překvapení nad neočekávaným faktem (Daneš a kol., 1987, s. 333; Zimová, 1994, s. 67; Grepl a kol., 1995, s. 412, 609; Štícha, 1995, s. 204-205; Štícha a kol., 2013, s. 393).

Popis funkce českých otázek s vyjádřeným podmětem v 1. a 2. osobě má několik úskalí. Jedním je již zmíněná motivovanost a příznakovost jejich užití. Dalším je plynulý funkční přechod mezi jednotlivými typy otázek: otázky lze charakterizovat jako „směs komunikačních funkcí“, tzn. směs „ověřování, námítky, podivu atp.“ (Štícha a kol., 2013, s. 763-764). Obtíž lze spatřovat i v tom, že i přes množství jazykových prostředků uplatňujících se při realizaci otázek „není ani jeden z nich plně explicitním výrazem některé z množiny rozlišitelných komunikačních funkcí“ (Štícha a kol., 2013, s. 762).

V našem hledání komunikačních funkcí otázek se zájmeny 1. a 2. osoby a jejich indikátorů se opíráme o metodu „sémantických zrcadel“ (Dyvik, 2011) a předpokládáme, že podobně jako se překrývají překladové protějšky sémanticky blízkých slov, budou se v překladu překrývat i protějšky struktur s obdobnou funkcí (Malá, 2013). Angličtinu používáme jako ‚zrcadlo‘, díky kterému můžeme zobrazit funkce zkoumaných českých otázek.

(1) Ty se eště ptáš? (Škvorecký)

You mean you have to ask?

(2) You put that on my conscience? (Clancy)

Tohle chcete, abych měl na svědomí?

Výzkum zahrnuje dva kroky. Nejprve jsme identifikovaly centrální opakující se anglické překladové protějšky českých zjišťovacích otázek se zájmeny 1. a 2. osoby. Jsou to zejména otázky s formou věty oznamovací (příklad 1), v nichž se často objevují výrazy vyjadřující kontextovou zapojenost a dialogičnost, postoje, emoce, kognitivní procesy nebo modalitu (srov. Dušková a kol., 2006). Tyto anglické protějšky mohou nést stejné funkce jako české otázky v 1. a 2. osobě a sloužit tedy jako indikátory daných funkcí. Toho využíváme ve druhém kroku, kdy sledujeme české protějšky anglických otázek s formou věty oznamovací (příklad 2). Vedle zjišťovacích otázek se zájmeny 1. a 2. osoby jsou mezi nimi také např. otázky se slovesy mluvení a myšlení, postojové ukazatele, otázky se subjektivním slovosledem apod. V takovémto ‚shluku‘ různých forem s podobnou komunikační funkcí vyvstávají do popředí podstatné vlastnosti českých otázek s podmětem vyjádřeným zájmeny 1. a 2. osoby.

## Bibliografie

- DANEŠ, F. (1999): *Jazyk a text I. Výbor z lingvistického díla Františka Daneše. Část 2.* Praha: Univerzita Karlova v Praze – Filozofická fakulta.
- DANEŠ, F. – GREPL, M. – HLAVSA, Z. (eds.) (1987): *Mluvnice češtiny 3. Skladba.* Praha: Univerzita Jana Evangelisty Purkyně.
- DUŠKOVÁ, L. a kol. (2006 [1988]): *Mluvnice současné angličtiny na pozadí češtiny.* Praha: Academia. Dostupné online z <https://mluvniceanglictiny.cz/>
- DYVIK, H. (2011): *Semantic Mirrors.* Dostupné online z <https://clara.w.uib.no/files/2011/06/semmirrors.pdf>
- GREPL, M. (1967): *Emocionálně motivované aktualizace v syntaktické struktuře výpovědi.* Brno: Státní pedagogické nakladatelství.
- GREPL, M. – KARLÍK, P. (1986): *Skladba spisovné češtiny.* Praha: Státní pedagogické nakladatelství.
- GREPL, M. – KARLÍK, P. – NEKULA, M. (1995): Syntax. In: M. GREPL a kol. (eds.), *Příruční mluvnice češtiny.* Praha: Nakladatelství Lidové Noviny, s. 369-698.
- JOHANSSON, S. (2007) *Seeing through Multilingual Corpora. On the Use of Corpora in Contrastive Studies.* Amsterdam/Philadelphia: John Benjamins.
- KARLÍK, P. (2017): PODMĚT. In: P. KARLÍK – M. NEKULA – J. PLESKALOVÁ (eds.), *CzechEncy - Nový encyklopedický slovník češtiny.* Dostupné online z <https://www.czechency.org/slovník/PODMĚT> (poslední přístup: 10. 3. 2024).
- MALÁ, M. (2013): Translation counterparts as markers of meaning: The case of English copular verbs in a parallel English-Czech corpus. *Languages in Contrast* 13, 2, s. 170-192.
- PANEVOVÁ, J. – ŠKRABALOVÁ, H. G. (2017): ELIPSA. In: P. KARLÍK – M. NEKULA – J. PLESKALOVÁ (eds.), *CzechEncy - Nový encyklopedický slovník češtiny.* Dostupné online z <https://www.czechency.org/slovník/ELIPSA>
- ŠTÍCHA, F. (1995): Otázky predikátové: inference, implicitnost a explicitní výrazy ilokučních funkcí. *Slovo a slovesnost* 56, s. 98-109, 204-218.
- ŠTÍCHA, F. a kol. (2013): *Akademická gramatika spisovné češtiny.* Praha: Academia.
- ZIMOVÁ, L. (1994): *Způsoby vyjadřování větných členů v textu: Konkurence pojmenování, pronominalizace a elize.* Acta Universitatis Purkynianae 5 – Studia Linguistica. Ústí nad Labem: Univerzita Jana Evangelisty Purkyně.

## Korpus

- ROSEN, A. – VAVŘÍN, M. – ZASINA, A.J. (2023): *Korpus InterCorp – čeština, verze 16 z 12.10. 2023.* Praha: Ústav

## České a polské modální výrazy – výzkum na materiálu InterCorpu

Karolína Skwarska (Slovanský ústav AV ČR)

Modalitě a modálním výrazům ve slovanských jazycích je věnováno mnoho pozornosti přibližně od 60. let 20. století (srov. např. [2]). Studium této problematiky a ověřování dřívějších závěrů na materiálu korpusu znamená novou výzvu již proto, že modalitu věty jako celku lze jen velmi těžko značkovat. Vyšší stupeň bádání pak představuje výzkum kontrastivní.

Základní výzkum se může odvíjet od lexikálních jednotek – modálních výrazů, které je možno celkem jednoduše vyhledávat. V předchozím výzkumu [5] jsme tak vyšli od lemmatu *powinien* a v InterCorpu v15 zkoumali jeho české ekvivalenty. Výsledkem byla široká škála českých ekvivalentů odpovídající různým modálním odstínům polského výrazu (srov. též [4], naše výsledky se však v určitém ohledu od práce [4] lišily). Mezi ekvivalenty lemmatu *powinien* patří v češtině nejen modální lexikální jednotky, ale také gramatické formy českého plnovýznamového slovesa (např. imperativ).

Bylo by jistě možné v kontrastivním duchu zkoumat také modalitu gramatických konstrukcí – např. konstrukcí s infinitivem plnovýznamových sloves. Takový výzkum by však byl ještě pracnější a prozatím ho necháváme stranou.

Cílem našeho příspěvku bude popsat z kontrastivního hlediska polské lexikální jednotky *trzeba*, *należy*, *wolno* a jejich české ekvivalenty na materiálu InterCorpu v16. Zaměříme se i na záporné jednotky *nie trzeba*, *nie należy*, *nie wolno* (srov. např. [3]). Je zřejmý rozdíl v gramatické struktuře takových vyjádření (povrchové vyjádření tvaru slovesa *być/być*), který na první pohled ukazuje na některé odlišné typologické rysy obou jazyků (srov. např. [1]).

Chtěli bychom také vyzkoušet, jaké výsledky přinese vyhledávání polských ekvivalentů k českému stejně znějícímu predikátu *je třeba* a jeho negované podobě *není třeba*.

Referát by měl přispět nejen k bližšímu poznání modálních výrazů v polštině a češtině, které je zásadní např. při překladech či ve výuce jazyka, ale také k metodologii práce s paralelními korpusy.

## Bibliografie

- [1] Lotko, Eduard. 2009. *Srovnávací a bohemistické studie*. Olomouc: Univerzita Palackého v Olomouci, Filozofická fakulta.
- [2] *Otázky slovanské syntaxe III*. Brno: Universita J. E. Purkyně 1973.
- [3] Panevová, Jarmila, Eva Benešová, Petr Sgall 1982. *Čas a modalita v češtině*. Praha: Univerzita Karlova.
- [4] Rytel, Danuta. 1982. *Leksykalne środki wyrażania modalności w języku czeskim i polskim*. Wrocław et al.: Zakład Narodowy im. Ossolińskich. Wydawnictwo PAN.
- [5] Skwarska, Karolína 2023. *Polské powinen a jeho české ekvivalenty*. Referát na zasedání Komise pro gramatickou stavbu slovanských jazyků při MKS.

## Specifika překladového jazyka v souvislosti se slovesy pohybu na příkladu češtiny a španělštiny

Petr Šlechta (Ostravská univerzita)

L. Talmy (1985, 1991) navrhl sémantickou typologii jazyků založenou na způsobu, jakým daný jazyk pracuje s pohybovými událostmi. Podle této typologie dělíme jazyky na satelitně rámcující (*satellite-framed*) a slovesně rámcující (*verb-framed*) v závislosti na tom, zda je sémantický komponent CESTA (PATH) vyjádřen pomocí slovesného kořene (např. šp. *salir*, „dostat se pohybem ven“), nebo pomocí tzv. satelitu, tedy prvku vyskytujícího se v blízkosti jádrového elementu, který poskytuje o události dodatečné informace jako umístění, způsob pohybu, cíl atd. (např. v českém *vyběhnout* předpona vy- udává směr ven). Pro jazyky první skupiny (např. angličtina, čeština) je typické, že disponují širokým repertoárem sloves kódujících způsob pohybu (*kráčet, cválat, běhat, plazit se ...*) a mluvčí pohybové události popisují dynamičtěji, zatímco u slovesně rámcujících jazyků je vyjadřování pohybových událostí spíše statické a inventář sloves se způsobovou složkou na kořeni je chudší (Slobin 1996, 2004). Aske (1989) dále poukázal na to, že slovesně rámcující jazyky v situacích, kdy dochází k překročení hranice (*boundary-crossing events*), mají omezené možnosti vyjadřování ZPŮSOBU (MANNER):

čeština:	<i>Včela</i>	<u><i>vyletěla</i></u>	<i>z úlu.</i>
angličtina:	<i>The bee</i>	<u><i>flew out</i></u>	<i>of the hive.</i>
španělština:	<i>*La abeja</i>	<u><i>voló</i></u>	<i>de la colmena.</i>
		letět.PFV.PST.3SG	
	<i>La abeja</i>	<u><i>salió (volando)</i></u>	<i>de la colmena.</i>
		pohyb ven.PFV.PST.3SG (letět.GER)	

V příspěvku se zabýváme právě tímto typem pohybových událostí, neboť při jejich překladu může docházet k posunům významů překládaného textu a ztrátě nebo naopak doplnění informací (srov. Filipović, 2008). Jednou z možností, jak ZPŮSOB ve španělštině zachovat, je využití konstrukce slovesa následovaného tzv. způsobovým gerundiem (např. *vyletět – salir volando*), které je však méně obvyklým fakultativním prvkem.

Na základě dat z českého a španělského jádrového segmentu paralelního korpusu InterCorp (Rosen, Vavřín & Zasina, 2020; Čermák & Vavřín, 2020) složeného z literárních textů jsme identifikovali španělské překladové protějšky pěti českých sloves označujících nejběžnější způsoby pohybu směrem ven: *vyjít/vycházet, vyjet/vyjíždět, vylézt/vylézat, vyběhnout/vybíhat* a *vyletět/vylétat*. Ačkoli ve více než 70 % příkladů byly tyto výrazy do španělštiny přeloženy pomocí kanonického *salir*, u dynamičtějších pohybových událostí (*vyběhnout* a *vyletět*) bylo využití způsobového gerundia častější (*salir corriendo, salir volando*). Porovnáním výskytu těchto výrazů v překladových a nepřekladových textech pak bylo zjištěno, že konstrukce s gerundii *corriendo* a *volando* jsou častější v překladových textech než ve španělských originálech.

## Bibliografie

Aske, J. (1989). Path predicates in English and Spanish: A closer look. *Annual Meeting of the Berkeley Linguistics Society*, 15(1), 1–14.

Čermák, P. & Vavřín, M. (2020). *Korpus InterCorp – španělština, verze 13*. Ústav Českého národního korpusu FF UK. <http://www.korpus.cz>

Filipović, L. (2008). Typology in action: applying typological insights in the study of translation. *International Journal of Applied Linguistics*, 18, 23–40. <http://dx.doi.org/10.1111/j.1473-4192.2008.00189.x>

Rosen, A., Vavřín, M. & Zasina, A. J. (2020). *Korpus InterCorp, verze 13*. Ústav Českého národního korpusu FF UK. <http://www.korpus.cz>

Slobin, D. I. (1996). From “thought and language” to “thinking for speaking”. In J. J. Gumperz & S. C. Levinson (eds.), *Rethinking linguistic relativity* (s. 70–96). Cambridge University Press.

Slobin, D. I. (2004). The many ways to search for a frog: linguistic typology and the expression of motion events. In S. Strömquist & L. Verhoeven (eds.), *Relating events in narrative, Vol. 2. Typological and contextual perspectives* (s. 219–257). Erlbaum.

Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In T. Shopen (ed.), *Grammatical categories and the lexicon. Language typology and syntactic description* (s. 57–149). Cambridge University Press.

Talmy, L. (1991). Path to realization: A typology of event conflation. *Annual Meeting of the Berkeley Linguistics Society*, 17(1), 480–520. <https://doi.org/10.3765/bls.v17i0.1620> Praha: Academia

## Sémantická a gramatická životnost v češtině a polštině

Karolina Wieserová (University of Salzburg)

Gramatickou kategorií životnosti, která rozlišuje mezi životnými a neživotnými podstatnými jmény, lze nalézt v češtině a ve většině slovanských jazyků. Zakladně se provádí sémantické rozdělení na živé a neživé podstatná jména, které se v singuláru projevuje synkretismem genitivu a akuzativu. V plurálu je tato distribuce rozdělena odlišně v jednotlivých slovanských jazycích [2, 7 a další]. Množné číslo však není v zájmu výzkumu.

Pro polštinu již existují pokusy vysvětlit tendence tohoto jevu [6]. Počet těchto příkladů je zejména ve západoslovanských jazycích vysoký. Rovněž pozornost na výjimky v rámci této kategorie je věnována v různých lingvistických pracích na toto téma [6, 8, 9, 10, 11 a další]. Moje studie se zaměřuje na neživotné objekty v češtině, které mají genitivně-akuzativní synkretismus, jenž by měl být podle očekávání vyhrazen pro předměty životné. Západoslovanské jazyky však mají tendenci toto rozlišení v akuzativním pádě některých podstatných jmen nečinít a označovat neživé předměty také genitivní koncovkou -a [8]. V příkladu *nastavil(a) jsem budík* by měl být podle gramatických pravidel použit nominativní akuzativní synkretismus. Čeští rodilí mluvčí však mají tendenci používat genitivní akuzativní synkretismus. Místo tvaru *nastavil(a) jsem budík* se používá tvar *nastavil(a) jsem budíka*. V akuzativu se setkáváme s genitivní koncovkou -a .

Rozdíl mezi životnými a neživotnými objekty se vyskytuje i v jiných světových jazycích a najdeme v nich i výjimky. García García [5] například předkládá studii o španělštině. Pomocí korpusu analyzoval různé charakteristiky jako např. určitost, pozici objektu a sémantiku slovesa, které mohou být zodpovědné za výjimky.

Provádím korpusovou studii, která pomáhá odhalit tendence v užívání genitivně-akuzativního synkretismu s neživými předměty. Pomocí korpusové studie byly analyzovány relevantní údaje podle různých kritérií. Nalezené důkazy jsou zkoumány z hlediska vlastností, jako je agentivita, etymologie, sémantika slovesa nebo pozice objektu. Tyto faktory byly vybrány na základě předchozích studií. Například nejrelevantnějším kritériem pro výjimky z životnosti v polštině je původ substantiv: Neživá vypůjčená substantiva mají statisticky častěji tendenci k akuzativnímu synkretismu v genitivu než neživá zděděná substantiva [6]. Naproti tomu ve španělštině hraje ústřední roli slovesná sémantika [5]. Z hlediska diferenčního označování objektů může být určitost dále zodpovědná za nepravdivost v označování podstatných jmen v akuzativu [1].

Působení kategorie životnosti v akuzativu singuláru je také vnímáno jako druh diferenčního označování objektu [4]. Jazykový fenomén diferenčního označování objektů se vyskytuje ve více než 300 jazycích světa [3]. V češtině je omezeno v rámci rodu a čísla. V souladu s tím je obtížné jej v češtině koncipovat jako univerzální jev. Pokud však stejná kritéria, která jsou považována za odpovědná za diferenční označování objektů, hrají roli i zde, byl by to argument ve prospěch zařazení češtiny mezi jazyky s diferenčním označováním objektů.

Polština je navíc ve srovnání s češtinou zajímavá tím, že se v ní vyskytují různé stupně gramatikalizace tohoto jevu [8]. U některých sémanticky neživotných objektů je genitivní akuzativní synkretismus již gramatikalizován a ve výuce cizích jazyků jsou považovány za výjimky z kategorie životnosti. Například *Jem banana*. Použití v akuzativu nominativu je vyloučeno ve spojení se slovesy *jíst + zelenina* a *odřůdy ovoce*. Jiné naopak mohou být použity v polštině i češtině jak v akuzativu genitivu, tak v akuzativu nominativu. Existují doklady na použití genitivu akuzativu synkretismu s podstatným jménem *mail*: *Poslal(a) jsem ti maila*. Samozřejmě by byl akceptován nominativní akuzativ paralelně: *Poslal(a) jsem ti mail*. Totéž platí pro polštinu: *Wystarczy wysłać maila* vs. *Wystarczy wysłać mail*.

Jak víme z četných studií ve výzkumné literatuře, vliv mohou mít různé faktory [5, 6, 8, 9, 10, 11]. Ty jsou analyzovány v polském a českém jazyce a následně porovnány. Jsou za tento specifický jev v obou jazycích zodpovědné stejné faktory?



## Bibliografie

- [1] Aissen, Judith. 2003. Differential object marking: Iconity vs. Economy. *Natural Language & Linguistic Theory* 21, 435–483.
- [2] Berger, Tilman. 2005. Nominale Klassifikationssysteme II: Belebtheits-Oppositionen. In David A. Cruse and Franz Hundsnurscher et al. (eds), *Lexikologie/Lexicology. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen 2*, 995–998. Berlin; New York: De Gruyter Mouton.
- [3] Bossong, Georg. 1985. *Empirische Universalienforschung. Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Narr (= Ars linguistica, 14).
- [4] Bossong, Georg. 1998. Le marquage différentiel de l'objet dans les langues d'Europe. In Jack Feuillet (ed.), *Actance et valence*, 193–258. Berlin: de Gruyter.
- [5] García García, Marco. 2014. *Differentielle Objektmarkierung bei unbelebten Objekten im Spanischen*. Berlin; Boston: De Gruyter (= Linguistische Arbeiten, 545).
- [6] Kiklewicz, Aleksander. 2023. „Niesforny” biernik rzeczowników męskonieżywotnych w liczbie pojedynczej: o ekspansji tego zjawiska we współczesnym języku polskim i jego czynnikach. *Język Polski*, 103(2), 5–25.
- [7] Klenin, Emily. 2009. Animacy, Personhood. In Sebastian Kempgen u.a. (eds), *Die slavischen Sprachen. Ein internationales Handbuch ihrer Struktur, ihrer Geschichte und Erforschung Halbband 1*, 152–161. Berlin; New York: De Gruyter Mouton (= Handbücher zur Sprach- und Kommunikationswissenschaft, 32/1).
- [8] Kosta, Peter. 2003. The New Animacy Category in Slavic Languages: Open Questions of Syntax, Semantics and Morphology. *Germanoslavica: Zeitschrift für germano-slawische Studien* 9, 179–198.
- [9] Scharnberg, Jessica. 2006. *Wörter in freier Wildbahn. Untersuchung zum Verhältnis von Grammatik und Lexik anhand der russischen substantivischen Kategorien Numerus, Genus und Belebtheit*. Hamburg: Dr. Kovač.
- [10] Swarska, Karolina. 2018. Zapnul už jsi toho mobíla? O akuzativu singuláru neživotných maskulin ve slovanských jazycích. *Slavistika* 22(1), 169–180.
- [11] Vondráček, Miloslav. 2011. Životnost. In Štícha, František et al. (ed.)

**Are film subtitles more spoken or written data?  
Using multidimensional model to quasi-spoken texts**

Adrian Jan Zasina (Charles University)

Parallel corpora provide a vast repository of translation equivalents across multiple languages. While predominantly focused on fiction, they also encompass press materials and other sources, including film subtitles, available in numerous translations. Film subtitles offer a unique dataset within parallel corpora, straddling the boundary between written and spoken registers [1, pp. 380–381]. They offer intriguing insights into vocabulary from spoken registers that are not readily available in traditional written texts. Furthermore, subtitles remain current with the introduction of new phrases into a language. However, empirical evidence confirming their ability to capture spoken register is lacking, leading to their classification as quasi-spoken language in some instances. The specific characteristics of this text type remain ambiguous. Therefore, this study seeks to elucidate the characteristics of subtitles using an additive multidimensional approach, as proposed by Berber Sardinha et al. [2, p. 165], allowing for the projection of data onto a model of Czech register variation [3]. Register variation is a key to reveal the shape of subtitles stemming from its linguistics characteristics.

The research draws upon a collection of subtitles from the parallel corpus InterCorp [4], sourced from its 16th release. Multidimensional analysis was employed to analyse the dataset. Initially, the subtitles were annotated in a manner consistent with the original Czech MDA [3], with 122 language features applied to them. The operationalization results were subjected to factor analysis to project specific texts onto the eight-dimensional space of the Czech model of register variation.

The findings demonstrate that film subtitles represent a distinct text type capable of simulating spoken communication. Across the majority of examples, particularly within the Czech subtitles division, the data confirm that film subtitles are characterized by language features typical of informal spoken communication.

## Bibliography

- [1] Zabawa, Marcin. 2009. A Corpus of Spoken Language and Its Usefulness in the Research on Language Contact. In Jana Levická & Radovan Garabík (eds.), *NLP, Corpus Linguistics, Corpus Based Grammar Research*, 377–388. Bratislava: Slovak National Corpus. [http://kassiopeia.juls.savba.sk/~slovko/2009/Proceedings\\_Slovko\\_2009.pdf#page=375](http://kassiopeia.juls.savba.sk/~slovko/2009/Proceedings_Slovko_2009.pdf#page=375).
- [2] Berber Sardinha, Tony, Marcia Veirano Pinto, Cristina Mayer, Maria Carolina Zuppari & Carlos Henrique Kauffmann. 2019. Adding registers to a previous multi-dimensional analysis. In Tony Berber Sardinha & Marcia Veirano Pinto (eds.), *Multi-dimensional analysis: Research methods and current issues*, 165–186. London: Bloomsbury Publishing.
- [3] Cvrček, Václav, Zuzana Laubeová, David Lukeš, Petra Poukarová, Anna Řehořková & Adrian Jan Zasina. 2020. *Registry v češtině*. Praha: Nakladatelství Lidové noviny.
- [4] F. Čermák and A. Rosen, ‘The case of InterCorp, a multilingual parallel corpus’, *International Journal of Corpus Linguistics*, vol. 17, no. 3, pp. 411–427, 2012 Čermák, František & Alexandr Rosen. 2012. The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics* 17(3). 411–427.

## Diskuse u kulatého stolu

### **Kontrastivní lingvistika jako průsečík různých lingvistických oborů: aktuální výzkumné a teoreticko-metodologické trendy a výzvy**

Viktor Elšík, Mirjam Fried, Olga Nádvořníková, Alexandr Rosen, Pavel Štichauer  
(Filozofická fakulta Univerzity Karlovy)

Cílem tohoto setkání u kulatého stolu je otevřít diskusi o teoreticko-metodologických otázkách kontrastivní lingvistiky a lingvistických oborů, které ji mohou inspirovat – nebo pro které může být kontrastivní přístup sám inspirací. Diskuse je zároveň pozvánkou na mezinárodní konferenci ICLC-11 (*International Contrastive Linguistics Conference*), která se bude v září roku 2025 konat v Praze (<https://iclc11.ff.cuni.cz/>) a která je právě takovou platformou pro interdisciplinární dialog.

Diskusní témata nejprve krátce otevřou čtyři panelisté a do následné diskuse se mohou zapojit také účastníci konference z publika – přímo nebo pomocí aplikace *slido*.

Panelisté:

1. Mirjam Fried (Ústav obecné lingvistiky) – kognitivní lingvistika a konstrukční gramatika
2. Viktor Elšík (Ústav obecné lingvistiky) – jazyková typologie
3. Alexandr Rosen (Ústav Českého národního korpusu) – mnohojazyčné korpusy a jednotné schéma značkování podle Universal Dependencies
4. Pavel Štichauer (Ústav románských studií) – kontrastivní lingvistika

Diskusi uvede a bude moderovat Olga Nádvořníková (Ústav románských studií).

18

## *Lingvistická anotace korpusů*

organizují

Hana Skoumalová

Marie Mikulová

Barbora Štěpánková

(CZE)

# Popis panelu

Panel se zaměří na všechny aspekty lingvistické anotace korpusů, a to jak z pohledu teoretického, tak praktického. Bude se na jedné straně zabývat principy a metodologií při vytváření anotační koncepce, uplatněním lingvistických teorií při tvorbě korpusů s ohledem na jejich cíl. Na druhé straně budou předmětem panelu otázky spojené s vytvářením vhodných anotačních nástrojů, jejich univerzálností, anotátorskou a uživatelskou vstřícností a s různými možnostmi jejich využití. Představeny mohou být konkrétní anotační nástroje a jazykové korpusy i příklady obohacování lingvistické teorie na základě řešení praktických problémů anotace.

Vzhledem k tomu, že se jedná o zastřešující téma, které se v současné době neomezuje jen na oblast korpusové lingvistiky, očekáváme příspěvky napříč všemi lingvistickými obory, zaměřené na anotaci jednotlivých jazykových rovin (od fonetiky a morfologie až po diskurz a pragmatiku), využívající jednojazyčná i paralelní data a zohledňující synchronní i diachronní přístup.

Příspěvky by měly popisovat původní práci (probíhající nebo dokončenou, spíše než zamýšlenou) na následující témata:

- role lingvistických teorií ve vývoji korpusů;
- využití anotovaných jazykových korpusů pro lingvistický výzkum;
- principy, metodologie a cíle lingvistické anotace;
- představení nebo aktualizace jazykového korpusu;
- nástroje pro tvorbu a spravování jazykových korpusů;
- hodnocení a kontrola kvality anotace;
- standardizace anotací, univerzální kategorie (např. UD);
- využití jazykových korpusů pro aplikace NLP;
- specializované korpusy (akviziční, paralelní, mluvené, diachronní apod.)
- co je to lingvistická anotace, co všechno je možné anotovat.

Zpracování přirozeného jazyka (NLP) se v poslední době významně posouvá vpřed. Velké jazykové modely (LLM) nevyužívají ke svému natrénování lingvisticky anotovaná data. Zdá se, že budování anotovaných korpusů tak přichází o jedno z možných využití. V panelu bychom rádi reagovali i na tento současný vývoj, speciální pozornost bude věnována příspěvkům na aktuální témata:

- lingvistická anotace korpusů a AI/velké jazykové modely
- budoucnost lingvistické anotace

U příspěvků bude posuzována jejich originalita, význam a relevance pro panel. Součástí panelu budou dvě zvané přednášky, jednu na téma *30 let budování korpusů a co dál* přednese Jan Hajič z Ústavu formální a aplikované lingvistiky (MFF UK), druhou na téma *Anotace korpusů a AI* přednese Pavel Rychlý z Centra zpracování přirozeného jazyka (FI MU). Autoři přijatých abstraktů budou mít možnost své příspěvky publikovat v 30. svazku edice *Studie z korpusové lingvistiky* v nakladatelství NLN.

## Literatura

*Proceedings of the 17th Linguistic Annotation Workshop (LAW-XVII)*. (Prange, J. – Friedrich, A.; eds.) Association for Computational Linguistics, Toronto, 2023. <https://aclanthology.org/2023.law-1.0>

*Proceedings of the 16th Linguistic Annotation Workshop (LAW-XVI)*. (Pradhan, S. – Kuebler, S.; eds.) European Language Resources Association, Marseille, 2022. <https://aclanthology.org/2022.law-1.0>

*Handbook on Linguistic Annotation*. (Ide, N. – Pustejovsky, J.; eds.). Springer Science+Business Media, Dordrecht, Netherlands. 2017. [https://doi.org/10.1007/978-94-024-0881-2\\_1](https://doi.org/10.1007/978-94-024-0881-2_1)

*Korpusová lingvistika Praha 2011*. 1 *InterCorp* (Čermák, Fr.; ed.). 2 *Výzkum a výstavba korpusů* (Čermák, Fr.; ed.). 3 *Gramatika a značkování korpusů* (Petkevič, V. – Rosen, A.; eds.). Lidové noviny, Praha, 2011.

# Obsah

*Lematizácia a morfosyntaktická anotácia korpusu metódou „miešaného zboru“*

Vladimír Benko (Slovenská akadémia vied a Univerzita Komenského v Bratislave)

*Morfologická defektivita and nevyváženost paradigmatu u českých substantiv, adjektiv a verb*

Dominika Kovářiková (Univerzita Karlova)

*Vícerovinný popis jazyka: Příklad částic*

Marie Mikulová (Univerzita Karlova), Barbora Štěpánková (Univerzita Karlova)

*Analýza mezianotátorských neshod jako cesta k interpretaci textu*

Anna Nedoluzhko, Šárka Zikánová, Eva Hajičová, Jiří Mírovský (Univerzita Karlova, ÚFAL MFF)

*Synkretismus českých adjektiv a jejich značení v korpusech: Několik poznámek k Minervině zradě*

Klára Osolsobě, Hana Žižková (Masarykova univerzita)

*Hierarchizace a granularita sémantické klasifikace cirkumstantů na příkladu způsobových určení*

Jarmila Panevová, Marie Mikulová (Univerzita Karlova)

*Korpusy DIA a jejich využití*

Klára Pivoňková, Martin Stluka (Univerzita Karlova)

*Anotace korpusů a AI*

Pavel Rychlý (Masarykova Univerzita, Fakulta informatiky)

*Morfologická anotace korpusu rané češtiny Chroma*

Jakub Sláma (Ústav pro jazyk český AV ČR, Praha), Anna Chromá, Klára Matiasovitsová (Filozofická fakulta UK, Praha)

*Základní argumenty ve světových jazycích optikou Universal Dependencies*

Daniel Zeman (Univerzita Karlova, Matematicko-fyzikální fakulta)



## Lematizácia a morfosyntaktická anotácia korpusu metódou „miešaného zboru“

Vladimír Benko (Slovenská akadémia vied a Univerzita Komenského v Bratislave)

Ako „zborové“ („ensemble“) sa v počítačovej lingvistike označujú metódy, pri ktorých sa určitá operácia uskutočňuje pomocou niekoľkých nezávislých nástrojov, pričom sa očakáva, že agregáciou ich výstupov sa docieli celkové zlepšenie výsledku. V kontexte lingvistickej anotácie korpusových dát potom hovoríme o „zborovej lematizácii“ („ensemble lemmatization“), prípadne „zborovom tagovaní“ („ensemble (PoS) tagging“). Využíva sa pri tom skutočnosť, že chyby, ktoré vykazujú jednotlivé nástroje sa obyčajne nevyskytujú na rovnakých pozíciách v texte.

V ideálnom prípade, ak sú k dispozícii viac než tri nástroje, sa agregácia dá uskutočniť aj jednoduchou metódou „hlasovania“, najmä ak sú jednotlivé nástroje kompatibilné, čiže používajú rovnakú množinu morfológických značiek („tagset“). Gell’Oretta [4] napríklad opisuje požitie šiestich takýchto tagerov pre taliančinu, kde agregovaný výsledok bol lepší ako výsledok ktoréhokoľvek z nich.

Pre mnohé jazyky je však situácia nie až taká priaznivá – existuje (prípadne je k dispozícii) menší počet nástrojov, pričom tieto produkujú nie celkom kompatibilné výsledky. Typickým prípadom je možnosť použitia tagerov *TreeTagger*<sup>13</sup> [8] a *UDPipe*<sup>14</sup> [10] a lematizátora *CSTlemma*<sup>15</sup> [5], pre ktoré existujú natréňované modely pre veľké množstvo jazykov<sup>16,17</sup>. Ich správanie sumarizuje Tabuľka 1.

**Tabuľka 5.** Činnosť jednotlivých nástrojov vzhľadom na prítomnosť slovného tvaru v lexikóne

slovný tvar	<i>TreeTagger</i>	<i>UDPipe</i>	<i>CSTlemma</i>
nájdenný v lexikóne	morfológická značka pridelená	morfológická značka odhadnutá	--
mimo lexikónu	morfológická značka odhadnutá	morfológická značka odhadnutá	--
nájdenný v lexikóne	lema pridelená	lema odhadnutá	lema pridelená <sup>18</sup>
mimo lexikónu	lema označená ako nerozpoznaná	lema odhadnutá	lema odhadnutá

V našom príspevku opíšeme skúsenosti so zborovou lematizáciou a morfosyntaktickou anotáciou pri spracovaní niekoľkých webových korpusov rodiny *Aranea* [3]. Vo svojej práci [6] napríklad autorka uskutočnila experiment s anotáciou anglických korpusových dát, kde agregovala výstupy tagerov *MorphoDiTa*<sup>19</sup> [9] (1 jazykový model), *TreeTagger* (4 modely) a *UDPipe* (4 modely) a lematizátora *CSTlemma* (1 model), pričom dosiahla najmä výrazné celkové zlepšenie lematizácie. Pri spracovaní novších korpusov *Aranea* bola zborová metóda úspešne použitá na anotáciu poľských, dánskych, nórskych, bieloruských, ukrajinských, gruzinských a perzských korpusových dát [1].

Najrozsiahlejšie využitie však našla zborová metóda pri anotácii najnovších verzií niekoľkých slovenských korpusov, pričom okrem štandardných tagerov a lematizátora sa využili aj lexikálne a morfológické dáta získané z niekoľkých slovenských elektronických slovníkov, korektora pravopisu<sup>20</sup> [6] a tiež anglického a českého jazykového modelu (na detekciu cudzojazyčných fragmentov v texte) – tento postup sme preto nazvali „miešaným zborom“. Vzhľadom na to, že naše korpusy sú určené na lexikografické použitie, aj v tomto prípade bola prioritou lematizácia (najmä) slov nenachádzajúcich sa v morfológických lexikónoch jednotlivých nástrojov, čiže potenciálnych neologizmov – kandidátov na zaradenie do heslára pripravovaného slovníka.

<sup>13</sup> <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

<sup>14</sup> <https://ufal.mff.cuni.cz/udpipe/1>

<sup>15</sup> <https://cst.dk/download/uk/>

<sup>16</sup> <https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-3131>

<sup>17</sup> <https://github.com/kuhumcst/texton-linguistic-resources>

<sup>18</sup> Ak pre daný jazyk nie je k dispozícii lexikón, *CSTlemma* odhaduje všetky lemy.

<sup>19</sup> <https://ufal.mff.cuni.cz/morphodita>

<sup>20</sup> <https://github.com/sk-spell/hunspell-sk>

## Bibliografia

- [1] Benko, Vladimír and Švedova, Marija Oleksijivna. 2023. Araneum Ucrainicum: A new Ukrainian web-crawled corpus. In *Leksykohrafična paradyhma 21 st.: teorija i metodolohija : do 100-lett'a vid dňa narodženňa Leonida Sydorovyča Palamarčuka*. Vidp. red. E.A. Karpilovska. – Kyjev: Instytut ukrajinskoji movy NAN Ukrajiny, 2023, pp. 540-557. ISBN 978-617-14-0110-5.
- [2] Benko, Vladimír. 2022. Aranea Go Middle East: Persicum. In *Recent Advances in Slavonic Natural Language Processing (RASLAN 2022)*. Brno: Tribun EU, pp. 113-121. ISBN 978-80-263-1752-4. ISSN 2336-4289.
- [3] Benko, Vladimír. 2024. The Aranea Corpora Family: Ten+ Years of Processing Web-Crawled Data. In *Text, Speech, and Dialogue. 27th International Conference, (TSD 2024)*. Brno, Czech Republic.
- [4] Dell'Orletta, Felice. 2009. Ensemble system for Part-of-Speech tagging. *Evalita 2009 POS Closed Task*. Nepochikované, dostupné na: [https://www.corpusitaliano.it/static/documents/POS\\_ILC.pdf](https://www.corpusitaliano.it/static/documents/POS_ILC.pdf).
- [5] Jongejan, Bart, and Dalianis, Hercules. 2009. Automatic training of lemmatization rules that handle morphological changes in pre-, in- and suffixes alike. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Suntec, Singapore*: Association for Computational Linguistics, pp. 145-153.
- [6] Németh, László, Trón, Viktor, Halácsy, Péter, Rung, András, and Szakadát, István. 2004. Leveraging the open source ispell codebase for minority language analysis. In *SALTMIL Workshop at LREC 2004: First Steps in Language Documentation for Minority Language*. Lisbon, Portugal.
- [7] Pavlova, Olga. 2023. Ensemble Approach to Morphosyntactic Annotation of Web-Crawled Corpora: The English Case. *Bachelor Thesis. Comenius University, Faculty of Arts*. (2023).
- [8] Schmid, Helmut. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of International Conference on New Methods in Language Processing, Manchester*.
- [9] Spoustová, Drahomíra „johanka“, Hajič, Jan., Raab, Jan, Spousta., Miroslav. 2009. Semi-Supervised Training for the Averaged Perceptron POS Tagger. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pp. 763-771, Athens, Greece, March. Association for Computational Linguistics.
- [10] Straka, Milan, Hajič Jan., Straková Jana. 2016. UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia.

## Pod'akovanie

Prezentovaný výskum čiastočne podporila Vedecká grantová agentúra VEGA, projekt číslo 2/0016/21.

## Morfologická defektivita and nevyváženost paradigmatu u českých substantiv, adjektiv a verb

Dominika Kovářiková (Univerzita Karlova)

Nedávné lingvistické výzkumy zaměřené na frekvenční variabilitu morfologického paradigmatu prohloubily naše chápání defektivitu v paradigmatu Defektivita označuje situaci, kdy některé slovní tvary nejsou v paradigmatu zastoupeny nebo doloženy (např. Sims 2015; Nikolaev a Bermel 2022; Bermel et al. 2023). Inherentní defektivita je identifikována na základě intuice rodilých mluvčích (zjišťované psycholingvistickými experimenty), kdežto korpusová defektivita, které se věnuje předkládaný příspěvek, je daná nedoložeností tvarů v jazykových korpusech (Brown a Evans 2021; Brown et al. 2023).

Tato studie rozšiřuje záběr zkoumání z pouhé defektivitu i na celkovou (ne)vyváženost paradigmatu u českých podstatných jmen, přídavných jmen a sloves, která zahrnuje nejen samotnou defektivitu, ale také frekvenčně anomální zastoupení určitých tvarů, zvláště frekvenci vyšší než očekávanou. Paradigmatickou defektivitu tak představuje jakožto součást složitější problematiky. Datový soubor 35 tisíc frekventovaných lemmat z morfologicky označovaných korpusů současně psané češtiny SYN2015 a SYN2020 je analyzován za pomoci aplikace GramatiKat (Kovářiková a Kovářik 2023) s využitím boxplotové analýzy, statistického nástroje, který je schopen zprostředkovat vizualizaci průměrného chování a celkové distribuce tvarů lexémů a identifikovat anomálie. Díky rozsáhlému a velmi spolehlivému morfologickému značkování v korpusech řady SYN, včetně značkování víceslovných slovesných tvarů (verbtage), stanovuje GramatiKat frekvenční kritéria pro vyvážené paradigma a identifikuje anomální frekvenci tvarů včetně defektivitu (nedoloženého tvaru).

Předkládaný výzkum ukazuje, že jednotlivá morfologická paradigmatu se často odchylojí od očekávaného chování stanoveného sledováním velkého počtu lemmat. Prvním typem odchylky jsou lemmata, která vykazují defektivitu, tedy jeden nebo několik tvarů není doloženo. Počet lemmat vykazujících defektivitu se snižuje společně se stoupající frekvencí lemmatu (minimální frekvence lemmat v našem materiálu je 100). Druhou možnou odchylkou je výskyt jednoho nebo několika tvarů s frekvencí, která významně přesahuje očekávání stanovené s využitím boxplotové analýzy. Díky těmto frekvenčním odchylkám lze konstatovat, že morfologické chování jednotlivých lexémů není nijak jednotné, naopak, lexémy se od sebe mohou velice odlišovat, což je dáno jejich individuálními sémantickými vlastnostmi a syntaktickým a kolokačním chováním. Příkladem může být neexistence imperativu u neagentních sloves, vysoká frekvence první osoby u sloves smyslového vnímání nebo převaha ženských tvarů adjektiva *rychlouvarný*, která je dána jeho monokolokabilitou (vyskytuje se v podstatě jen v kolokaci se substantivem *konvice*).

Výsledky studie naznačují, že nevyvážená paradigmatu se vyznačují nejen nedoloženými tvary (tedy defektivitou), ale také jemnějšími formami frekvenční variability paradigmatu, tedy jakýmkoli odchylkami od očekávané frekvence tvarů. Konstatováním těchto nedeterministických rysů lexikálních jednotek studie přispívá k nuancovanému chápání morfologických paradigmatu a naznačuje, že struktura a užívání jazyka jsou ovlivňovány řadou nepředvídatelných a proměnlivých jazykových faktorů.

## Bibliografie

Bermel, N., Knittl, L., Nikolaev, A. (2023): Uncertainty in the production of Czech noun and verb forms. *Word Structure* 16(2-3), pp. 258–283.

Brown, D., Evans, R. (2021): Two kinds of nothing: Exploring Corpus-based Approaches to Defectiveness in Russian. Presentation, *Feast & Famine Workshop*.

Brown, D., Bermel, N. & D. Kovářiková (2023): What we know, don't know, and want to know about defectivity and overabundance. Presentation, *SLE Conference Athens 2023*.

Kovářiková, Dominika & Kovářik, Oleg. 2023. *GramatiKat. Tool for Research of Grammatical Categories and Grammatical profiles*. Prague: FF UK. Available at WWW: <<http://www.korpus.cz/gramatikat>>.

Nikolaev, A., & Bermel, N. (2022). Explaining uncertainty and defectivity of inflectional paradigms. *Cognitive Linguistics*, 33(3), 585–621. <https://doi.org/10.1515/cog-2021-0041>

Sims, A. D. (2015). *Inflectional defectiveness*. Cambridge University Press.

## Použité korpusy

Křen, M. et al. (2015): *SYN2015: reprezentativní korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha. Dostupný z WWW: <http://www.korpus.cz>

Křen, M. et al (2020): *SYN2020: reprezentativní korpus psané češtiny*. Ústav Českého národního korpusu FF UK, Praha. Dostupný z WWW: <http://www.korpus.cz>

Víceřovinný popis jazyka: Příklad částic<sup>21</sup>

Marie Mikulová, Barbora Štěpánková (Univerzita Karlova)

Při tvorbě korpusů se zvažuje otázka, jakým způsobem zachytit množství jazykových jevů různých typů, od morfologie přes větnou syntax až po diskurzní vztahy. Jednou z možností je víceřovinná anotace. (Různé přístupy k víceřovinné anotaci srov. např. Silvano a kol. 2021, Ide – Pustejovsky 2017.) Teoretické výhody víceřovinné anotace jazyka byly už nejméně zmiňovány (Hajičová a kol. 2022). Patří mezi ně např. vzájemná opora a možnost kontroly při anotaci jednotlivých rovin a komplexní uchopení popisu jazyka. V tomto příspěvku se zaměříme na víceřovinnou anotaci v Pražských závislostních korpusech (PDT, Hajič a kol. 2020) a její využití při lingvistickém výzkumu chování částic v češtině.

Víceřovinná struktura PDT je tvořena třemi základními rovinami<sup>22</sup> (Hajič a kol. 2017): lineární morfologickou rovinou a dvěma závislostními syntaktickými rovinami - analytickou (formální) a tektogramatickou (sémantickou).

Kontroly konzistence anotace pro připravované vydání konsolidované verze PDT-C 2.0 vyjevily značný nesoulad v manuální anotaci neohebných slov mezi všemi rovinami a poukázaly na problematická místa, tedy ta, která nejsou v souladu s očekávaným/převažujícím chováním těchto slov v jazyce. Odráží se zde především nedostatečnost samostatné/oddělené morfologické anotace neohebných slov, tj. slov, jejichž slovní druh je definován i syntakticky a sémanticky. Teprve kombinace manuální anotace na všech třech rovinách umožňuje (i) relevantní komplexní popis těchto slov, zejména s ohledem na jevy, které jsou tradičními přístupy dosud nevyřešené nebo neřešitelné, navíc také upozorňuje (ii) na jejich nové, dosud nepopsané funkce.

V oblasti částic je to ad (i) například částicově-navazovací chování souřadících spojek nebo spojovací funkce zdůrazňovacích částic (srov. překrývání významů viz už Bauer (1964), Vondráček (1999), Štěpánková (2010)), ad (ii) proces významového/funkčního posunu mezi adverbii a částicemi (gramatikalizace, pragmatikalizace).

Jako příklad (i) uvádíme spojku *ale* a částici *hlavně*. Anotací na třech různých rovinách lze adekvátně podchytit, že spojka ( $J^{\wedge}$ -----) *ale* může vedle své hlavní funkce spojovat věty (Coord) plnit i funkce náležející částicím (AuxZ), a sice navazování na předcházející text (PREC); srov. i její chápání jako diskurzního markeru (např. Hoffmannová 2019: 314). Na výrazu *hlavně* pak demonstrujeme využití primárně zdůrazňovací částice (RHEM) pro modifikaci významu souřadící spojky (CM). Srov. v tab. 1.

Jako příklad (ii) poslouží výraz *stěží*, který má ve slovníku (srov. Slovník spisovné češtiny) kromě významu jistotní částice (předpokládaná kombinace anotace: TT-----; AuxZ; MOD) také význam způsobového příslovce (Db-----; Adv; MANN). V případech, kdy tomuto zařazení neodpovídá anotace, resp. kdy se kombinace anotací na jednotlivých rovinách vymyká předpokládanému řešení, je pravděpodobné, že se jedná o další funkci. V datech se tak projevilo užívání výrazu *stěží* ve funkci měrového určení (EXT), a to ve spojení výrazu s číselným údajem (podobně jako je to u *asi*). Srov. v tab. 1.

Na příkladu neohebných slov, tj. slov, která ač mají jen jeden paradigmatický tvar, mohou plnit několik různých syntaktických funkcí a vyjadřovat různé, i poměrně odlišné sémantické distinkce, ukazujeme, že právě víceřovinná struktura umožňuje komplexní popis těchto relací mezi morfologickým slovním druhem, syntaktickou funkcí a vyjadřovanou sémantikou. Zvolené výrazy jsou přitom reprezentanty větších skupin a podobné chování lze očekávat i u dalších slov.

<sup>21</sup> Příspěvek byl vytvořen s podporou projektů GAČR GA23-05240S a GA23-05238S a projektu MŠMT LM2023062 LINDAT/CLARIAH-CZ.

<sup>22</sup> Teoreticky je opřena o východiska Funkčního generativní popisu (Sgall a kol. 1986, Panevová a kol. 2020).

	Morfologická rovina	Analytická rovina	Tektogramatická rovina	Příklad
stěží	TT-----	AuxZ	MOD	<i>To však v Dánsku <b>stěží</b> přesvědčí protimaastrichtskou lobby.</i>
	TT-----	Adv	EXT	<i>Do roku 1990 totiž zaměstnávaly například Závody těžkého strojírenství (ZTS) na 88 tisíc lidí, zatímco nyní jich u nástupnických organizací pracuje <b>stěží</b> polovina.</i>
	Db-----	Adv	MANN	<i>Bylo příjemné poslouchat výtečný jazz, aniž byste museli sedět v zakouřeném klubu, <b>ale</b> stejně jako v první polovině byla většina skladeb snadno přijatelná a nakonec <b>stěží</b> zapamatovatelná.</i>
ale	J^-----	Coord	ADVS GRAD	<i>Hlavní město by podle Prokopa mělo být nejen všestranným centrem státu, <b>ale</b> hlavně domovem milionu Pražanů.</i>
	J^-----	AuxZ	PREC	<i>V řadě případů bylo zahájeno správní řízení. Většinou byly <b>ale</b> přestupky řešeny pokutami.</i>
hlavně	TT-----	AuxZ	RHEM	<i>Dlažba z Bety je vhodná <b>hlavně</b> do center historických měst.</i>
	TT-----	AuxZ	CM	<i>V nových prostorách, vybavených klimatizací a dobře zabezpečených, budou cenné materiály údajně lépe využitelné a <b>hlavně</b> odolnější proti ztrátě či zničení.</i>

Tabulka 1: Vícerovinový popis částic v PDT

## Bibliografie

- Bauer, Jaroslav (1964): Spojky a částice. In: *Sborník prací filosofické fakulty brněnské university*, A 12, 131–39.
- Filípec, Josef a kol. (2003): *Slovník spisovné češtiny pro školu a veřejnost*. (2003). Praha: Academia.
- Hajič, Jan – Bejček, Eduard – Hlaváčová, Jaroslava – Mikulová, Marie – Straka, Milan – Štěpánek, Jan – Štěpánková, Barbora (2020): Prague Dependency Treebank - Consolidated 1.0. In: *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*, European Language Resources Association, Marseille, 5208–5218,
- Hajič, Jan – Hajičová, Eva – Mikulová, Marie – Mírovský, Jiří (2017): Prague Dependency Treebank. In: *Handbook of Linguistic Annotation*. (N. Ide and J. Pustejovsky eds.), Dordrecht: Springer. 555–594.
- Hajičová, Eva – Mikulová, Marie – Štěpánková, Barbora – Mírovský, Jiří (2022): Advantages of a complex multilayer annotation scheme: The case of the Prague Dependency Treebank. In: *Proceedings of The 16th Linguistic Annotation Workshop (LAW-XVI) within LREC2022*, Marseille. 70–78,
- Hoffmannová, Jana – Homoláč, Jiří – Mrázková, Kamila, eds. (2019). *Syntax mluvené češtiny*. Praha: Academia.
- Ide, Nancy – Pustejovsky, James, eds. (2017): *Handbook of Linguistic Annotation*, Dordrecht: Springer.
- Panevová, Jarmila – Hajičová, Eva. – Kettnerová, Václava – Kolářová, Veronika – Lopatková, Markéta – Mikulová, Marie – Ševčíková, Magda (2020): Funkční generativní popis – rámeček pro konzistentní popis gramatiky. *Naše řeč*, 103 (1–2), 55–78.
- Silvano, Purificação – Leal, António – Silva, Fátima – Cantante, Inês – Oliveira, Fatima – Jorge, Alípio Mario (2021): Developing a multilayer semantic annotation scheme based on ISO standards for the visualization of a newswire corpus. In *Proceedings of the 17th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, Groningen. Association for Computational Linguistics.
- Sgall, Petr – Hajičová, Eva – Panevová, Jarmila (1986): *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: D. Reidel Publishing Company.
- Štěpánková, Barbora (2010): K možnostem zpracování spojek a částic ve výkladovém slovníku (k tzv. polyfunkčnosti spojek a částic a nejednoznačnosti při určování slovních druhů). *Bohemica Olomucensia*, 2, 111–118.
- Vondráček, Miloslav (1999): Příslovce a částice – hranice slovního druhu. *Naše řeč*, 82 (2), 72–78.

### Analýza mezinotátorských neshod jako cesta k interpretaci textu

Anna Nedoluzhko, Šárka Zikánová, Eva Hajičová, Jiří Mirovský (Univerzita Karlova, ÚFAL MFF)

Interpretace textu není a nemůže být úplná, jednoznačná ani objektivní, což je dáno povahou přirozeného jazyka. Na tento fakt poukazují jak práce českých a zahraničních jazykovědců (srov. např. Hausenblas 1988, Daneš 1988, Hoffmannová 2017 aj.), tak i počítačové modely věnující se například automatickému překladu, a to jak starší teoretické práce z počátků tohoto oboru (viz např. Melčuk 1974), tak i současné analýzy výsledků automatických překladů (Kloudová et al. 2023).

Nicméně při anotaci jazykových dat nezřídka implicitně předpokládáme jen jedinou interpretaci, již vyžadujeme po anotátorech. Čím větší je pak mezinotátorská shoda, tím je anotace daného jazykového jevu často považována za úspěšnější. Za poslední desetiletí anotace rozsáhlých jazykových korpusů se stal tento přístup samozřejmostí, a to na všech jazykových rovinách, od morfologie až po anotaci diskurzu. Naráží na jistá úskalí, jichž jsou si výzkumníci čím dál více vědomi (srov. Artstein 2017) a kterým se chceme věnovat v daném příspěvku.

Zdroje odlišného porozumění textu mívají odlišný charakter. Zatímco některé jazykové prostředky jsou víceznačné, což znamená, že mohou vyjadřovat více jasně odlišných významů, jiné jsou vágní, neodkazují přesně k detailům skutečnosti, zahrnují jistou „nedořečenost“ (van Rooij 2011, Zwicky – Sadock 1975). Na morfologické a syntaktické úrovni se projevuje spíše víceznačnost, srov. častou homonymií tvarů, která se však ve většině případů náležitě interpretuje v kontextu věty. Na úrovni syntaxe může být porozumění dále různě ovlivněno aktuálním členěním věty, intonačními důrazy, slovosledem apod. Na úrovni textu pak do hry výrazněji vstupuje vágnost. Je možné tuto vágnost v textu nějak systematicky dohledat? Věříme, že částečně ano, a to pomocí analýzy mezinotátorských neshod (s odhlédnutím od zjevných chyb v anotaci).

Podívejme se na dva příklady z Pražského závislostního korpusu (PDT 3.0, Hajič a kol., 2020), ve kterém jsou ručně anotovány mj. koreferenční a diskurzivní vztahy.

- (1) *Naše metodika a techniky vyšetřování mohou u jednotlivých vyjmenovaných profesí pouze předpovědět **úspěšnost** v rozdílné **míře pravděpodobnosti**. Nejvyšší  $\emptyset$  bude u svářeče a zedníka, nižší  $\emptyset$  u manažera a redaktora.*
- (2) *Ve středu vychází Slunce v 05:06 a zapadá ve 21:05, Měsíc vychází ve středu v 00:43 a zapadá v 16:27 hodin letního času.*

V příkladu (1) může nevyjádřený podmět ve druhé větě odkazovat jak na výraz *úspěšnost*, tak na výraz *míře pravděpodobnosti* ve větě předchozí. V tomto případě vede různá volba antecedentu v koreferenčním vztahu k odlišné interpretaci textu. Diskurzivní vztah v příkladu (2) mezi větou o Slunci a větou o Měsíci pak může být pochopen buď jako konfrontační (mezi větami můžeme doplnit spojku *kdežto*), nebo jako konjunktivní (s možností doplnit spojku *a*); na rozdíl od (1) zde však různé čtení nemá praktické důsledky pro celkové porozumění.

Podobné případy víceznačnosti a vágnosti textu je často obtížné identifikovat, pokud text čte a analyzuje pouze jeden čtenář. Nejlépe vystupují na povrch při analýze mezinotátorských neshod (Basile aj. 2021, Planc 2022). Vzhledem k tomu, že diskurzivní vztahy a koreference v Pražském závislostním korpusu jsou částečně anotovány více anotátory a že neshodám v těchto anotacích již byla věnována jistá výzkumná pozornost (Jinová et al. 2012, Nedoluzhko – Mirovský 2013), máme takový materiál k dispozici. V našem příspěvku se zaměříme na analýzu mezinotátorských neshod diskurzivních jevů a koreference a pokusíme se ukázat, jak analýza mezinotátorských neshod může přispět k pochopení interpretace textu. Vycházíme z předpokladu, že neshoda mezi anotátory indikuje různá porozumění textu, což je nepochybně funkční a obohacující.

## Bibliografie

- Artstein, R. (2017). Inter-annotator agreement. In: Ide N., Pustejovsky, J. (eds.), *Handbook of linguistic annotation*. Dordrecht: Springer, pp. 297-313.
- Basile V., Fell M., Fornaciari T., Hovy D., Paun S., Plank B., Poesio M., Uma A. et al. (2021). We need to consider disagreement in evaluation. In: *Proceedings of the 1st Workshop on Benchmarking: Past, Present and Future*. Association for Computational Linguistics, pp. 15–21. <https://aclanthology.org/2021.bppf-1.3.pdf> <cit. 23. 3. 2024>
- Daneš F. (1988). *Předpoklady a meze interpretace textu*. In: *Slavica Pragensia: Fungování textu ve společenské komunikaci*. Praha: UK.
- Hoffmannová J. (2017). Neurčitost textu. In: Karlík P. et al. (eds.), *Nový encyklopedický slovník češtiny*. URL: <https://www.czechency.org/slovník/NEURČITOST TEXTU> <cit. 23. 3. 2024>
- Hajič J., Bejček E., Hlaváčová J., Mikulová M., Straka M., Štěpánek J., Štěpánková B. (2020). Prague Dependency Treebank – Consolidated 1.0. In: *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*. Marseille, France: European Language Resources Association, pp. 5208–5218.
- Jínová P., Mírovský J., Poláková L. (2012). Analyzing the Most Common Errors in the Discourse Annotation of the Prague Dependency Treebank. In: *Proceedings of the 11th TLT*. Lisboa, Portugal, pp. 127–132.
- Hausenblas K. Interpretace textu a její druhy v současné komunikaci. (1991). In: *Konfrontačný a komparatívny výskum jazykovej a literárnej komunikácie: Jazykovedný zborník 8*. Bratislava: SPN.
- Kloudová V., Mráček D., Bojar O., Popel M. (2023). Možnosti a meze tvorby tzv. optimálních referenčních překladů: po stopách „překladatelštiny“ v profesionálních překladech zpravodajských textů. *Slovo a slovesnost*, 84 (2), s. 122–156.
- Melčuk I. (1974). *Opyt teorii lingvističeskich modelej „Smysl ⇔ Tekst“* [Toward a Theory of Meaning-Text Linguistics Models]. Moscow: Nauka.
- Nedoluzhko A., Mírovský J. (2013). Annotators' Certainty and Disagreements in Coreference and Bridging Annotation in Prague Dependency Treebank. In: *Proceedings of the Second International Conference on Dependency Linguistics (DepLing 2013)*. Prague, Czech Republic: Charles University in Prague, Matfyzpress, pp. 236–243. <https://aclanthology.org/W13-3726.pdf> <cit. 23. 3. 2024>
- Plank B. (2022). *The "Problem" of Human Label Variation: On Ground Truth in Data, Modeling and Evaluation*. <https://arxiv.org/abs/2211.02570> <cit. 23. 3. 2024>
- van Rooij R. (2011). Vagueness and linguistics. In Ronzitti G. (ed.), *Vagueness: A Guide*. Logic, Epistemology, and the Unity of Science, vol. 19. Springer, pp. 123–170. [https://doi.org/10.1007/978-4-007-0375-9\\_6](https://doi.org/10.1007/978-4-007-0375-9_6) <cit. 23. 3. 2024>
- Zwicky A. M., Sadock J. M. (1975). Ambiguity tests and how to fail them. In: Kimball J. (ed.), *Syntax and Semantics*, vol. 4, Seminar Press, pp. 1–36. <https://kb.osu.edu/server/api/core/bitstreams/e10019c8-0461-5d37-9dcf-5ad771a222aa/content> <cit. 23. 3. 2024>



**Synkretismus českých adjektiv a jejich značení v korpusech:  
Několik poznámek k Minervině zradě**

Klára Osolsobě, Hana Žižková (Masarykova univerzita)

Cílem naší korpusové sondy bylo analyzovat chování posesivních adjektiv v otázce pádové homonymie v češtině. Zaměřili jsme se na dvouslovné oslovení v pátém pádě, kde se setkáváme s konstrukcí posesivního adjektiva následovaného substantivem / substantiva následovaného posesivním adjektivem (Minervin[ao] zrado / zrado Minervin[ao]). Zkoumali jsme použití této konstrukce jak v případě depropriálních, tak v případě deaplativních posesiv, a položili jsme si i otázku, zda může být synkretismus/homonymie ovlivněn/a pořadím zkoumaných slov (např. *Minervina zrado* vs. *zrado Minervino*). V úzu jsme okrajově narazili na kolísání u vokativu posesiv na *-in*, ale objevili jsme i řídké doklady posesiv na *-ov*, které mají ve vokativu koncovku *-o*, kterou v případě, že řídicí substantivum je femininum, lze interpretovat jako tvar vokativu podle jmenné flexe u feminin, který není homonymní s tvarem nominativu.

Obecně lze výskyt vokativu předpokládat spíše v mluvených korpusech, ty jsou ale malé. Proto jsme se pro náš výzkum obrátili ke korpusům psaného jazyka. Ze všech dostupných jsme vybrali tři: SYN v12, CsTenTen12+17+19 a Araneum Bohemicum Maximum. Při excerpci dat se ukázalo, že pádový synkretismus/homonymie do značné míry ovlivňuje i kvalitu morfologického značení a vyhledávání podle morfologické značky je nespolehlivé. Dotazy jsme proto v CQL formulovali tak, abychom vyhledali tvar slova a předešlé či následující substantivum v pátém pádě a určeném rodě. Ke každému typu (antepozice vs. postpozice vs. posesiva od maskulin vs. posesiva od feminina) jsme sestavili dotaz v CQL, prošli jsme dohledané výskyty, ručně jsme vyřídili nežádoucí spojení.

Naše korpusová sonda přinesla dvě zjištění: 1) ukázalo se, že počty výskytů v antepozici a postpozici jsou frekvenčně srovnatelné, zakončení na *-o* (*babiččino prdýlko* vs. *zrado Minervino*) se však mnohem častěji se vyskytovalo v antepozici. Vzhledem k povaze zkoumaných dat není možné udělat analýzu, ve kterých typech textů se pozorovaná sousloví vyskytují nejčastěji; 2) zjistili jsme, že hledaná sousloví vykazovala velkou míru nesprávného morfologického značení, a to jak ve webových korpusech csTenTen, které využívají guessery (slova zakončená na *-ino* jsou interpretována jako posesivní adjektiva v nominativu singuláru neutra, přičemž se jedná o italská maskulina), tak v korpusech řady Araneum a SYN, kde jsou tvary končící na *-ovo/-ino* interpretovány jako tvary v nominativu/akuzativu singuláru neutra v případě, že rozvíjeným substantivem je femininum.

Věříme, že naše studie přináší hlubší porozumění chování posesivních adjektiv v češtině a poukazuje na oblast, ve které možná uživatelé korpusů chybné značení neočekávají.

## **Bibliografie**

Benko, V., 2013-2023. Aranea: Family of Comparable Web-crawled Corpora. L. Štúr Institute of Linguistics, Bratislava. <http://aranea.juls.savba.sk/guest/> and <http://unesco.uniba.sk/guest/>.

Křen, M. Et al., 2023. Korpus SYN, verze 12 ze 29. 12. 2023. Ústav Českého národního korpusu FF UK, Praha. <https://www.korpus.cz>.

Suchomel, V., 2018. csTenTen17, a Recent Czech Web Corpus. In: Horák, A., Rychlý, P., Rambousek, A. (Eds). Proceedings of the Twelfth Workshop on Recent Advances in Slavonic Natural Languages Processing RASLAN 2018. Tribun EU, Brno, pp. 111-123.

## Hierarchizace a granularita sémantické klasifikace cirkumstantů na příkladu způsobových určení

Jarmila Panevová, Marie Mikulová (Univerzita Karlova)

Problematika sémantické klasifikace cirkumstantů zahrnuje řadu aspektů včetně hierarchizace a granularity významových kategorií. V příspěvku představíme na materiálu široce způsobových určení **dvouúrovňovou sémantickou klasifikaci cirkumstantů** založenou na teorii Funkčního generativního popisu (Sgall a kol. 1986) a popsanou poprvé v Panevová 1980 (rozpracovanou pro cirkumstanty prostorové a časové v Mikulová – Panevová 2021, měrové v Mikulová 2023, prostředku v Mikulová 2024). Jedná se o základní hrubou klasifikaci na tzv. funktoxy a jemnou klasifikaci na tzv. subfunktoxy. Funktoxy v souladu s tradičním pojetím příslovečných určení (srov. Šmilauer 1947) vymezujeme pomocí otázek, kterými se ptáme na jednotlivé okolnosti, např. *kde*, *kdy*, *proč* apod. Subfunktor pak popisuje jemné sémantické rozdíly uvnitř jednoho funktoxy, přičemž sdílí základní sémantiku funktoxy (odpovídá na stejnou otázku po okolnosti). Cirkumstanty ohodnocené různými funktoxy nejsou substituovatelné, tj. otázka *jak* nemůže být zodpovězena cirkumstantem s významem *kde* (funktor LOC) jako v (7). Tato otázka může být zodpovězena cirkumstantem způsobu (MANN), který může vyjadřovat různé dílčí významy (subfunktoxy: property-of-event v (1), property-of-actor+ / property-of-actor– v (2)). **Granularita klasifikace okolnostních významů** je zakotvena ve struktuře daného jazyka. Stanovujeme jen takové dílčí významy (subfunktoxy), které daný jazyk (v našem případě čeština) vyjadřuje, pro které existuje formální opora. Jemná klasifikace cirkumstantů je založena na detailní analýze bohatého jazykového materiálu prezentovaného v PDT-korpusech (Hajič a kol. 2020).

- (1) Pracoval *rychle* / v *rychlosti*.
- (2) Pracoval v *rukavicích* / *bez rukavic*.
- (3) Pracoval s *ohledem na své schopnosti*.
- (4) Pracoval s *lopatou*.
- (5) Pracoval *podle plánu*.
- (6) *Co do výkonnosti* pracoval velmi *dobře*.
- (7) Pracoval v *garáži*.

Situace však není u všech cirkumstantů stejná. Zatímco u určení prostorových a časových je systém otázek návodný pro určení příslušného funktoxy (*kde* (LOC), *kam* (DIR3), *kdy* (TWHEN), *odkdy* (TSIN) atd.), u široce způsobových určení (jako je vlastní způsob (MANN), zřetel (REG), prostředek (MEANS), kritérium (CRIT) ad.) formulace otázky nebývá snadná. Na základní způsobovou otázku *jak* lze odpovědět různě; srov. (1) – (5). Přitom však neplatí, že na všechna široce způsobová určení se lze vždy zeptat *jak*, a tedy hodnotit všechna tato významově variabilní určení jako reprezentanty jednoho funktoxy; srov. příklad (6), ve kterém je odpovědí na otázku *jak* pouze cirkumstant *dobře* a nikoliv zřetelové určení *co do výkonnosti* (viz i Daneš – Knapová 1972). Pro specifická způsobová určení je třeba naformulovat specifické otázky: *se zřetelem k čemu* (REG; (3)); *pomocí čeho* (MEANS; (4)); *podle jakého vodítka* (CRIT; (5)).

Podobně komplikovaná je i situace na úrovni subfunktoxy. Zatímco pro vyjádření časově-prostorových významů existuje v češtině systém formálních prostředků (*po* vs. *před*, *nad* vs. *pod*), pro rozlišení různých dílčích významů v rámci široce způsobových určení čeština zpravidla nedisponuje speciálními (primárními) prostředky (výjimkou je např. vyjádření +/- opozice; srov. (2)), ale využívá prostředky s významem primárně prostorově-časovým a „spoléhá“ na lexikální význam užitých plnovýznamových slov (srov. užití předložky *v+6* v (1) a (2)). To však neznamená, že forma vyjádření způsobových určení je libovolná; typické formy lze stanovit pro jednotlivé dílčí významy. Navíc pro explicitní vyjádření některých způsobových významů se v češtině ustálila celá řada předložek sekundárních ((3), (5), (6)), které jsou vodítkem pro vyčleňování jednotlivých dílčích významů.

## Poděkování

Tato studie je podporována projektem GAČR GA23-05238S a využívá jazykové zdroje a nástroje vyvinuté, uchovávané a distribuované v rámci projektu MŠMT LM2023062 LINDAT/CLARIAH-CZ.

## Bibliografie

Daneš František, Knappová Marie (1972): Příslovce a příslovečné určení zřetelově vymezovací. *Slovo a slovesnost*, 33(2):106-115.

Hajič Jan, Bejček Eduard, Bémová Alevtina, Buráňová Eva, Fučíková Eva, Hajičová Eva, Havelka Jiří, Hlaváčová Jaroslava, Homola Petr, Ircing Pavel, Kárník Jiří, Kettnerová Václava, Klyueva Natalia, Kolářová Veronika, Kučová Lucie, Lopatková Markéta, Mareček David, Mikulová Marie, Mírovský Jiří, Nedoluzhko Anna, Novák Michal, Pajas Petr, Panevová Jarmila, Peterek Nino, Poláková Lucie, Popel Martin, Popelka Jan, Romportl Jan, Rysová Magdaléna, Semecký Jiří, Sgall Petr, Spoustová Johanka, Straka Milan, Straňák Pavel, Synková Pavlína, Ševčíková Magda, Šindlerová Jana, Štěpánek Jan, Štěpánková Barbora, Toman Josef, Urešová Zdeňka, Vidová Hladká Barbora, Zeman Daniel, Zikánová Šárka, Žabokrtský Zdeněk (2020): *Prague Dependency Treebank - Consolidated 1.0 (PDT-C 1.0)*. LINDAT/CLARIAHCZ, Charles University, Prague, <http://hdl.handle.net/11234/1-3185>.

Mikulová Marie (2023): Expressing Measure in Czech (a Corpus-based Study). *Jazykovedný časopis / Journal of Linguistics*, 74(1): 108-118.

Mikulová Marie (2024): Fine-grained Classification of Circumstantial Meanings within the Prague Dependency Treebank Annotation Scheme. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, Torino.

Mikulová Marie, Panevová Jarmila (2021): *Formy a funkce okolnostních určení v češtině. Určení prostorová a časová*. Praha: Institute of Formal and Applied Linguistics, Charles University.

Panevová Jarmila (1980): *Formy a funkce ve stavbě české věty*. Praha: Academia.

Sgall Petr, Hajičová Eva, Panevová Jarmila (1986): *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht: D. Reidel Publishing Company.

Šmilauer Vladimír (1947): *Novočeská skladba* [vydání I.]. Praha: Ing. Mikuta.

## Korpusy DIA a jejich využití

Klára Pivoňková, Martin Stluka (Univerzita Karlova)

Vytvořit kvalitně zpracovaný anotovaný diachronní korpus představuje stále i v době rozvoje neuronových sítí a AI specifickou výzvu. I když historický textový materiál nemusí pro rodilého mluvčího představovat překážku z hlediska porozumění, bývá od soudobé podoby jazyka natolik odlišný, že prosté využití automatických nástrojů natrénovaných na současném stavu jazyka nemusí představovat uspokojivé řešení. Netriviálnost vytvoření diachronního korpusu a potřeba nalézt co nejvhodnější metodu zpracování materiálu představuje výzvu i pro diachronní korpusy jiných jazyků než češtiny (např. korpusy staré polštiny nebo angličtiny aj.).

V tomto příspěvku představíme práci s českým materiálem 19. století. Při zpracování textů z tohoto období bylo zapotřebí najít řešení pro velkou míru rozkolísanosti úzu. Jelikož synchronní anotační nástroje neposkytovaly při anotaci korpusového materiálu uspokojivé výsledky, byl pro účely anotace češtiny 19. století zcela nově vytvořen rozsáhlý morfologický slovník, který umožnil přizpůsobit lemmatizaci a tagset koncepci s velkou mírou podrobnosti, se zaměřením na využitelnost z hlediska (historické) lingvistiky.

Rádi bychom prezentovali zkušenosti s kompletní tvorbou korpusu od přípravy samotného textového materiálu až po vytvoření nástrojů pro anotaci, které nám opakovaně ukazovaly, že není možné vytvořit naprosto univerzální korpus ideální pro všechny typy uživatelských přístupů. Zásadní se nám v tomto směru jeví míra podrobnosti zpracování jednotlivých jevů, resp. rozhodnutí ohledně jejich autonomie a hierarchie.

Představíme konkrétní příklad z naší praxe: tentýž textový materiál se podařilo připravit pro různé typy uživatelského využití díky trojímu způsobu lemmatizace a morfologické anotace. Jedná se o anotaci diachronních dat v rámci Ústavu Českého národního korpusu pokrývajících 1. a 2. polovinu 19. století.

Výchozí koncepce má jednoúrovňovou lemmatizaci a při klasifikaci jazykových jevů zohledňuje poznatky o vývoji češtiny. Například k pravidelným hláskovým změnám přistupuje při lemmatizaci odlišně ve srovnání s variabilitou slovtvorných prostředků (např. slova *ovocní a ovocný* mají samostatná lemmata, *vejr* a *výr* mají totožné lemma *výr*; slovo *spůsob* získá lemma *spůsob* a bude pomocí našeptavače spojeno s lemmatem *způsob*). Tato koncepce bude využita pro tvorbu korpusů DIA1850 (1. polovina 19. století) a DIA1900 (2. polovina 19. století).

Díky své podrobnosti se výchozí zpracování stane základem pro víceúrovňovou lemmatizaci podle současných anotačních standardů korpusů řady Syn (slovo *vejr* získá lemma *výr* a sublemma *vejr*) a vedle toho také poslouží jako východisko pro UD anotaci v rámci projektu Hičkok, která bude použita při tvorbě korpusu pokrývajícího celé období psané češtiny (slovo *spůsob* získá novočeské lemma *způsob*).

Chceme ukázat specifika těchto tří koncepcí jak z hlediska přípravy korpusů, tak z pohledu uživatele. Ukážeme, jaké typy variability mohou v tomto období komplikovat určování tvarů (např. varianty v oblasti pravopisu, morfologie a slovtvorby) a poskytneme kvantifikaci těchto typů u tvarů, které zvolenými nástroji (automatické) anotace nebyly rozpoznány. Poskytneme srovnání tří pojetí lemmatizace od přidělování lemmat sadám paradigmát až po různý stupeň formálního/funkčního přístupu při určování slovních tvarů.

## Bibliografie

Benešová, L., Kučera, K., Najbrtová, K., Pivoňková, K., Stluka, M. (2019). Korpus DIA1900: jeho koncepce a vytváření. *Časopis pro moderní filologii*, 105, pages 121–140.

Hana, J., Lehečka, B., Feldmann, A., Černá, A., Oliva, K. (2011). Building a Corpus of Old Czech. In *Proceedings of the Adaptation of Language Resources and Tools for Processing Cultural Heritage Objects Workshop associated with the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 9–15.

*Helsinki Corpus of English Texts*. Dostupné z:

<[https://varieng.helsinki.fi/series/volumes/14/rissanen\\_tyrkko/](https://varieng.helsinki.fi/series/volumes/14/rissanen_tyrkko/)>

Kieraś, W., Woliński, M. (2018). Manually Annotated Corpus of Polish Texts Published between 1830 and 1918. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki, Japan: ELRA, pages 3854–3859.

Kučera, K., Najbrtová, K., Pivoňková, K., Řehořková, A., Stluka, M. (2019). Korpus českého jazyka 2. poloviny 19. století. *Časopis pro moderní filologii*, 101, pages 92–97.

Osolsobě, K., Hlaváčová, J., Petkevič, V., Šimandl, J., Svášek, M. (2017): Nová automatická morfologická analýza češtiny. *Naše řeč*, 4, pages 225–234.

Synková, P. (2017): Popis staročeské apelativní deklinace (se zřetelem k automatické morfologické analýze textů Staročeské textové banky). Praha: Filozofická fakulta UK, 2017.

Tichý, O. (2017): Nástroj na tvaroslovnou analýzu staré angličtiny. *Časopis pro moderní filologii*, 99, 1, pages 40–54.

**Anotace korpusů a AI**

Pavel Rychlý (Masarykova Univerzita, Fakulta Informatiky)

V přednášce stručně zmíníme historii vytváření korpusů a jejich anotací, zejména z pohledu jejich použití pro dřívější i současné automatické zpracování přirozeného jazyka. Dotkneme se některých koncepčních mýtů, které u anotací korpusů stále přetrvávají. Dále vysvětlíme strukturu a klíčové vlastnosti současných velkých jazykových modelů, které reprezentují umělou inteligenci (AI) na poli zpracování textů. Ukážeme jakým způsobem a z jakých dat se velké jazykové modely učí, a co z toho plyne pro jejich použití. V závěru přednášky uvedeme příklady použití jazykových modelů pro anotaci korpusů a nastíníme velký potenciál AI pro anotaci v blízké budoucnosti.

## Morfologická anotace korpusu rané češtiny Chroma

Jakub Sláma (Ústav pro jazyk český AV ČR, Praha)

Anna Chromá (Filozofická fakulta UK, Praha)

Klára Matiasovitsová (Filozofická fakulta UK, Praha)

Korpus Chroma je akviziční korpus, který longitudinálně zachycuje produkci 7 monolingvních dětí osvojujících si češtinu v raném předškolním věku. Jeho poslední verze (z července 2023) zahrnuje i lematizaci a morfologickou anotaci. Korpus je publikován v databázi CHILDES (MacWhinney, 2000), která shromažďuje korpusy zachycující osvojování různých jazyků. Slovanské jazyky (včetně korpusu Chroma) však představují pouze 4 % všech jazyků v databázi, zařazení češtiny jako morfologicky bohatého jazyka tak představuje významný přínos. Zároveň ale ovlivňuje podobu korpusu, která se tak odlišuje od jiných korpusů vzniklých v českém prostředí, sdružených především pod hlavičkou Českého národního korpusu (<https://www.korpus.cz/>). Rozdíly se týkají jak způsobu přepisu, tak morfologické anotace.

Korpus Chroma tvoří přepisy audionahrávek zachycující spontánní mluvenou interakci dětí ve věku mezi 19 až 49 měsíci s jejich blízkými pečovateli v domácím prostředí. Nahrávky jsou zaměřeny na běžné denní aktivity, jako je hra, prohlížení knížek nebo domácí práce. Celkem obsahuje 99 388 tokenů v 42 103 dětských promluvách a 238 211 tokenů v 61 252 promluvách dospělých; průměrný objem nahrávek u každého dítěte je 36 minut za měsíc. Korpus používá pro přepis [systém CHAT](#) (MacWhinney, 2000) a každá promluva je doprovázena řádkem s morfologickou anotací:

\*CHI: to je speciální letadlo , který umí hasit a stříká .

%xmor: pro:dem|ten-1&SG&N v:cop|být-3&SG&ind&pres&akt&impf adj|speciální-1&SG&N n|letadlo-1&SG&N , pro:rel|který-1&SG&N v|umět-3&SG&ind&pres&akt&impf v|hasit-inf&impf conj:coord|a v|stříkat-3&SG&ind&pres&akt&impf .

Morfologická anotace byla provedena pomocí volně dostupného nástroje MorphoDiTa (Spoustová et al., 2009; Straková et al., 2014) a následně převedena do formátu zohledňujícího podobu morfologického tagu podle [systému MOR používaného v databázi CHILDES](#) (MacWhinney, 2000), který umožňuje prohledávání a analýzu pomocí [specializovaného softwaru CLAN](#) (MacWhinney, 2000). Morfologická anotace dětských promluv byla následně manuálně zkontrolována, zároveň byla doplněna o specifikaci jazykových jevů podstatných pro jazykový vývoj, avšak nediferencovaných automatickou anotací (např. rozlišení vztažných a tázacích zájmen, slovesa *být* ve funkci sponového, pomocného a plnovýznamového slovesa, sloves s modálním použitím). Úspěšnost automatické anotace byla na velmi vysoké úrovni (Chromá et al., 2024). Během porovnání s manuálně zkontrolovanou anotací u tří dětí z hlediska klasifikace slovních druhů a u jednoho dítěte s ohledem na anotaci hodnot morfologických kategorií se jako nejvíce chybová ukázala 1. klasifikace nesklonných slovních druhů (spojky, částice a citoslovce), 2. anotace homonymních tvarů ohebných slov a 3. anotace jevů typických pro dětský jazyk (tvarotvorné a slovotvorné inovace, idiosynkratická slova). Z toho důvodu byla morfologická anotace promluv dospělých manuálně kontrolována pouze výběrově.

Morfologická anotace umožňuje korpus Chroma prohledávat a analyzovat podle lemat, slovních druhů, hodnot morfologických kategorií nebo jejich kombinací ve výpovědi. Zároveň je díky ní možné posuzovat lexikální či morfosyntaktický vývoj u jednotlivých dětí pomocí různých měřítek, jako je počet různých slov (number of different words) či tzv. index produktivní syntaxe (Index of productive syntax) – srov. Matiasovitsová et al. (2023; 2024).

Příspěvek se po představení korpusu Chroma zaměří na způsob a specifika jeho morfologické anotace a možnosti jejího využití, včetně ukázek některých analýz.



## Bibliografie

- Chromá, A., Sláma, J., Matiasovitsová, K., & Treichelová, J. (2024). A morphologically annotated longitudinal corpus of spoken Czech child-adult interactions. *Language Resources and Evaluation*. <https://doi.org/10.1007/s10579-023-09710-y>
- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3<sup>rd</sup> Edition. Lawrence Erlbaum Associates.
- Matiasovitsová, K., Čechová, P., Sláma, J., Treichelová, J., & Smolík, F. (2023): The validity of a transcript-based measure of child language development in Czech. In xxx (Eds.), *BUCLD 47: Proceedings of the 47<sup>th</sup> Annual Boston University Conference on Language Development* (pp. 533–547). Cascadilla Press. <https://www.lingref.com/buclid/47/BUCLD47-44.pdf>
- Matiasovitsová, K., Čechová, P., Sláma, J., Homolková, K., & Smolík, F. (2024). Mean Length of Utterance in Czech Toddlers: Validity of Estimates and Comparison of Words, Morphemes, and Syllables. *Journal of Speech, Language, and Hearing Research*, 67(3), pp. 837–852. [https://doi.org/10.1044/2023\\_JSLHR-23-00251](https://doi.org/10.1044/2023_JSLHR-23-00251)
- Spoustová, D., Hajič, J., Raab, J., & Spousta, M. (2009). Semi-Supervised Training for the Averaged Perceptron POS Tagger. In A. Lascarides, C. Gardent, & J. Nivre (Eds.), *Proceedings of the 12th Conference of the European Chapter of the ACL* (pp. 763–771). Association for Computational Linguistics.
- Straková, J., Straka, M., & Hajič, J. (2014). Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition. In K. Bontcheva, & J. Zhu (Eds.), *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 13–18). Association for Computational Linguistics. <http://dx.doi.org/10.3115/v1/P14-5003>

## Základní argumenty ve světových jazycích optikou Universal Dependencies

Daniel Zeman (Univerzita Karlova, Matematicko-fyzikální fakulta)

Universal Dependencies (UD) (de Marneffe et al., 2021)<sup>23</sup> představují morfosyntakticky orientovaný přístup k lingvistické anotaci, založené na binárních závislostních relacích mezi větnými členy. Závislosti se anotují primárně mezi plnovýznamovými slovy, zatímco funkční slova se považují za nositele morfosyntaktických rysů, které typicky „patří“ k některému plnovýznamovému slovu. Tento přístup v UD zvyšuje paralelismus mezi syntaktickými strukturami v různých jazycích. Cílem projektu UD byla v jeho počátcích harmonizace existujících korpusů a usnadnění vícejazyčného počítačového zpracování jazyků. Postupně se ale UD proměnily v soustavně se zvětšující sbírku korpusů, obohacovanou o data z dříve nepokrytých jazykových rodin, a staly se tak neocenitelným zdrojem i pro srovnávací jazykovědu a typologii (Futrell et al., 2015; Alzetta et al., 2018; Levshina, 2019; atd.). Tato výhoda ještě více vynikne, když se jazykověda spojí s jazykovými technologiemi a využije parser natrénovaný na datech UD k analýze dalších dat.

Zatím poslední verze, UD 2.14, se může pochlubit pokrytím 161 jazyků z 31 rodin (popř. 154 jazyků a 28 rodin, jestliže pomineme zvláštní „rodiny“ pro kreolštiny, znakové jazyky a střídání kódů). Tato sbírka je nicméně stále poměrně nevyvážená: Největší rodina, indoevropská, je reprezentována 74 jazyky, zatímco 18 rodin má jen po 1 jazyku. Ještě méně vyvážené jsou UD co do velikosti dat: indoevropské jazyky mají 24 mil. slov, uralské 1 mil., přes 100 tis. slov najdeme ještě v dalších 7 rodinách (afroasijská, turkická, sinotibetská, korejská, japonská, austronéská, baskická), zatímco na druhém konci škály je 9 rodin, které zatím nepřekročily laťku 10 tis. slov. Toto omezení je třeba mít na paměti a UD není možné považovat za proporční odraz trendů ve světových jazycích. I tak jde ale o zajímavý zdroj dat pro srovnání jazyků, které už pokryty jsou.

Jednou z klíčových vlastností anotačního schématu UD je, že rozlišuje tzv. základní argumenty (*core arguments*, S, A, P) od nepřímých rozvití (*oblique dependents*). Tím se UD liší od jiných závislostních korpusů, ve kterých se rozlišují argumenty od adjunktů, popř. přímo sémantické role. Vychází se z předpokladu, že většina jazyků používá specifické strategie pro kódování a specifická gramatická pravidla cílící na základní argumenty (Andrews, 2007; Zeman, 2017), čímž je opodstatněno jejich zachycení pomocí vyhrazených typů závislostí. Strategie pro kódování základních argumentů pracují s prostředky z poměrně malé množiny (slovosled, pádové morfémy, morfologie slovesa), nejsou však ve všech jazycích stejné. Přesná kritéria pro rozlišení základních argumentů od nepřímých rozvití musí být stanovena pro každý jazyk zvlášť.

Díky normalizované sadě značek používané v UD můžeme nyní studovat strategie kódování základních argumentů v jednotlivých jazycích. V případě konkurujících si strategií v rámci jednoho jazyka pak můžeme kvantifikovat jejich využití. UD například obsahuje 10 jazyků s ergativními morfologickými rysy: baskičtinu, gruzínštinu, abazinštinu, hindštinu, urdštinu, gudžarátštinu, marátštinu, bhódžpurštinu, warlpiri a jupičtinu. Mnoho dalších jazyků v UD vykazuje nominativně-akuzativní morfologii a nemálo jazyků pády nemá. Kolem 100 jazyků má anotace slovesné morfologie, která odráží rysy (např. osobu, číslo nebo rod) podmětu; 17 z nich promítá do téhož slovesa rysy dvou nebo více základních argumentů (podmětu a předmětu/ů).

Data z UD jsou také potvrzením, že slovosled hraje při kódování základních argumentů v některých jazycích větší roli než v jiných. Namísto striktní klasifikace jazyků jako SVO, SOV atd. můžeme pozorovat trendy a vyhodnotit, jak silně se ten který jazyk přiklání ke konkrétnímu slovoslednému typu. Obr. 1 je příkladem takového pozorování a ukazuje, že některé jazykové rodiny silně tíhnou ke konkrétnímu typu, zatímco jiné jsou rozprostřenější nebo zcela nevyhraněné, pokud jde o pořadí argumentů.

<sup>23</sup> <https://universaldependencies.org/>



19

*Speech and language  
development in children  
acquiring Czech*

organizers

Filip Smolík

Kateřina Chládková

(ENG)

# Workshop description

The workshop will examine the processes and mechanisms of speech and language acquisition in children acquiring Czech. Compared to the more widely studied languages in developmental literature, Czech has unique features at all linguistic levels, whose acquisition has not yet been fully understood (or not even considered in studies so far). At the level of phonetics and phonology, it has phonological vowel length and lacks word-level stress in connected speech, and for both these characteristics there is little data in the literature to make inferences about how they are acquired (Tsuji & Cristia 2014, Peperkamp 2004, Singh et al. 2023). The workshop will cover these phenomena. Embracing the morphological complexity of Czech, the workshop will touch upon broader questions that relate to the acquisition of a language with strong syntax-morphology interactions.

The study of language acquisition in Czech not only helps to understand language development and language representation in this particular language but addresses topics that have important implications both in the theoretical and applied domains. By focusing on the acquisition of a relatively understudied language (Kidd & Garcia 2022), this workshop will contribute to clarifying the basic questions about how knowledge of language is implemented in the human mind, and at the same time, it will provide data about the role of language in the developing social and cognitive mechanisms in infants and children, including children that go through typical as well as atypical developmental trajectories. The workshop will include, but will not be limited to, topics such as:

- Early speech perception, language-specific attunement to suprasegmental and segmental information in speech.
- The earliest stages of grammatical knowledge in children: sensitivity to grammatical number, gender, case marking and word order, as demonstrated in comprehension or early productions.
- Assessment of language development, tools available for examining children's language.
- Corpora of child Czech and similar languages, comparative work using child language corpora.
- Disorders of language development, linguistic structures that can serve as markers of language development.
- Semantic and formal factors in lexical acquisition: how factors such as imageability, semantic complexity, phonotactic probability or morphological complexity affect the age of word acquisition.

The presented studies will not only deal with a wide range of linguistic levels but will also cover a variety of methods used for assessing child language development: from the more traditional behavioural looking tasks or picture naming tasks to eyetracking paradigms and measures of neural activity. The attendees will get insights into the latest advancements of the study of speech and language acquisition with particular focus on the specificities of Czech. The individual presentations will cover various areas of language development and will complement each other such that a comprehensive picture of the early development of Czech will become available. Apart from the presentations listed below, and additional ones that will be contributed in the submission & review process, the workshop will include a final 15-min closing and discussion slot, led by the workshop organisers with contributions from the plenary speaker and other selected speakers.

## References

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, 42(6), 703-735.

Peperkamp, S. A. (2004). Lexical exceptions in stress systems: Arguments from early language acquisition and adult speech perception. *Language*, 80(1), 98-126.

Singh, L., Cristia, A., Karasik, L. B., Rajendra, S. J., & Oakes, L. M. (2023). Diversity and representation in infant research: Barriers and bridges toward a globalized science of infant development. *Infancy*.

Tsuji, S., & Cristia, A. (2014). Perceptual attunement in vowels: A meta-analysis. *Developmental Psychobiology*, 56(2), 179-191.

# Table of Contents

*Reference to the speaker and the addressee in Czech children: longitudinal perspective on spontaneous and elicited production*

Anna Chromá (Charles University)

*Newborns' processing of rhythm in native-language speech: an fNIRS experiment with Czech-learning infants*

Martina Dvořáková<sup>1</sup>, Natálie Kikořová<sup>1</sup>, Josef Urbanec<sup>2</sup>, Antonia Goetz<sup>3</sup>, Kateřina Chládková<sup>1</sup>  
(<sup>1</sup>FF UK a PSÚ AVČR, <sup>2</sup>LF HK UK a Nemocnice Havlíčkův Brod, <sup>3</sup>Western Sydney University)

*Language-based social preferences in Czech-learning infants and toddlers: a toy-choice experiment*

Lucie Jarůšková (Faculty of Arts, Charles University)

*Voicing in Czech sibilants: what children say, what adults hear*

Tanja Kocjančič, Tomáš Bořil (Faculty of Arts, Charles University)

*Sentence imitation and its relation to language skills in children with developmental language disorder*

Klára Matiasovitsová (FF UK), Filip Smolík (FF UK, PsÚ AV ČR)

*Verbal and conceptual development in Czech-learning children with cochlear implants*

Michaela Svoboda (Faculty of Arts, Charles University; Institute of Psychology, Czech Academy of Sciences)

*The mouse is pulling the hedgehog. Or the other way around? Role of case marking and word order in interpreting subject and object in Czech and German four-year-olds*

Jolana Treichelová<sup>1</sup>, Anna Chromá<sup>1</sup>, Filip Smolík<sup>1,2</sup>, Claudia Friedrich<sup>3</sup> (<sup>1</sup>Charles University, <sup>2</sup>Institute of Psychology CAS, <sup>3</sup>University of Tuebingen)

## Reference to the Speaker and the Addressee in Czech Children: Longitudinal Perspective on Spontaneous and Elicited Production

Anna Chromá (Charles University)

**Background:** Speaker (Spe) and Addressee (Add) are two key conversational roles and perspectives that shape human language interactions and are reflected in grammatical structure of languages. Reference to Spe and Add is typically realized by 1<sup>st</sup> and 2<sup>nd</sup> person singular pronouns and verb forms. In languages like Czech (pro-drop, without an obligatory subject but with person and number agreement on verb), the verb form is often used alone without the pronoun. In utterances produced by or directed to children, nouns (names) along with 3<sup>rd</sup> person verb forms are often used to refer to Spe and Add as well, which is a highly salient feature of this register. Acquiring the shifting system of 1<sup>st</sup> and 2<sup>nd</sup> grammatical person is challenging for first-language learners and the usage of non-shifting reference by names in children's input is probably helpful in this process (Smiley et al. 2011).

**Method:** The present study is concerned with spontaneous and elicited production of Spe and Add reference in Czech children in a longitudinal perspective. Sixty toddlers were seen in a laboratory setting at two time points at mean age of 29 and 43 months. Person reference produced by them was analyzed from three sources: two structured tasks and a 15-minute-long sample of spontaneous-play interaction with a close caregiver (mostly mother). In the tasks, children could have produced either names or pronouns in reference to themselves as the speakers or to the examiner as the addressee. In the Photo task at the first run, two pictures of the child and the examiner were taken and displayed at a screen; the child was asked who is in the picture. In the Kissing task at both runs, actions with toy animals were acted out and directed either towards the examiner or the child; the child was asked, e.g., who did the frog kiss. Ratio of (1) pronouns produced, (2) names produced, and (3) sum of pronouns and names produced (out of all stimuli) were calculated and used as measures of elicited performance.

From the spontaneous samples, all occurrences of names, pronouns, and verbs referring to the Spe and Add were extracted, their frequency was counted and numbers of tokens per utterance were used as measures of spontaneous performance. As a control, well-established measure of general linguistic development, mean length of utterance in words (MLUw, Brown 1973) was calculated from the samples as well.

**Results:** Table 1 shows a large difference in production of names within the spontaneous vs elicited context. In the spontaneous production at age 2.5 and later, names are very rare, especially in reference to Spe. In Figure 1, a weak tendency can be seen of names in Add reference to be negatively related to the concurrent MLUw. Spontaneously produced pronouns and verb forms are positively related to the concurrent MLUw (see Figure 1). In the elicited production, names are frequent, and pronouns alone relate neither to the concurrent MLUw nor to any (elicited) measure in the future. Only the sum of pronouns and names seems to be relevant in the elicited context (see Figure 2 for the relation to the concurrent MLUw).

In the spontaneous as well as elicited production, a pattern can be observed that early Spe reference (from the first run) does not predict well itself in the future (second run), but it predicts future MLUw as well as future Add reference. Early spontaneous Add reference is not related to the future measures (it has also very low frequency), however, early elicited Add reference predicts well itself, future Spe reference and MLUw as well.

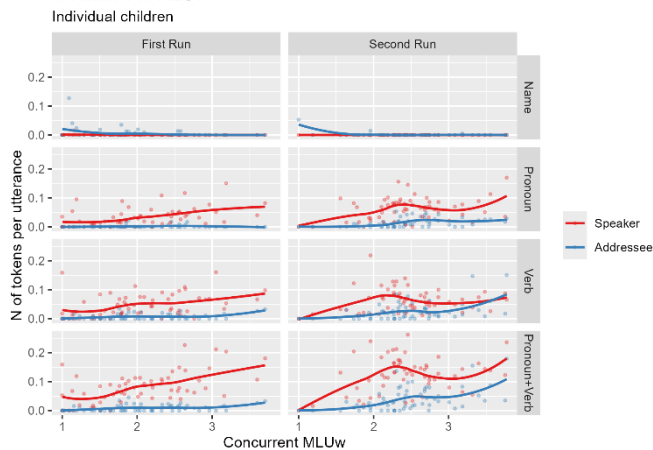
**Conclusions:** Spontaneous Spe and Add reference by names in Czech 2.5-year-olds may indicate lower level of linguistic development but in an eliciting task, both names and pronouns function together and indicate advanced development. Even at 3.5, names are still important referring means. Spontaneous Add reference at 2.5 is much less frequent than Spe reference, but the future performance in both Spe and Add reference is predictable from the frequency of Spe reference alone.



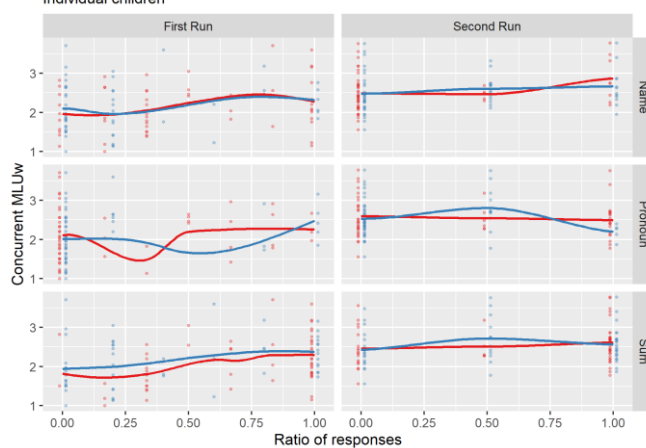
	Sample Spontaneous				Photo task Elicited		Kissing task Elicited	
		Name	Pronoun	Verb (without pronoun)	Name	Pronoun	Name	Pronoun
		Spe	Add	Spe	Add	---	---	---
<b>First run</b>	Spe	1	47	49	43	12	24	19
<b>(29 months)</b>	Add	12	7	22	21	6	16	15
<b>Second run</b>	Spe	0	56	59	---	---	15	25
<b>(43 months)</b>	Add	3	39	44	---	---	26	16

**Table 1** Number of children (out of 60) producing at least one token of each category.

**Figure 1** Spontaneous production: pronouns, names, and verb forms in relation to MLUw



**Figure 2** Elicited production: pronouns and names in relation to MLUw



## References

Brown, R. (1973). *A first language: The early stages*. Harvard University Press.

Smiley, P. A., Chang, L. K., & Allhoff, A. K. (2011). Can Toddy give me an orange? Parent input and young children's production of I and you. *Language learning and development*, 7(2), 77–106.

**Newborns' processing of rhythm in native-language speech:  
an fNIRS experiment with Czech-learning infants**

Martina Dvořáková<sup>1</sup>, Natálie Kikořová<sup>1</sup>, Josef Urbanec<sup>2</sup>, Antonia Goetz<sup>3</sup>, Kateřina Chládková<sup>1</sup>  
(<sup>1</sup>FF UK a PSÚ AVČR, <sup>2</sup>LF HK UK a Nemocnice Havlíčkův Brod, <sup>3</sup>Western Sydney University)

Virtually any child with normal hearing develops spoken language. The human brain seems to be specialised to speech already at birth such that newborn infants discriminate between speech and non-speech sounds and preferentially listen to speech [1,2]. Newborns have also been shown to discriminate between their native language and an unfamiliar, foreign language [3,4]. Arguably, the newborns' ability to discriminate between languages stems from their sensitivity to global prosodic cues, such as melody and rhythm [5]. As temporal and intonational speech cues are veridically transmitted to the intrauterine environment, humans have several weeks to months prior to birth to sensitise to the rhythm and melody of the ambient language [6,7]. Studies on newborn language discrimination to date tested the infants' sensitivity to rather salient language differences, such as forward vs. backward speech or distinct languages with very different rhythm [3,4,5]. We hypothesised that if humans sensitise to native-language rhythm patterns *in utero*, they should recognize those patterns at birth even if they merely represent a rhythmic deviation within the native language played forward. To this end, we tested whether newborns can discriminate between native-language speech with typical rhythm and native-language speech with atypical rhythm solely based on the temporal rhythmic differences.

Forty-two Czech-learning newborns (aged 1–5 days) were tested while asleep, listening to natural recordings of rhythmically typical, i.e. native-accented, and rhythmically atypical, i.e. non-native, Czech. The rhythmically non-native Czech had enlarged temporal ratios between stressed and unstressed syllables. (The stimuli were adopted from prior work demonstrating behavioural discrimination in Czech 4- and 6-month olds [8].) Each infant listened to 12 blocks of native-accented and 12 blocks of non-native accented Czech, in pseudo-random order. The blocks were between 15 and 17 s long and were interspersed by 27- to 32-s long blocks of silence. The newborns' neural activation in response to the speech materials was assessed using functional near-infrared spectroscopy (fNIRS) that measured changes in hemodynamic activity across 20 channels in the fronto-temporal regions bilaterally. The data has now been preprocessed using the Homer3 [9] Matlab toolbox, employing steps common in fNIRS speech perception studies with newborns [10]. Statistical analyses are underway.

Visual inspection of the data averaged per condition, across infants, indicates a different pattern of hemodynamic response to the native-accented and non-native accented blocks. The response appears to have a somewhat larger amplitude and a more canonical shape in the native-accented condition than for the non-native condition. However, these are only impressionistic descriptions and final interpretation and conclusions will be drawn after statistical modelling; these will be presented at the conference.

The present experiment will contribute to the current knowledge because it investigates newborns' rhythm discrimination of continuous naturally produced speech, extending previous research that tested newborns neural processing of prosodic rhythm differences with isolated words [11]. The results will be further discussed in relation to prior behavioural findings on rhythm perception with Czech-learning 4- and 6-month olds [8,12].

## Bibliography

- [1] Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental science*, 10(2), 159-164.
- [2] Chládková, K., Urbanec, J., Skálová, S., & Kremláček, J. (2021). Newborns' neural processing of native vowels reveals directional asymmetries. *Developmental Cognitive Neuroscience*, 52, 101023.
- [3] May, L., Gervain, J., Carreiras, M., & Werker, J. F. (2018). The specificity of the neural response to speech at birth. *Developmental science*, 21(3), e12564.
- [4] Pena, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences*, 100(20), 11702-11705.
- [5] Ramus, F. (2002). Language discrimination by newborns: Teasing apart phonotactic, rhythmic, and intonational cues. *Annual Review of Language Acquisition*, 2(1), 85-115.
- [6] Eggermont, J. J., & Moore, J. K. (2011). Morphological and functional development of the auditory nervous system. In *Human auditory development* (pp. 61-105). New York, NY: Springer New York.
- [7] Granier-Deferre, C., Bassereau, S., Ribeiro, A., Jacquet, A. Y., & DeCasper, A. J. (2011). A melodic contour repeatedly experienced by human near-term fetuses elicits a profound cardiac reaction one month after birth. *PLoS One*, 6(2), e17304.
- [8] Paillereau, N., Podlipský, V. J., Smolik, F., Šimáčková, Š., & Chládková, K. (2021). The development of infants' sensitivity to native versus non-native rhythm. *Infancy*, 26(3), 423-441.
- [9] Huppert, T., Diamond, S., Franceschini, M., Boas, D. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied optics* 48(10). <https://dx.doi.org/10.1364/ao.48.00d280>
- [10] Sato, H., Hirabayashi, Y., Tsubokura, H., Kanai, M., Ashida, T., Konishi, I., ... & Maki, A. (2012). Cerebral hemodynamics in newborn infants exposed to speech sounds: A whole-head optical topography study. *Human brain mapping*, 33(9), 2092-2103.
- [11] Abboub, N., Nazzi, T., & Gervain, J. (2016). Prosodic grouping at birth. *Brain and language*, 162, 46-59.
- [12] Chládková, K., Podlipský, V., Nudga, N., Paillereau, N., Kynčlová, K. & Šimáčková, Š. (2023). Infants' learning of novel segments is modulated by prosody. *Proceedings of the 20th International Congress of Phonetic Sciences*, pp. 2358-2362.

**Language-based social preferences in Czech-learning infants and toddlers:  
a toy-choice experiment**

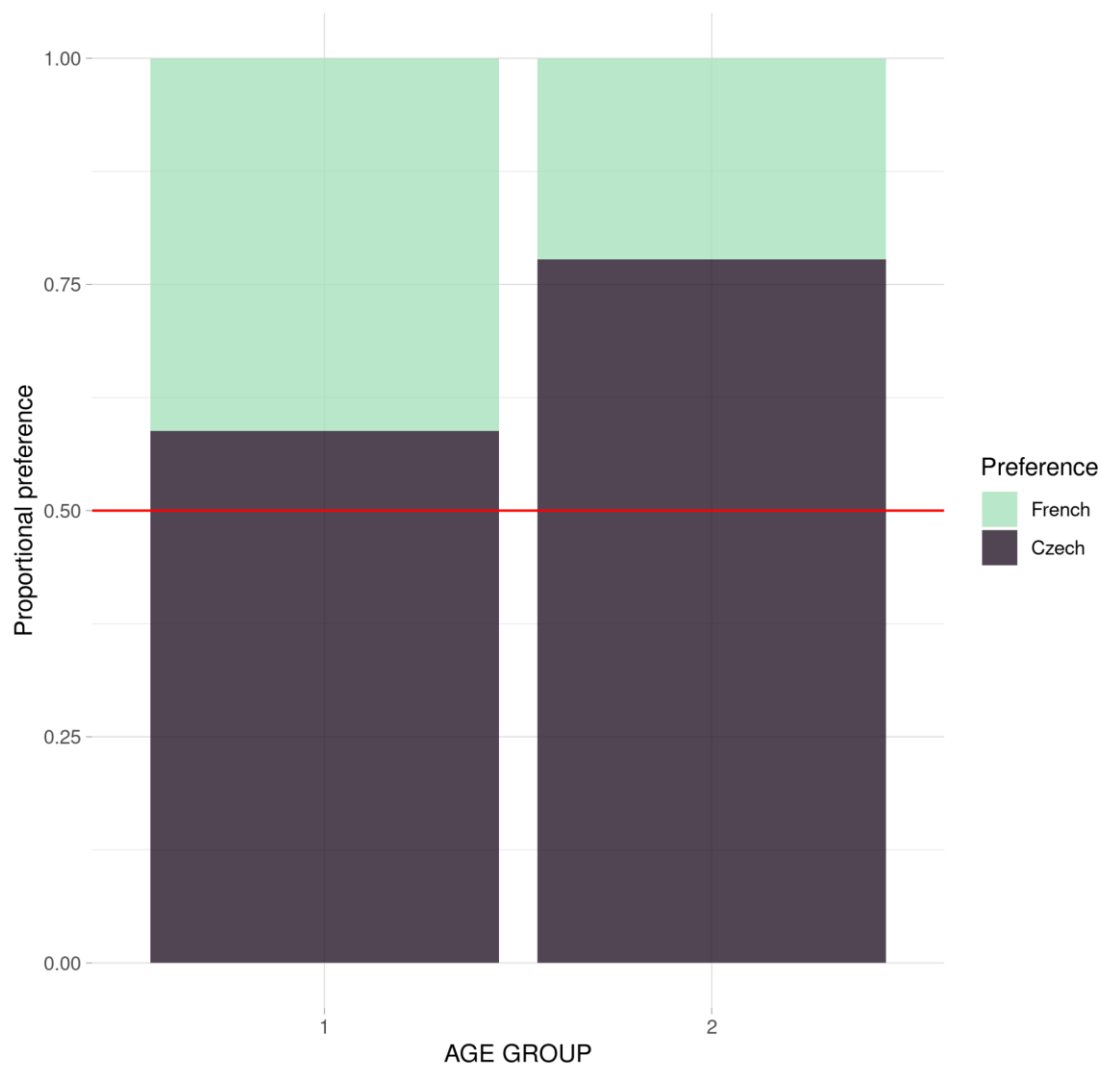
Lucie Jarůšková (Faculty of Arts, Charles University)

**Background** The way we speak, such as our language and dialect, are part of our identity. Atypical or unfamiliar-sounding speech is easy to perceive from an early age and affects our social behavior and cultural learning [1]. It has been repeatedly shown that perceptual biases favoring L1- (known as native) over L2- (second-language) talkers are present from early childhood. Infants are more likely to take a toy or food from an adult who speaks their native language rather than from a speaker of a foreign language. This observation was replicated with dialects: 12-month-old Japanese-learning children preferred to take a toy from a talker who spoke the same local dialect as the child's caregivers (i.e. the child's native dialect) than from a talker who spoke a dialect unfamiliar to the child [2]. At preschool age, children in laboratory settings prefer to be friends with L1-accented peers over with L2-accented ones [3]. Some studies indicate that multilingual exposure may attenuate these biases [4], while others fail to find such effects [5]. Multilingual exposure might reduce language-based social biases, for instance, because the child is used to interacting with people who speak with various atypical accents, or because the child themselves is being brought up with multiple languages. In the Czech-speaking society, the preference for L1-accented friends later in life has been replicated in preschool children speaking Czech and growing up in a relatively linguistically homogeneous community, but the accent-based social bias was significantly reduced in bilingual children [6]. Here we ask whether the bias towards L1- or L2-accented talkers is present at a much earlier age, namely in Czech-learning infants and toddlers, and whether it is modulated by the infants' bilingual exposure. Studying language-based social preferences in a so-far understudied language group and culture, will contribute new insights as to the role of accent and language in cultural learning and development of in-group biases across early childhood.

**Method** The present study employed a behavioral toy-taking experiment, inspired by [3]. Two age groups were recruited: monolingual and bilingual infants at 11–13 months of age and toddlers at 24–30 months of age. Bilingual language background was determined from a parental questionnaire about language and accent exposure. During the experiment, the child sat in a parent's lap behind a table, in front of two monitors. Each of the monitors displayed a video of one of two talkers, an L1-accented woman and an L2-accented woman. The talkers appeared in turns, saying identical phrases (a greeting and two poems), and at the end, they both simultaneously handed over a toy. The toys appeared in their physical form beneath the monitors so that the child could reach and take one of them. If the child did not perform any action for 10 seconds, they were prompted by the parent to take one of the toys. The child's reaction was videotaped. The order and side of presentation of the talkers were counterbalanced across children. In line with prior work (e.g. [2], [3]) we predicted that children will more often select a toy from an L1-accented talker than from an L2-accented talker. Based on our prior findings with preschool children, we predicted that the effect of accent will be smaller in bilingual than in monolingual group.

**Preliminary results** Data collection is currently underway, with data from 17 monolingual infants and 18 monolingual toddlers already collected (the aim is to collect  $n = 28$  in each age group and each language group). Statistical analysis (model *glm*, R Studio) revealed a significant preference for the native-accented talker ( $p = .0254$ ); see Figure 1.

**Conclusions** The present study will provide insights on accent-based pro-social behavior in infants and toddlers acquiring Czech, which is a to-date understudied population. Preliminary data suggest a preference for L1-accented talkers. The study will further reveal whether and how the accent-based preference is modulated by bilingual experience. We will further discuss how social preferences based on language develop in childhood and how we can assess them.



**Figure 1.** Preliminary results: Counts of toy-taking preferences in 1-year-old and 2-year-old children ( $N = 35$ ).

## Bibliography

- [1] Shutts et al. (2009). Social information guides infants' selection of foods. *Journal of Cognition and Development*, 10(1-2).
- [2] Okumura et al. (2014). [Twelve-month-old infants show social preferences for native-dialect speakers]. *Shinrigaku kenkyu : The Japanese journal of Psychology*, 85(3).
- [3] Kinzler et al. (2007). The native language of social cognition. *PNAS*, 104(30).
- [4] Spence et al. (2021). Something about the way you speak: A meta-analysis on children's linguistic-based social preferences. *Child Development*, 92(2).
- [5] Byers-Heinlein et al. (2017). Monolingual and bilingual children's social preferences for monolingual and bilingual speakers. *Developmental Science*, 20(4).
- [6] Jarůšková & Chládková. (2023). Preference for native-accented peers in monolingual and bilingual children acquiring Czech, *Journal of Multiling. and Multicul. Development*, 1-12.

**Voicing in Czech sibilants: what children say, what adults hear**

Tanja Kocjančič, Tomáš Bořil (Faculty of Arts, Charles University)

Difficulties with the production of Czech sibilants /s, z, ʃ, ʒ/ are one of the main reasons for children entering speech and language therapy. During the assessment, a child's production is perceptually evaluated, and any observed errors are marked. They are typically classified in terms of place of articulation (e.g. interdental, lateral) and loss of voicing (Anjos et al., 2018; Neubauer, 2018). While the identification of the place of articulation error is supported by visible articulatory mechanisms, guidelines for the recognition of voicing errors are lacking. In spontaneous adult speech, Czech voiced sibilants are relatively consistently fully voiced, however partial or full devoicing has been observed in 8 % of /z/ and 10 % of /ʒ/ (Skarnitzl, 2011). It can be expected that the realization of voicing in children's speech is significantly more varied due to the ongoing growth and maturation of the speech-motor system (Koenig et al., 2008). A better understanding could lead to more reliable identification of voicing errors in clinical speech assessment. However, to improve assessment, we also need to understand how well adults perceive voicing in children's fricatives, and how they respond to different voicing errors. A recent study of whispered adult Czech has shown that listeners are successful in identifying sibilants even when they are all produced voiceless (Svatošová & Bořil, 2023). The aim of the study was twofold: (1) how do Czech children realize voicing in sibilants, and (2) how adults perceive this voicing?

Ten typically developing monolingual Czech children aged between 3;5 and 5;11 performed a picture-naming task in which they produced four repetitions of six isolated words per sibilant, resulting in 857 productions. The target consonants were all in a word-initial position in CV or CCV syllables. The realization of voicing was measured as F0 and evaluated in terms of its presence within the duration of a sibilant (expressed as a percentage of voicing). The analysis uncovered two interesting results. First, voiced, and voiceless sibilants differed in the amount of fully voiced/voiceless productions. 64 % of all /z/ and 60 % of /ʒ/ were fully voiced, while 47 % of /s/ and 42 % of /ʃ/ were fully voiceless. Second, they also differed in the amount of present voicing. The productions of /z, ʒ/ targets included examples ranging from fully voiced to fully voiceless. The percentage of voicing in the /s, ʃ/ targets ranged from 0 – 70 %, however only 17 out of 428 productions had voicing present in more than 40 % of the segment duration. These results highlight a greater variability of voicing realization in children's sibilants. To furthermore improve voicing error detection in clinical practice, we need a better understanding of how adults perceive different voicing realizations. To explore adult perception of children's sibilants, we are preparing a two-alternative forced choice (2AFC) perception test exploring the differences in voicing and place of articulation of the target sibilants. Test items consist of real children's examples of sibilants in CV syllables, as well as synthesized examples with manipulated ratio of voiced/voiceless duration within the target fricatives. The results of these experiments will contribute to our understanding of the difficulties with auditory perception of voicing and voicing error detection in a clinical speech assessment.

## Bibliography

Anjos, I., Grilo, M., Ascensão, M., Guimarães, I., Magalhães, J., & Cavaco, S. (2018). A model for sibilant distortion detection in children. In Proceedings of the 2018 international conference on digital medicine and image processing (pp. 42-47).

Koenig, L. L., Lucero, J. C., & Perlman, E. (2008). Speech production variability in fricatives of children and adults: Results of functional data analysis. *The Journal of the Acoustical Society of America*, 124(5), 3158-3170.

Neubauer, K. (2018). Vývojové a přetrvávající poruchy artikulace a fonologického rozlišování hlásek. In K. Neubauber (Ed.), *Kompendium klinické logopedie: diagnostika a terapie poruch komunikace* (pp.316-341). Portal.

Skarnitzl, R. (2011). Znělostní kontrast nejen v češtině. Epoque. Praha.

Svatošová, M., & Bořil, T. (2023). Duration as a cue for phonological voicing contrast in whispered Czech. Proc. 20th ICPHS.

**Sentence imitation and its relation to language skills  
in children with developmental language disorder**

Klára Matiasovitsová (FF UK), Filip Smolík (FF UK, PsÚ AV ČR)

Sentence imitation (SI) is a sensitive marker of developmental language disorder (DLD; Riches, 2012), revealing difficulties in syntactic structures, as children should only be able to imitate mastered syntactic structures (Poll et al., 2013). This study compares the ability of children with DLD and typically developing (TD) children to imitate, understand, and produce the same syntactic structures. The study also assesses the effects of general language knowledge on sentence imitation performance. Considering that children with DLD often struggle with relative clauses (RCs; Adani et al., 2016, Frizelle & Fletcher, 2014), the research focuses on their ability to imitate RCs compared to adjectival nominal phrases. In Czech, these structures vary in morphosyntactic complexity but share some inflectional characteristics (both the wh-word *který* 'which' and the adjectives follow adjectival declension), allowing us to distinguish between problems with inflection and overall difficulties with processing complex syntax (Stavrakaki, 2015).

Sixty-three monolingual Czech-speaking children with DLD (6;5-9;6,  $M = 8;10$ ) with receptive vocabulary below the 16th percentile were gender- and vocabulary- matched with TD children (3;7-6;7,  $M = 6;1$ ). They completed the SI task involving 32 RCs and 16 simple sentences with adjectival phrases. First, the SI was scored for the number of changes (with 5 points for more than 4 changes). Then, the grammaticality of using relative pronouns/adjectives within the core of the structure was scored, and the used avoidance strategies and types of errors were analyzed. The children also completed a truth-value judgment (comprehension) task and an elicited production task involving 16 items from both structures.

First, a cumulative link mixed model revealed a significant interaction between the effects of vocabulary, group, and construction ( $z = -2.10$ ,  $p = .036$ ), suggesting that the relation between scores in SI and vocabulary is weaker in children with DLD compared to TD children, but stronger in children with DLD when imitating relative clauses (see Figure). The analyses also revealed the unique effects of comprehension and production of relative clauses on the number of changes in imitation of both syntactic structures, above and beyond the effect of the vocabulary task. The number of changes decreased with better scores (see Table). Generalized linear mixed-effects models revealed a greater tendency in children with DLD to avoid the target structure ( $z = -3.076$ ,  $p = 0.002$ ), and a significant interaction between the effects of group and construction ( $z = -3.006$ ,  $p = 0.003$ ), suggesting that children with DLD avoided sentences with RCs more often than adjectival phrases compared to TD children. Children with DLD also used the target relative pronouns/adjectives less grammatically ( $z = -5.099$ ,  $p < 0.001$ ). However, the difference between the scores for both structures was smaller compared to TD children (interaction group\*construction:  $z = 2.564$ ,  $p = 0.010$ ). During the imitation of RCs, the most frequent avoidance strategies were using RCs modifying an isolated noun phrase (especially in TD children) and simple sentences (see Adani et al., 2016). Children with DLD exhibited higher rates of wh-word omissions (consistent with Contemori & Garraffa, 2010) and non-inflected wh-word substitutions (214 cases vs. 94 cases) than TD children (see Diessel & Tomasello, 2005).

The results revealed the relation between both general and specific language knowledge and sentence imitation performance, although the effect of general knowledge differed between children with DLD and TD children. Proficiency in both comprehension and production of specific morphosyntactically complex structures was related to the repetition of both those structures and simpler ones that share similar features. More common avoidance of RCs than adjectival phrases and omissions of relative pronouns indicated greater syntactic difficulties with RCs in children with DLD. Challenges with adjectival declension were suggested by non-inflected wh-word substitutions and more frequent ungrammatical use of the target wh-word/adjective in the core of the sentence. The most common avoidance strategies and error types used by children with DLD were consistent with previous research.



Figure. Relations between sentence imitation of relative clauses and simple sentences with adjectival phrases, vocabulary and group. The darkest and lightest band show the probability of zero-error or five-plus error responses, respectively. Increases in the dark blue bands show the increase of probability for imitations with lower number of errors.

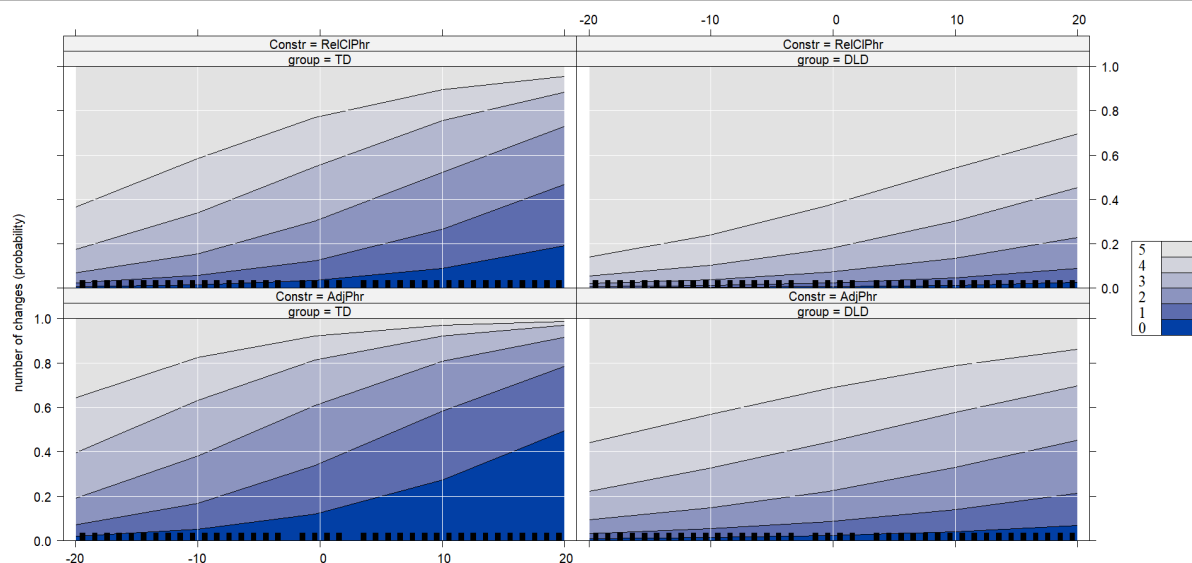


Table. Relations between imitation of relative clauses and simple sentences with adjectival phrases, their use in elicited production (EP) and truth value judgement task (TVJT) and vocabulary. No significant interactions between variables.

Predictors	SI (AdjPhrs)		SI (RCs)	
	Estimate	p	Estimate	p
group [DLD]	0.82	<b>0.004</b>	0.71	<b>0.007</b>
TVJ (AdjPhrs)	-0.89	0.150	-0.12	0.838
TVJT (RCs)	-1.90	<b>0.002</b>	-1.32	<b>0.023</b>
EP (AdjPhrs)	-0.00	0.982	-0.00	0.907
EP (RCs)	-0.07	<b>&lt;0.001</b>	-0.07	<b>&lt;0.001</b>
vocabulary	-0.03	<b>0.045</b>	-0.03	<b>0.015</b>

## Bibliography

- Adani, F., Stegenwallner-Schütz, M., Haendler, Y., & Zukowski, A. (2016). Elicited production of relative clauses in German: Evidence from typically developing children and children with specific language impairment. *First Language*, 36(3), 203-227.
- Contemori, C., & Garraffa, M. (2010). Comparison of modalities in SLI syntax: A study on the comprehension and production of non-canonical sentences. *Lingua*, 120(8), 1940-1955.
- Diessel, H., & Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language*, 882-906.
- Frizelle, P., & Fletcher, P. (2014). Relative clause constructions in children with specific language impairment. *International Journal of Language and Communication Disorders*, 49(2), 255–264. <https://doi.org/10.1111/1460-6984.12070>
- Poll, G. H., Miller, C. A., Mainela-Arnold, E., Adams, K. D., Misra, M., & Park, J. S. (2013). Effects of children's working memory capacity and processing speed on their sentence imitation performance. *International Journal of Language and Communication Disorders*, 48(3), 329–342.
- Riches, N. G. (2012). Sentence repetition in children with specific language impairment: An investigation of underlying mechanisms. *International Journal of Language and Communication Disorders*, 47(5), 499–510.
- Stavrakaki, S., Tasioudi, M., & Guasti, T. (2015). Morphological cues in the comprehension of relative clauses by Greek children with specific language impairment and typical development: A comparative study. *International journal of speech-language pathology*, 17(6), 617-626.

**Verbal and conceptual development in Czech-learning children with cochlear implants**

Michaela Svoboda

(Faculty of Arts, Charles University; Institute of Psychology, Czech Academy of Sciences)

A cochlear implant (CI) is a neuroprosthetic device that provides hearing to individuals with sensorineural hearing loss. Especially in prelingually deaf children, its contribution to successful L1 acquisition is undeniable, although language outcomes vary widely within this population. The objective of our study is to assess the vocabulary development of implanted children of hearing parents in spoken Czech and in Czech sign language and to identify some of the factors that may contribute to the variation in language outcomes. Following Thal et al. (2007) who showed that the American-English CDI provides valid measures of language development in CI children throughout the first three years post-implantation, we use the Czech adaptation of MacArthur-Bates CDI:WS (Smolík et al. 2017). We adjusted the questionnaire to include answer options “says”, “says and signs”, “signs”. The questionnaire is aimed at early implanted children whose hearing age (HA) is 16 to 30 months. The questionnaires are being completed online by parents in quarterly intervals. To date we have collected data from sixteen children in a first session, seven of whom already contributed data also in a second session (mean age 45.5 months; HA 24 months at first assessment; mean age at first CI surgery 21 months). Data collection is still underway, by the time of the conference we expect to have data from at least one session from 30 children, with a second or even a third session from approximately half of them. The preliminary data confirm the great variability in outcomes observed in CI children acquiring other languages, with 3 children out of the current 16 performing above the 98th percentile of the hearing population matched for HA. The data will be discussed in more depth at the conference, following a quantitative analysis of the contributing factors and a qualitative inspection of the individual trajectories.

**Bibliography**

Smolík, F., Turková, J., Marušincová, K. & Malechová, V. (2017). Dotazník vývoje komunikace II. Dovyko II. Dotazník pro diagnostiku jazykového vývoje ve věkovém rozmezí 16 až 30 měsíců. Příručka a normy. Univerzita Karlova, Filozofická fakulta.

Thal, D., Desjardin, J. L., & Eisenberg, L. S. (2007). Validity of the MacArthur-Bates Communicative Development Inventories for measuring language abilities in children with cochlear implants. *American journal of speech-language pathology*, 16(1), 54–64. [https://doi.org/10.1044/1058-0360\(2007/007](https://doi.org/10.1044/1058-0360(2007/007)

**The mouse is pulling the hedgehog. Or the other way around? Role of case marking and word order in interpreting subject and object in Czech and German four-year-olds**

Jolana Treichelová<sup>1</sup>, Anna Chromá<sup>1</sup>, Filip Smolík<sup>1,2</sup>, Claudia Friedrich<sup>3</sup>  
(<sup>1</sup>Charles University, <sup>2</sup>Institute of Psychology CAS, <sup>3</sup>University of Tuebingen)

The preference for the canonical word order (often SVO) over the non-canonical one (often OVS) has been shown in many languages allowing for word-order variation, including Czech and German (Slobin & Bever 1982; Ditmar et al. 2008; Smolík 2015). However, Czech-learning children appear to comprehend the non-canonical OVS word order earlier than their German-learning peers. In a cross-linguistic study, we directly compare Czech- and German-learning children's comprehension of closely matched sentences with varying word orders.

We use identical pictorial stimuli, parallel Czech and German picture descriptions, and an identical experimental procedure. Each item combines one unambiguous noun (der Igel/ježek) and one ambiguous noun (die Maus/myš). There are 8 items and each of them appear in 4 conditions (see Table 1 below), resulting in 32 stimuli. Children are presented with all stimuli in pseudorandomized order, each time seeing two pictures with reversed role assignment and listening to one relevant verbal description (correct for one of the pictures), while having their gaze tracked. After each trial, children are asked to point at the correct picture.

Twenty three German-learning children (mean age 50.8 months, SD age 6.18) and 30 Czech-learning children (mean age 51.5 months, SD age 4.84) have been tested so far. Another 7 German-learning participants are to be tested yet.

For the initially-ambiguous sentences, neither gaze nor pointing of German-learning children do not differ across the word-order conditions (see both Figures below). The gaze direction of Czech-learning children suggest preference for the SVO interpretation across word-order conditions during processing and only the pointing results show that – to a certain extent – they finally arrive at the correct interpretation for each word order.

For the initially-unambiguous sentences, two separate models for both, gaze and pointing revealed significant effects of word order in German-learning children: the analysis of gaze using linear mixed-effects model yielded an effect of word order within the sentence-final time window (2ndNP in the figure): children spent less time looking at the SVO picture if the word order was OVS than if it was SVO. The analysis of pointing using logit mixed-effects model yielded a corresponding effect: children pointed less to the SVO picture if the word order was OVS than if it was SVO. Models for the Czech data were not run yet, however, both the figures below present clear evidence of large difference between the word-order conditions in Czech-learning children during processing (gaze) as well as in the final decision (pointing).

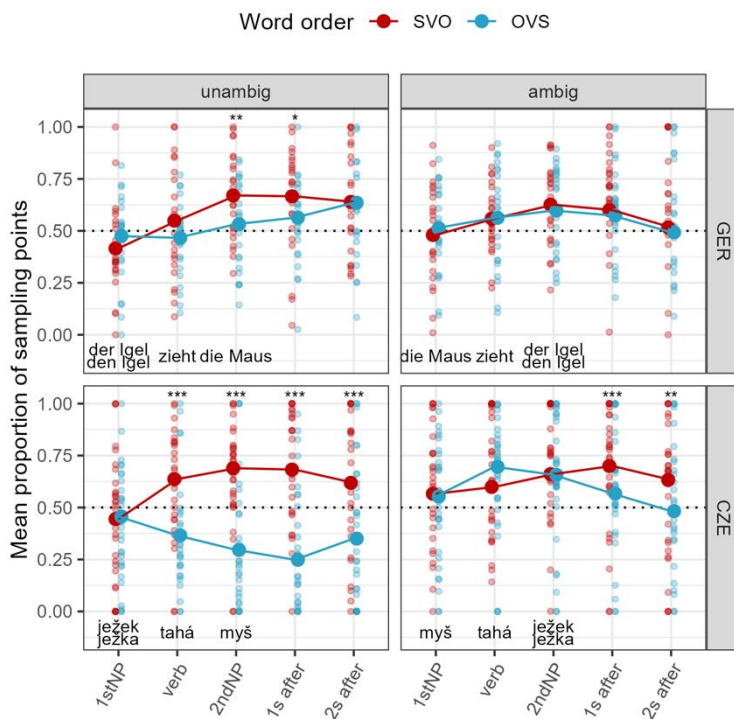
With an initial unambiguous morphological cue, German-learning four-year-olds can thus distinguish SVO from OVS word order, however the condition difference in looking as well as in pointing is quite small. It is clearly to see that Czech-learning children at the same age can use the morphological cue to a much larger extent.

The implicit (gaze direction) and the explicit (pointing) measure in our experimental setting mostly agree with each other. Nevertheless, in the initially-ambiguous sentences in Czech children, the two measures provide an interesting comparison: the correct distinction of conditions is observable only in the final decision (pointing) but not during processing (gaze).

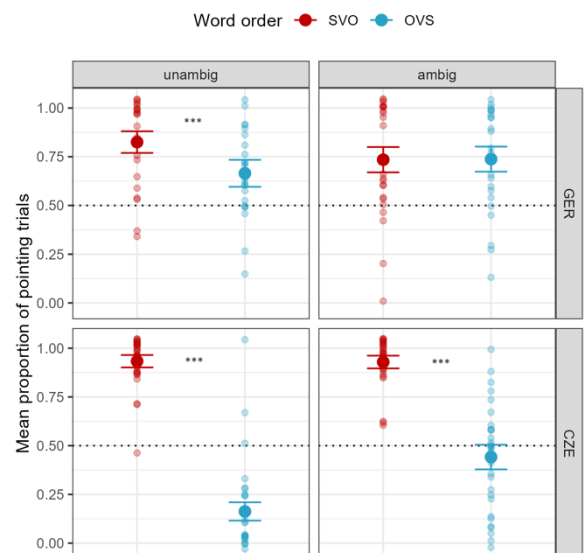
Condition	German	Czech	English translation
SVO initially unambiguous	Der Igel zieht die Maus.	Ježek táhne myš.	<i>The hedgehog is pulling the mouse.</i>
OVS initially unambiguous	Den Igel zieht die Maus.	Ježka táhne myš.	<i>The mouse is pulling the hedgehog.</i>
SVO initially ambiguous	Die Maus zieht den Igel.	Myš táhne ježka.	<i>The mouse is pulling the hedgehog.</i>
OVS initially ambiguous	Die Maus zieht der Igel.	Myš táhne ježek.	<i>The hedgehog is pulling the mouse.</i>

Table 1 The four conditions in one example item.

**Figure 1** Looking at the SVO picture *Individual participant means and group means. Asterisks mark significance level of the word-order effect in the respective model (\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ )*



**Figure 2** Pointing to the SVO picture *Individual participants and group means with 95% CI. Asterisks mark significance level of the word-order effect in the respective model (\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ )*



## Bibliography

- Slobin, D. I., & Bever, T. G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, 12(3), 229-265.
- Dittmar, M., Abbot-Smith, K., Lieven, E., & Tomasello, M. (2008). German children's comprehension of word order and case marking in causative sentences. *Child development*, 79(4), 1152-1167.
- Smolík, F. (2015). Word order and information structure in Czech 3-and 4-year-olds' comprehension. *First language*, 35(3), 237-25.

20

*L2 learning*  
*– a path to bilingualism*

*Osvojování dalšího jazyka*  
*– cesty k bilingvistice*

organizují / organizers

Šárka Šimáčková

Jonáš Podlipský

(CZE, ENG)

Bienále české lingvistiky 2024  
Biennial of Czech Linguistics 2024

# Workshop description / Popis panelu

**(ENG)** The theme of the panel is the path to bilingualism (linguistic multicompetence). We are interested in the dynamic changes of linguistic competence during the life of a bilingual individual. We ask questions about the development of bilingualism in childhood and adulthood and the resulting nature of bilingual language competence. The panel connects two main themes:

## 1. THE ROLE OF LANGUAGE EXPOSURE

The way in which an individual is exposed to a language fundamentally influences their linguistic competence at a given stage of language development (Luk & Bialystok 2013). There is a range of life experiences that can lead to the development of more than one language. Some people acquire multilingual competence ‚naturally‘ from exposure to the environment in which they grow up and are socialized (multilingual communities and/or families) or in which they find themselves later in life (emigration, study abroad). A completely different path to bilingualism is learning a foreign language at school. Bilingual language experience and the resulting linguistic multicompetence are therefore necessarily heterogeneous. Our panel will reflect on the importance of language immersion: if immersion leads to more complete language acquisition, is it possible to simulate immersion learning processes in a non-immersion context?

For example, we are interested in the following topics (though not exclusively):

- the relationship between language use, the process of language learning and the resulting linguistic multicompetence
- the influence of bi-vs.-uni-modal language experience on linguistic multicompetence and language processing (perception and production)
- the influence that the multilingual experience of an adult or child has on their social functioning in society.

Although we emphasize the importance of language exposure, we also want to give space to the idea that some language acquisition processes may be universal.

## 2. INTERLINGUAL INTERACTIONS

Arguments that the linguistic competence of bilingual speakers should be examined as a whole, not as a sum of separate monolingual competences, have been made in bilingualism studies (Grosjean 1989), second language acquisition theory (Cook 1995) and linguistic subfields such as phonology (Flege 1995) since the 1990s. Current empirical research demonstrates the interconnectedness of languages in the mind of the speaker across language plans (e.g., Kroll et al. 2012) and highlights the uniqueness of multilingual language competence and language processing (e.g., Kroll et al. 2014, Schmid & Köpcke 2017). The panel will offer a dynamic perspective on the linguistic competence of bilingual individuals.

Among other topics, we intend to address the following topics:

- Cross-linguistic interactions on all linguistic levels, differences between monolingual and bilingual linguistic competence and performance.

- Changes in the bilingual speaker's language system caused, for example, by prolonged residence in a second language environment and possible shifts in language dominance: Improvements in the L2 (e.g. due to study abroad), changes in the first language (the process of weakening of language competence (language attrition)).
- Short-term language shifts due to changes in the language environment. An example of such shifting is „phonetic drift“, which affects the sound of a bilingual's L1: exposure to a new language causes L1 sounds to become acoustically similar to equivalent sounds in the foreign language.

We understand the themes of language exposure and bilingual cross-language effects as connected in interesting ways and we believe that they can be approached from a variety of perspectives. Our goal as the panel organisers is to offer a platform where these perspectives can be discussed.

## References

- Cook, V. (1995). Multi-competence and the learning of many languages. *Language, Culture and curriculum*, 8(2), 93-98.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 92, 233-277.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and language*, 36(1), 3-15.
- Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Kroff, J. R. V. (2012). Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In *Psychology of learning and motivation* (Vol. 56, pp. 229-262). Academic press.
- Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current directions in psychological science*, 23(3), 159-163.
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25(5), 605-621.
- Schmid, M. S., & Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism*, 7(6), 637-667.

## Obsah / Table of Contents

*Komplexní predikáty s kategoriálními slovesy dělat, dávat, dostávat, brát a mít v češtině nerodilých mluvčích na úrovni B2–C1 – konstruktivistický pohled*

Andrea Hudousková (Ústav bohemistiky pro cizince a komunikace neslyšících, FF UK)

*L2 learners are more likely to imitate a native than a non-native speaker displaying native-like phonetic patterns*

Daniel Kopecký, Václav Jonáš Podlipský, Šárka Šimáčková (Palacký University Olomouc)

*Preschoolers' production of L2 vowels is affected by input quality: A longitudinal study*

Monika Kučerová (Institute of Psychology, The Czech Academy of Sciences), Šárka Šimáčková (Department of English and American Studies, Palacký University Olomouc)

*Statistical learning from visual speech cues in a novel artificial language*

Kateřina Kynčlová, Kateřina Chládková (Charles University, Czech Academy of Sciences)

*Cross-language phonological neighborhood effects*

Eva Maria Luef (University of Hamburg, Charles University)

*Gender processing in a non-gendered L2: The role of L1 grammar, gender stereotypicality and L2 proficiency*

Mikuláš Preininger<sup>1</sup>, James Brand<sup>2</sup>, Adam Kříž<sup>1</sup>, Markéta Ceháková<sup>1</sup> (<sup>1</sup>Charles University, <sup>2</sup>Edge Hill University)

*Korpus primární slovní zásoby češtiny jako dalšího jazyka z diachronního pohledu*

Hana Svobodová, Kateřina Malá, Simona Koryčánková (Masarykova univerzita)

*Using ultrasound in L2 Czech pronunciation teaching: /r/ and /l/ sounds*

Jitka Veroňková, Tanja Kocjančič (Charles University)



**Komplexní predikáty s kategoriálními slovesy *dělat, dávat, dostávat, brát a mít* v češtině nerodilých mluvčích na úrovni B2–C1 – konstruktivistický pohled**

Andrea Hudousková (Ústav bohemistiky pro cizince a komunikace neslyšících, FF UK)

Príspevek se zaměřuje na komplexní (analytické, verbonominální) predikáty s kategoriálními slovesy *dělat, dávat, dostávat, brát a mít* v češtině nerodilých mluvčích na úrovni B2–C1. Cílem výzkumu bylo zjistit, jakou roli hrají frekvence a míra asociace kolokátů při osvojování těchto predikátových struktur. Komplexní predikát je spojení predikativního jména s vlastním lexikálním významem se sémanticky oslabeným kategoriálním slovesem (Radimský, 2010; Kolářová, 2010; Kettnerová, 2017). Tyto predikátové struktury se nacházejí na pomezí syntaxe a lexika, a proto jejich kolokabilita a syntaktické vlastnosti nejsou plně předvídatelné.

Z hlediska konstrukční gramatiky jsou tedy komplexní predikáty typickými konstrukcemi, tj. takovým spojením formy a významu, v němž některé aspekty formy či významu nejsou přímo vyvoditelné z jednotlivých komponentů konstrukce ani z jiných konstrukcí (Goldberg, 1995, s. 4). Konstrukce jsou zde chápány jako postupně vznikající shluky informací v naší paměti, které jsou vzájemně propojovány na základě společných formálních a funkčních rysů (Goldberg, 2019, s. 6). Pro osvojování jazyka jsou pak klíčové dva parametry: a) jak důležitá je určitá lexikální jednotka (LJ) pro danou konstrukci (tj. jaký je její vztah k ostatním LJ, které se v konstrukci vyskytují) a b) jak důležitá je daná konstrukce pro určitou LJ (tj. v jakých jiných konstrukcích se tato LJ vyskytuje) (Goldberg, 2006, s. 94; Herbst, 2020, s. 68). Tyto údaje lze na základě korpusových dat aproximovat pomocí a) absolutní frekvence dané kolokace a b) asociční míry MI (mutual information), která udává těsnost kolokace neboli míru vázanosti LJ na tuto či jiné konstrukce (srov. obdobnou argumentaci v BNC).

Na základě těchto předpokladů lze stanovit následující hypotézy: a) čím frekventovanější LJ a jejich kolokace jsou, tím pravděpodobněji vytvářejí konstrukce (shluky vzájemně provázaných informací) v kognitivním prostoru nerodilého mluvčího; b) čím vyšší je MI kolokátu, tím více je tato LJ v kolokaci fixována a tím snáze je z paměti i vybavována.

Za účelem ověření těchto hypotéz byly kolokace sloves *dělat, dávat, dostávat, brát a mít* na základě údajů z korpusu SYN2020 převedeny na reprezentace zohledňující jejich frekvenci a MI (viz obr. 1). Z těchto rámců byly vybrány příklady s různými hodnotami těchto parametrů (viz zvýrazněné LJ v obr. 1). Na tomto základě byl vytvořen dotazník pro nerodilé mluvčí (celkem 20), který testoval jejich znalost kolokací s ohledem na frekvenci kolokace, frekvenci kolokátů a MI. Odpovědi byly vyhodnoceny s ohledem na přesnost (shodu s původním zněním), variabilitu správných možností a chybovost. Provedený výzkum vedl k následujícím závěrům:

- (1) Jako kolokáty kategoriálních sloves se přednostně asociují LJ s vysokou MI nebo vysokou frekvencí.
- (2) Čím nižší je MI, tím větší je variabilita správných odpovědí.
- (3) Platnost tvrzení (2) je ovlivněna frekvencí predikativního jména: a) pokud je MI relativně nízká a zároveň frekvence predikativního jména je vysoká, pak přesnost odpovědi je vysoká; b) pokud je MI vysoká a zároveň frekvence predikativního jména je nízká, pak přesnost je nízká a chybovost vysoká.
- (4) Tendence (2) a (3) u konkrétních kolokací vstupují do vztahů s jinými kolokacemi, jejich frekvencemi a MI ve vztahu ke kategoriálnímu slovesu či predikativnímu jménu.
- (5) Nejvyšší míra chybovosti u predikativních jmen je v případě substantiv s nízkou frekvencí.
- (6) Nejvyšší míra chybovosti u kategoriálních sloves je v kolokacích s nízkou frekvencí.
- (7) V případě kolokací s vysokou MI kolokátu a zároveň velmi nízkou frekvencí kolokace i predikativního jména je míra správnosti odpovědí výrazně vyšší u kategoriálního slovesa než u predikativního jména.

Lze shrnout, že provedený výzkum potvrdil výchozí hypotézy, tj. že kolokace a lexikální jednotky s vysokou frekvencí jsou výraznou součástí jazykové kompetence nerodilých mluvčích a jejich jazyková znalost odráží i vzájemné poměry LJ v dané konstrukci i vzhledem ke konstrukcím jiným.

f/MI	10+	9+	8+	7+	6+	5+	4+
1000+							
900-999							
800-899							
700-799						PRÁCI	
600-699							
500-599							
400-499							
300-399				CHYBY, potíže		problémy	
200-299				poznámky		dojem	
100-199			ostudu		chybu, závěry, maximum	rozhovor	ROZHODNUTÍ, kroky
90-99			pokroky, kompromisy	reklamu		politiku	
80-89					hluk, vědu	úkoly	
70-79						karléru	
60-69	drahoty			zázraky, hlouposti		pokusy, byznys, scény	divadlo
50-59	naschvály		rámus			sport, pohyby	rozdíly, zkoušky
40-49			kravál	divy	nepořádek	návrhy, rozhovory, pořádek, průzkum, testy	VÝZKUM, plány
30-39		fóry		bibosti, ústupky, dobrotu, bordel	vtipy, přestávky, muziku	gesta, opak	hudbu, kampaň, test, filmy, fotky
20-29	šoufky	parádu, posunky, neplechu, kotrmelce		POVYK, osvětu, randát	doktorát, klíky, interview	sporty, marketing, rozruch, tajnosti, experimenty	drama, zásoby

Obr. 1. Kolokace kategoriálního slovesa *dělat* s ohledem na jejich frekvenci a MI predikativního jména

## Bibliografie

BNC (British national corpus). Collocates (cf. Sketch Engine). PDF guide. Dostupné z: <english-corpora.org> [cit. 23. 3. 2024]

GOLDBERG, Adele E. *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago Press, 1995.

GOLDBERG, Adele E. *Constructions at work: the nature of generalization in language*. New York: Oxford University Press, 2006.

GOLDBERG, Adele E. *Explain me this: creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press, 2019.

HERBST, Thomas. Constructions, generalizations, and the unpredictability of language: Moving towards colloconstruction grammar. *Constructions and frames*. 2020, roč. 12, č. 1, s. 56–95.

KETTNEROVÁ, Václava. Syntaktická struktura komplexních predikátů v češtině. *Slovo a slovesnost*. 2017, roč. 78, č. 1, s. 3–24.

KOLÁŘOVÁ, Veronika. *Valence deverbativních substantiv v češtině (na materiálu substantiv s dativní valenci)*. Praha: Karolinum, 2010.

RADIMSKÝ, Jan. *Verbo-nominální predikát s kategoriálním slovesem*. České Budějovice: Editio Universitatis Bohemiae Meridionalis, 2010.

**L2 learners are more likely to imitate a native than a non-native speaker  
displaying native-like phonetic patterns**

Daniel Kopecký, Václav Jonáš Podlipský, Šárka Šimáčková (Palacký University Olomouc)

Speakers spontaneously and subconsciously shift their pronunciation in the direction of surrounding speech. These adjustments occur in the long term [4, 9] but also in the short term [2], and they are modulated by social factors, such as speaker attractiveness [2] or role in conversation [8]. Although it is clear that phonetic imitation plays a role in L2 speech learning, the interplay between language distance [6] and social factors in L2 phonetic imitation is not yet well understood.

We focused on the varying vowel duration (henceforth VVD) before a voiced or voiceless coda in English. In minimal pairs such as *bat* and *bad*, speakers produce a longer vowel in the latter word due to the following underlyingly voiced consonant. Although the effect is salient in English, it seems to be marginal in many other languages, including Polish [5] or Czech [10]. A study with Polish learners of English [11] measured how VVD was imitated upon exposure to a native speaker displaying VVD and a non-native speaker neutralising VVD. By keeping the native status congruent with the presence vs absence of VVD, the study did not separate VVD from the social factor of native-speaker status.

In the present study, we remove this confound. Twenty-four advanced Czech learners of English first heard a short English text (without coda obstruents) recorded either by an American or by a Czech speaker intended to enhance their perception of the model speaker's nativeness status. Then they completed a word shadowing experiment, in which they produced 10 English CVC minimal pairs differing in word-final voicing and 28 fillers upon hearing them by the model speaker. The word stimuli, recorded by the same speaker, were manipulated (in Praat [3]) to display either native-like or completely neutralised VVD. We asked (1) whether the learners' L2 proficiency would predict VVD presence in English and (2) whether the imitation of VVD depended on the model speaker's (non-)native status. To factor out possible local speech tempo influence, the V/C ratio between log-transformed vowel durations and log-transformed coda constriction durations was measured instead of raw vowel durations.

First, we modelled the participants' pretest VVD (i.e., their values before exposure to stimuli), using the lexical decision task LexTALE [7] as a proxy of L2 English proficiency. The LexTALE scores reliably predicted the participants' pretest V/C ratios: highly proficient learners employed VVD, while less proficient learners less so. Next, the differences between the pretest and shadowing V/C ratios were modelled [V/C ratio differences ~ VVD presence \* underlying coda voicing \* speaker status \* LexTALE + (1 + underlying coda voicing | participant) + (1 | word)]. Overall, speakers shadowing the native model speech with VVD adjusted their V/C ratio productions to a significantly higher degree than those shadowing the non-native model with the same VVD values (see Fig. 1), showing an effect of speaker status. Proficient speakers shifted their pronunciations minimally, probably due to them already being close to the model's speech (cf. [1]). The imitation of either model speaker without VVD was minimal.

We conclude that for implicit imitation to occur, sufficient initial phonetic difference between the model and the imitator was needed, and the model's nativeness status and the target-language VVD patterns had to be aligned. Our results confirm that phonetic imitation is influenced by a complex interplay of both linguistic and social-psychological factors.

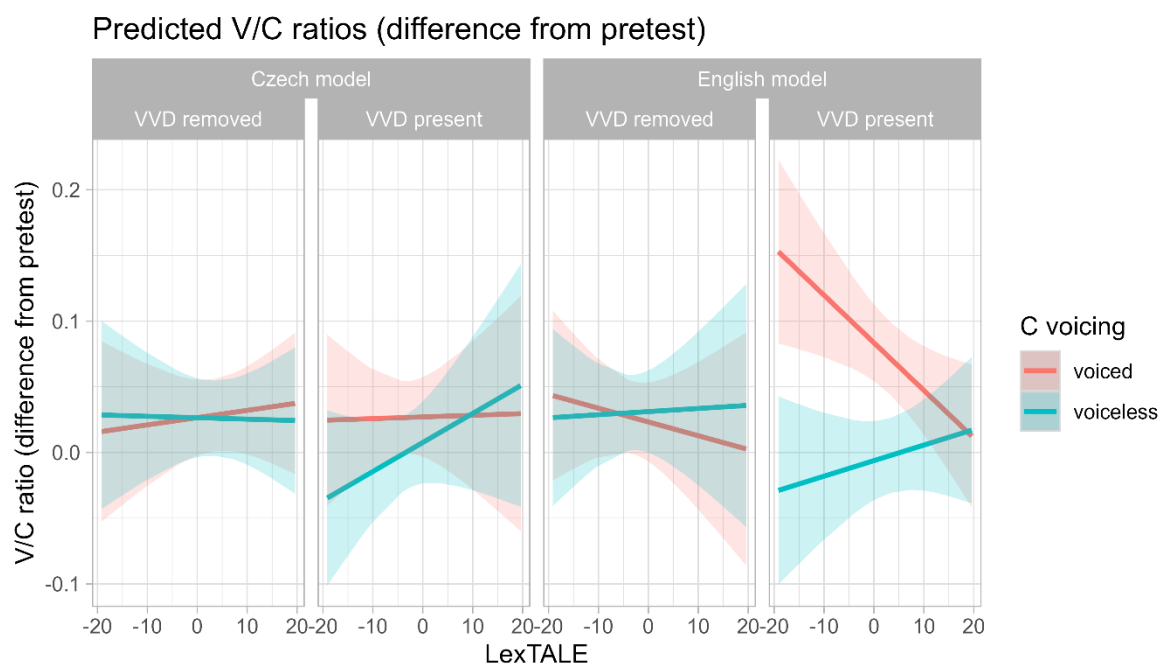


Figure 1: Plot of predicted values of the V/C ratio difference between pretest and shadowing. Split by accent of the model speaker, presence or absence of VVD, and coda voicing.

## Bibliography

- [1] Babel, Molly. 2012. "Evidence for phonetic and social selectivity in spontaneous phonetic imitation." *Journal of Phonetics*, 40(1):177–189. doi:10.1016/j.wocn.2011.09.001.
- [2] Babel, Molly, Grant McGuire, Sophia Walters, and Alice Nicholls. 2014. "Novelty and social preference in phonetic accommodation." *Laboratory Phonology*, 5(1):123–150. doi:10.1515/lp-2014-0006.
- [3] Boersma, Paul, and David Weenink. 2022. Praat: doing phonetics by computer [Computer program]. Version 6.3.03, available at <http://www.praat.org/>.
- [4] Harrington, Jonathan, Sallyanne Palethorpe, and Catherine I. Watson. 2000. "Does the Queen speak the Queen's English?" *Nature*, 408:927–928.
- [5] Jassem, Wiktor, and Lutosława Richter. 1989. "Neutralization of voicing in Polish obstruents." *Journal of Phonetics*, 17(4):317–325. doi:10.1016/S0095-4470(19)30447-4.
- [6] Kim, Midam, William S. Horton, and Ann R. Bradlow. 2011. "Phonetic convergence in spontaneous conversations as a function of interlocutor language distance." *Laboratory Phonology*, 2(1):125–156. doi:10.1515/labphon.2011.004.
- [7] Lemhöfer, Kristin, and Mirjam Broersma. 2012. "Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English." *Behavior Research Methods*, 44:325–343. doi:10.3758/s13428-011-0146-0.
- [8] Pardo, Jennifer S. 2006. "On phonetic convergence during conversational interaction." *The Journal of the Acoustical Society of America*, 119:2382–2393. doi:10.1121/1.2178720.
- [9] Pardo, Jennifer S., Rachel Gibbons, Alexandra Suppes, and Robert M. Krauss. 2012. "Phonetic convergence in college roommates." *Journal of Phonetics*, 40:190–197. doi:10.1016/j.wocn.2011.10.001.
- [10] Podlipský, Václav Jonáš, and Kateřina Chládková. 2007. "Vowel duration variation induced by coda consonant voicing and perceptual short/long vowel categorization in Czech." In *Speech Processing*, edited by Robert Vích, 68–74. Prague: Institute of Photonics and Electronics AS CR.
- [11] Zajac, Magdalena, and Arkadiusz Rojczyk. 2014. "Imitation of English vowel duration upon exposure to native and non-native speech." *Poznan Studies in Contemporary Linguistics*, 50(4):495–514. doi:10.1515/psicl-2014-0025.

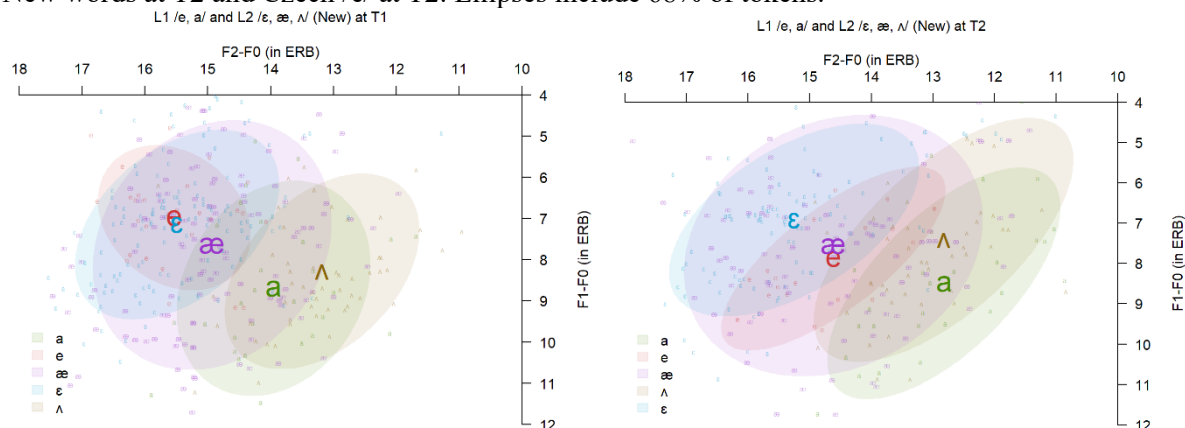
## Preschoolers' production of L2 vowels is affected by input quality: A longitudinal study

Monika Kučerová (Institute of Psychology, The Czech Academy of Sciences),

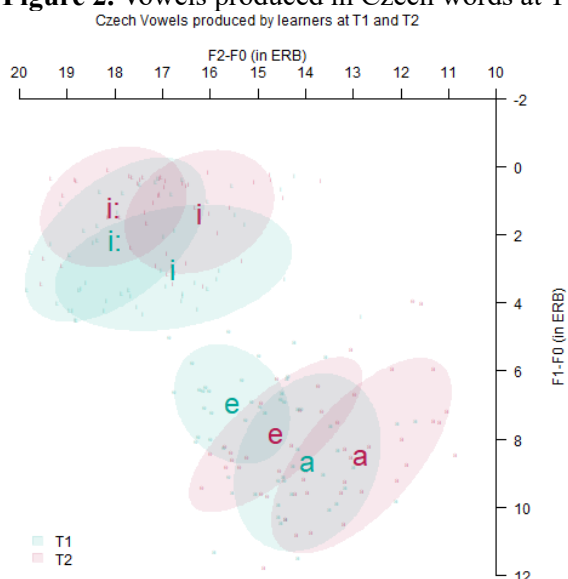
Šárka Šimáčková (Department of English and American Studies, Palacký University Olomouc)

We are interested in how three-to-five-year-old classroom learners of English use limited amount of L2 input to learn to produce L2 vowels, and how this learning process affects their L1 vowel production. When novice learners first encounter L2 speech, they process it relying on their L1 phonology. Learning of sound forms involves not only creating representations of novel words from input, but also updating the existing representations (Llompart, 2021). The learner's ability to notice and store the phonetic detail of sound is contingent upon the quantity and quality of L2 input (Flege, 2008). We assume that speakers store words as abstract strings of phonemes, as well as multiple phonetically-detailed tokens (exemplars), aggregated into exemplar clouds and weighted according to their frequency and recency (Nijveld et al., 2022). The categories are continuously modulated by language experience as new exemplars are stored. Accumulation of L2 tokens in an exemplar cloud previously consisting of primarily L1 tokens may increase variability and shift the centre of the cloud, causing phonetic drift of the entire category. We explore the effect of the teacher's pronunciation on the spectral quality of vowels in newly-learned words and on updating the pronunciation of earlier-learned words. Over 16 weeks, we recorded seven Moravian Czech EFL learners' English word production. In the course of data collecting, they experienced an input flood targeting the /æ/-/ɛ/ and /i/-/ɪ/ contrasts from the teacher, who provided quality contrasts across multiple tokens of the vowels in several newly-introduced words. At home they practised earlier-learned words with parents who potentially merged the /ɛ/-/æ/ contrast, difficult for Czech EFL speakers (e.g. Skarnitzl and Rumlová 2019), as well as the /i/-/ɪ/ contrast absent from the learners' L1 dialect (Šimáčková et al. 2012). Further, we test whether the non-immersive classroom input affected the quality of L1 vowels phonetically close to the targeted L2 sounds, and if so, whether such change was assimilatory or dissimilatory. The participants produced words containing Moravian Czech /i:, i, e, a/ in two sessions approximately nine weeks apart. Using linear mixed effects models, we modelled vowel height (F1-F0 in ERB) and retraction (F2-F0 in ERB) in English and Czech words for both Input sources (teacher, parents) and during the first (T1) and last (T2) half of the experiment. Recent input that targeted selected vowels was found to affect the learners' L2 output. However, the effect was lexically constrained. In words newly-learned from the teacher, /ɛ/ and /æ/ were separated in quality, due to the raising of /ɛ/. In familiar, earlier-learned words influenced by accented parental input, evidence of the spectral differentiation between /ɛ/ and /æ/ was not sufficient. In addition, some /æ/ tokens were [ɛ]-like, indicating lexical misrepresentation. We did not find evidence that the children's pronunciation of words with /i, ɪ, æ, ʌ/ from in-class and home input differed. We found phonetic drift for /e/, which retracted at T2 compared to T1, started overlapping more with L2 /æ/ and dissimilated from L2 /ɛ/. The L1/e/-L2/ɛ/ dissimilation was clearly achieved only in the newly-learned words. L1 /a/ retracted at T2 with its relative distance from /e/ remaining stable. L1 /a/ dissimilated from L2 /æ/ and assimilated to /ʌ/. The L1 vowels /i/ and /i:/ raised to similar height as L2 /i/, suggesting assimilatory drift. Our findings illustrate the influence of input characteristics on child learners' production of L2 vowels; the observed phonetic drift suggests that even time-constrained FL classroom exposure may result in changes of the learners' vowels.

**Figure 1.** Left: English /ε, æ, ʌ/ in New words at T1 and Czech /e/ at T1. Right: English /ε, æ, ʌ/ in New words at T2 and Czech /e/ at T2. Ellipses include 68% of tokens.



**Figure 2.** Vowels produced in Czech words at T1 (blue) and T2 (red). Ellipses include 68% of tokens.



## Bibliography

- Flege, J. (2008). Give input a chance!. In T. Piske & M. Young-Scholten (Ed.), *Input Matters in SLA* (pp. 175-190). Bristol, Blue Ridge Summit: Multilingual Matters.  
doi:10.21832/9781847691118-012
- Llompart, M. (2021). Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second-language phonological contrasts into the lexicon. *Bilingualism: Language and cognition*, 24(3), 481-496. doi:10.1017/S1366728920000656
- Nijveld, A., Ten Bosch, L., & Ernestus, M. (2022). The use of exemplars differs between native and non-native listening. *Bilingualism: Language and Cognition*, 25(5), 841-855.  
doi:10.1017/S1366728922000116
- Skarnitzl, R., & Rumlová, J. (2019). Phonetic aspects of strongly-accented Czech speakers of English. *AUC PHILOLOGICA*, 2019(2), 109-128. doi: 10.14712/24646830.2019.21
- Šimáčková, Š., Podlipský, V. J., & Chládková, K. (2012). Czech spoken in Bohemia and Moravia. *Journal of the International Phonetic Association*, 42(2), 225-232.  
doi:10.1017/S0025100312000102

**Statistical learning from visual speech cues in a novel artificial language**

Kateřina Kynčlová, Kateřina Chládková (Charles University, Czech Academy of Sciences)

Previous literature on implicit learning provides evidentiary support that statistical learning (SL), a phenomenon characterised by the unintentional adaptation to regularities in the environment, is exploited by both infants and adults during word boundary detection, a vital step in deciphering continuous, novel-language speech (see Christiansen, 2019; and Frost et al., 2019, for review). Studies investigating SL in speech have been limited to the auditory domain, i.e. the primary modality of spoken communication, which provides the learner with anchors aiding the word segmentation and possibly even facilitating the process of SL. Recent studies brought to attention the flaws of the typical exposure-posttest design which might fail to record the entire learning process (Dal Ben et al., 2021). Monitoring the process online with neuroimaging methods, such as EEG, could possibly eliminate this issue (Batterink and Paller, 2017).

To test the robustness of SL in speech without the bootstrapping by auditory cues and the limitations of the exposure-posttest design, the present study examined online neural indices of SL in speech devoid of all auditory and contextual cues. The aim was to investigate whether SL is a key process during novel language learning regardless of its modality or whether evidence of SL depends on the presence of auditory anchors in speech. Adult participants were exposed to naturally mouthed silent syllable streams in an artificial language in two within-subject conditions (adapting Batterink and Paller's 2017 auditory design). In one condition, twelve syllables occurred in a pseudo-random order with no underlying structure. In the other condition, the same twelve syllables were structured into four repeating trisyllabic statistical words. The structured stream provided no cues to the trisyllabic word boundaries apart from statistical cues.

In the EEG signal, phase synchronisation of neural oscillations and the speech stimulus was assessed as inter-trial phase coherence (ITPC) at the syllabic (3.3 Hz) and word (1.1 Hz) rate occurring in the exposure stream. Sensitivity to the underlying word structures (word learning index, or WLI) was quantified as the ITPC at the word frequency divided by the ITPC at the syllable frequency, and was also additionally tested by a post-exposure forced-choice task.

ITPC at word frequency increased significantly in the second block of the structured stream (see Fig. 1). WLI was greater in the structured compared to the random condition, indicating sensitivity to the underlying trisyllabic statistical words. Crucially, WLI increased significantly over the course of the first and second block in the structured condition, suggesting that the sensitivity to the underlying structures increases with exposure as a result of the gradual process of SL (see Fig. 2). This suggests that the statistics underlying the visual structured syllable streams have been detected and gradually acquired and that this process was successfully captured by the online assessment. A *t*-test on the behavioural word recognition accuracy did not detect a difference from the chance level. However, a positive medium-sized correlation was found between word recognition and WLI.

The present neural tracking results indicate that SL occurs during exposure to visual speech of a novel artificial language, despite there being no effect of SL in the behavioural posttest. This suggests that the brain's ability to synchronise to speech and detect underlying statistical patterns is rapid even in the absence of auditory and contextual cues, but might not be detected by behavioural posttests perhaps due to cognitive limitations. Since familiarity with the input immensely affects the process of learning, it is possible that the successful storing of detected statistical items would be affected by one's native modality. A follow-up study investigating this effect in hearing-impaired adults is currently underway. Our preliminary results show better performance in the behavioural posttest compared to the hearing adults, suggesting that familiarity with the input's modality might affect the success rate or might decrease the time needed for a successful storing of the acquired statistical items.

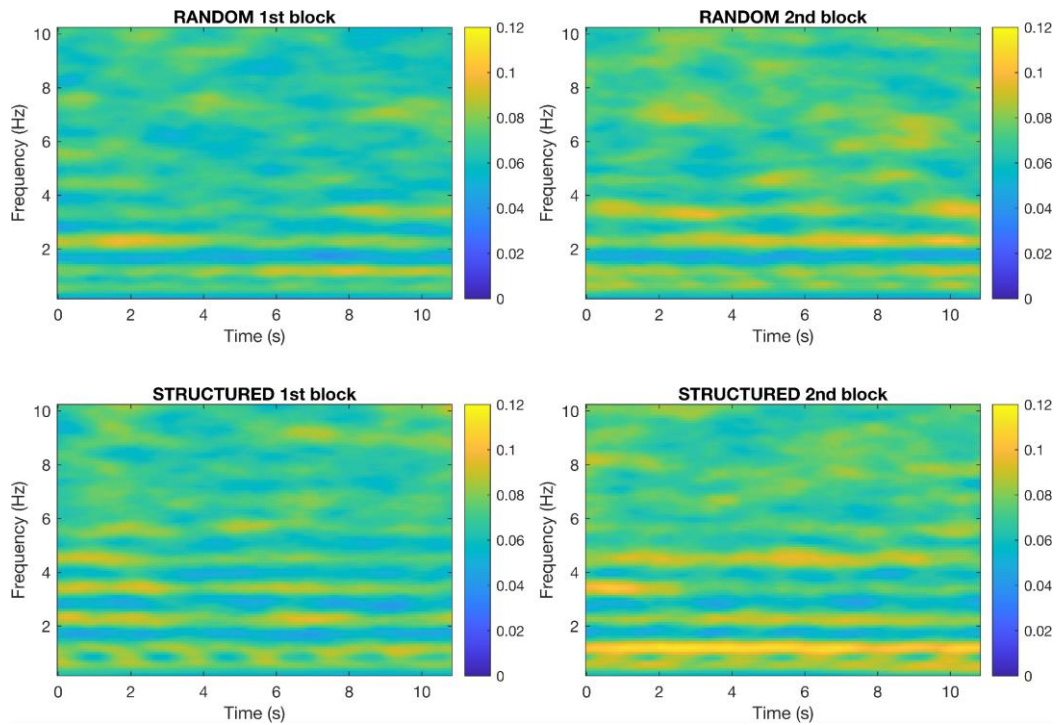


Figure 1: ITPC values in the first and the second exposure block of each condition, averaged across the 7 analysed channels (Pz, O1, O2, T3, T4, T5, T6).

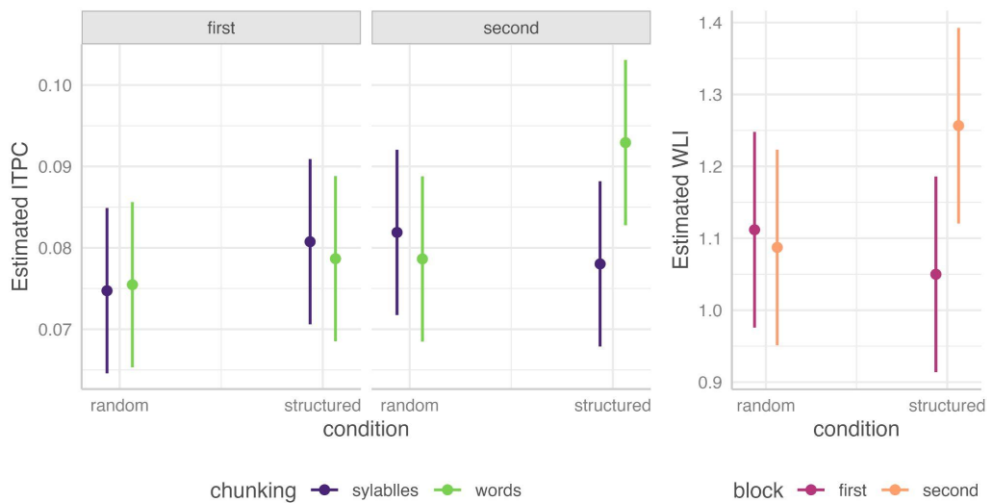


Figure 2 Left: Estimated ITPC per condition, exposure block, and chunking rate. Right: Estimated WLI per condition and exposure block (means and 95% confidence intervals)

### Bibliography

- Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, 90, 31–45.
- Christiansen, M. H. (2019). Implicit Statistical Learning: a tale of two literatures. *Topics in Cognitive Science*, 11(3), 468–481.
- Dal Ben, R., De Hollanda Souza, D., & Hay, J. F. (2021). When statistics collide: The use of transitional and phonotactic probability cues to word boundaries. *Memory & Cognition*, 49(7), 1300–1310.
- Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin*, 145(12).



**Cross-language phonological neighborhood effects**

Eva Maria Luef (University of Hamburg, Charles University)

Typologically related languages are often characterized by sets of words that are phonologically similar to a given target word in one language. For instance, the English word “cough” shares the majority of its phonemes (except for one) with its German neighbor “Koffer”/ Engl. *suitcase* (see Figure 1). The significance of phonological neighborhoods for speech processing is well documented in the psycholinguistic literature (e.g., Dijkstra and Van Heuven, 2002; Van Wijnendaele and Brysbaert, 2002, Vitevitch, 2012). When exposed to stimuli in one language, multilingual speakers also activate similar word forms in their other language(s) (Dijkstra, Grainger, & Van Heuven, 1999; Shook & Marian, 2013; Van Wijnendaele & Brysbaert, 2002). One frequently studied manifestation of cross-lingual activation is the 'cognate processing advantage' (de Groot & Nas, 1991; Schulpen, Dijkstra, Schriefers, & Hasper, 2003). This effect describes the observation that cognate words, which share both form and meaning across different languages (e.g., English "guitar" [gi'tar] and German "Gitarre" [gi'tarə]), are processed more rapidly and accurately in bilingual speakers of the languages. This facilitated processing of cognates suggests that multilingual speakers simultaneously activate word representations in all known languages, even when they are contextually focused on one specific language (Dijkstra & Van Heuven, 2002).

Co-activation of neighboring words (differing by one phonological segment) represents a core feature of speech processing models: a large number of phonological neighbors generally slows down lexical recognition (e.g., Neighborhood Activation Model, see Vitevitch and Luce, 2016). Thus, words with greater phonological overlap with numerous others across various known languages cause delays in lexical recognition. The competition effect arising from multiple neighbors must be resolved before successfully recognizing a word (Norris, McQueen, & Cutler, 1995). While phonological neighborhood effects are well-established in monolingual studies (e.g., Chan & Vitevitch, 2009; Chen & Mirman, 2012; Vitevitch & Luce, 2016), there is some evidence that cross-linguistic phonological neighborhoods appear to exert a similar influence on speech processing (Marian, Bartolotti, Chabal, & Shook, 2012; Marian et al., 2008; Shook & Marian, 2013).

The majority of research on bilingual activation spreading has focussed on visual word recognition and early bilingualism (i.e., learning two first languages). The present study was designed to investigate the effect of cross-language phonological neighborhoods in late bilingualism of English as a second language in German first-language users. In auditory lexical decision tasks, participants were presented with L2-English stimuli that either had many or no German phonological neighbors. Results showed that the presence of German neighbors delayed English speech recognition, a finding in line with traditional speech recognition models. In addition, lexical frequency of German words had an effect on how fast the English words were recognized. Understanding cross-language phonological neighborhood effects can provide insights into the nature of the late bilingual lexicon and the degree of cross-language integration that governs speech processes.

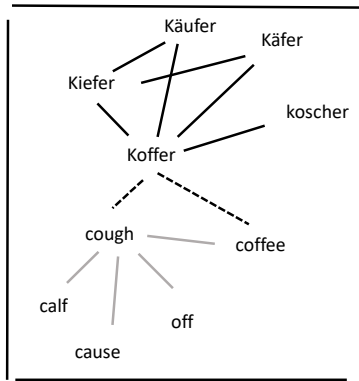


Figure 5: Phonological neighborhoods across the two languages German and English.

## Bibliography

- Chan, K. Y., & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive Science*, 34, 685-697.
- Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, 119(2), 417-430.
- de Groot, A. M., & Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory & Language*, 30, 90-123.
- Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory & Language*, 41(4), 496-518.
- Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175-197.
- Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy access resource for phonological and orthographic neighborhood densities. *PLoS ONE*, 7(8), e43230.
- Marian, V., Blumenfeld, H. K., & Boukrina, O. V. (2008). Sensitivity to phonological similarity within and across languages. *Journal of Psycholinguistic Research*, 37(3), 141-170.
- Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1209-1228.
- Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 616-627.
- Schulpen, B., Dijkstra, T., Schriefers, H., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human perception and performance*, 29(6), 1155-1178.
- Shook, A., & Marian, V. (2013). The Bilingual Language Interaction Network for Comprehension of Speech. *Bilingualism: Language and Cognition*, 16(2). doi:doi: 10.1017/S136672891200046
- Vitevitch, M. S. (2012). What do foreign neighbors say about the mental lexicon? *Bilingualism: Language and Cognition*, 15(1), 167-172.
- Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics*, 2, 75-94.

**Gender processing in a non-gendered L2:  
The role of L1 grammar, gender stereotypicality and L2 proficiency**

Mikuláš Preininger<sup>1</sup>, James Brand<sup>2</sup>, Adam Kříž<sup>1</sup>, Markéta Ceháková<sup>1</sup>

(<sup>1</sup>Charles University, <sup>2</sup>Edge Hill University)

**Introduction** When processing gender information, speakers rely on both linguistic and non-linguistic cues. Gygax et al. (2008) demonstrated that when interpreting the gender of a role-noun, native L1 speakers of a language without grammatical gender (English) rely on gender stereotypicality to guide their processing, whereas speakers of grammatically gendered languages, where almost all role-nouns carry gender marking, e.g. German, do not rely on stereotypicality and interpret the gender of the referent solely based on the grammatical gender of the noun. This was the case even though generically intended masculine forms were used. Sato et al. (2013) expanded on this work by focusing on L2 processing of English by speakers with a grammatically gendered L1, showing that the L2 speakers are also guided primarily by gender stereotype information. Interestingly, the more proficient speakers did so to a higher degree, whereas the less proficient speakers still carried a male bias from their L1. Preininger et al. (in prep) have, however, demonstrated that the utilization of stereotypicality cues is already present when speakers process sentences in their grammatically gendered L1. In this paper, we conduct a partial replication of Sato et al. (2013) with L1 Czech speakers in their L2 English, to investigate the influences of grammatical gender, stereotypicality, and L2 proficiency. Additionally, we address issues related to coreference in the original stimuli and their overall naturalness, which might have confounded the results.

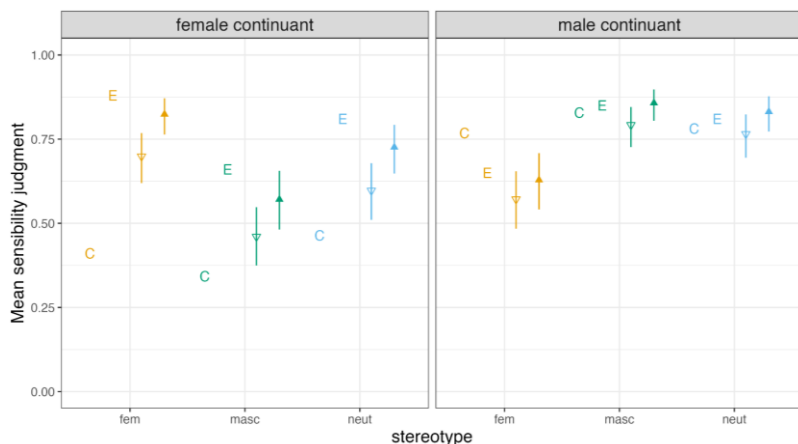
**Methods** In Exp. 1, L1Czech-L2English speakers from two proficiency levels ( $n_{low} = 160$ ,  $n_{high} = 148$ ) read pairs of English sentences and their task was to assess whether the second sentence was a sensible continuation of the first one, in a binary choice task. The first sentence always included a role-noun in a plural form (e.g. *journalists*), and the second sentence included a coreferring gendered noun (e.g. *some of the men*). Gender stereotypicality of the role-noun was manipulated (feminine/masculine/neutral), as well as gender of the anaphoric noun (feminine/masculine). We assumed that lower proportion of positive responses and higher response times in positive responses indicate anaphora resolution issues. In Exp. 2, L1Czech-L2English speakers ( $n_{low} = 181$ ,  $n_{high} = 127$ ) completed the same task, only this time with a new set of stimuli pretested for coreference strength and with a 5-point Likert scale (1 = not at all sensible, 5 = highly sensible).

**Results** In Exp. 1, we observed a significant two-way interaction between continuation and stereotype and a main effect of proficiency in glmer models predicting sensibility. The stereotype congruent continuations were more sensible and there was generally higher sensibility for responses by more proficient speakers (see Fig. 1). For response times, we observed the same significant effects and interactions with the positive responses being made quicker in stereotype congruent continuations and by more proficient speakers (Fig. 2). In Exp. 2, we only observed a significant effect of proficiency, with fewer sensible responses in less proficient speakers (Fig. 3).

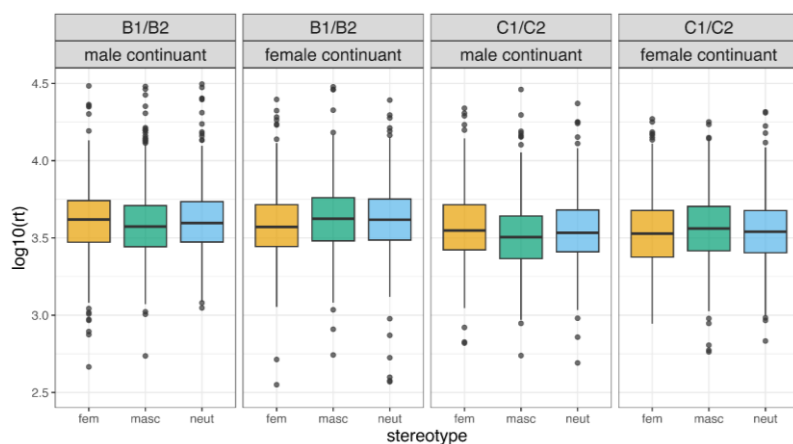
**Discussion** The results show that when processing anaphora in English, native speakers of Czech rely on the gender stereotypicality of the antecedent and do not carry a male bias from their L1, where masculine generic forms are often used. Findings from Exp. 1 are in contrast with Sato et al. (2013) since we have not observed any gender incongruent continuation to be preferred over the congruent one in neither of the proficiency groups. The only difference found between the groups was the overall greater ease of processing in more proficient speakers. Exp. 2 further shows that if the coherence of the

two sentences increases and the way sensibility is measured changes, even the stereotypicality effect disappears.

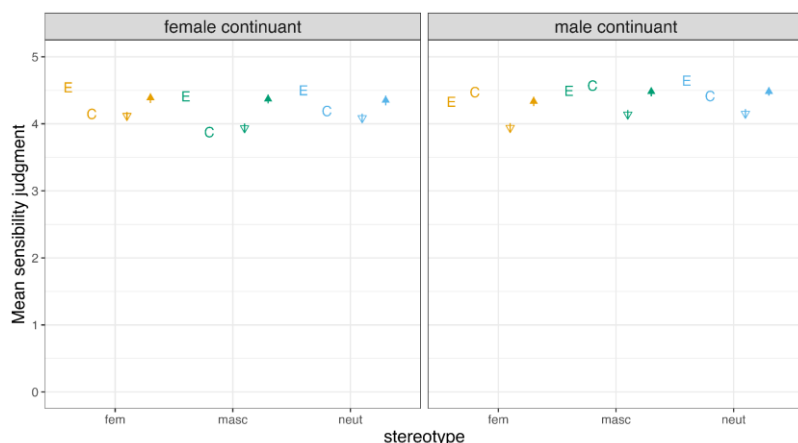
## Figures



**Figure 1.** Mean sensibility judgments in Exp. 1. Triangles represent condition means in L2 data, the empty one facing downwards for B1/B2 speakers and the filled one facing upwards for C1/C2 speakers. Letters also represent condition means, C for L1 Czech data from Preininger et al. (in prep) and E for L1 English data from Gygax et al. (2008). The plot is faceted by the gender in the continuation. The colors represent gender stereotypicality of the role-noun (feminine in orange, masculine in green, neutral in blue).



**Figure 2.** Response times in exp. 1. Each box represents logged response times in milliseconds for each condition and proficiency group. The plot is faceted by proficiency and the gender in the continuation. The colors represent gender stereotypicality of the role-noun (feminine in orange, masculine in green, neutral in blue).



**Figure 3.** Mean sensibility judgments in exp. 2. Triangles represent condition means in L2 data, the empty one facing downwards for B1/B2 speakers and the filled one facing upwards for C1/C2 speakers. Letters also represent condition means, C for L1 Czech data and E for L1 English data, both from Preininger et al. (in prep). The plot is faceted by the gender in the continuation. The colors represent gender stereotypicality of the role-noun (feminine in orange, masculine in green, neutral in blue).

## Bibliography

Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and cognitive processes*, 23(3), 464-485.

Preininger, M., Brand, J., Kříž, A., Ceháková, M. & Chromý, J. (in prep). Genericity on a scale: Offline processing of masculine generic role-nouns in Czech.

Sato, S., Gygax, P. M., & Gabriel, U. (2013). Gender inferences: Grammatical features and their impact on the representation of gender in bilinguals. *Bilingualism: Language and Cognition*, 16(4), 792-807.

## Korpus primární slovní zásoby češtiny jako dalšího jazyka z diachronního pohledu

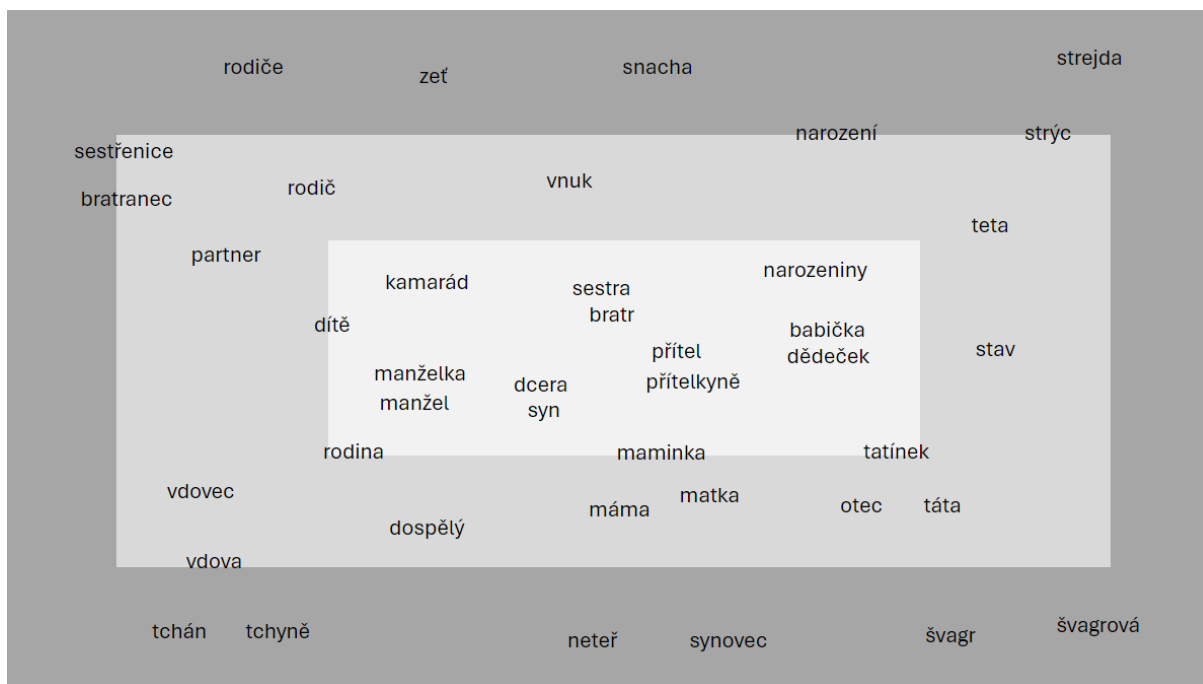
Hana Svobodová, Kateřina Malá, Simona Koryčánková (Masarykova univerzita)

Příspěvek je zaměřen na představení korpusu primární slovní zásoby vytvořeného na základě vybraných učebnic českého jazyka jako dalšího jazyka na úrovni A1 a věnovaného tématům podporujícím integraci jedinců s odlišným mateřským jazykem (OMJ) do naší společnosti. Pro výběr učebnic byla zvolena následující kritéria: učebnice jsou určeny pro výuku českého jazyka jako jazyka cizího nebo druhého na úrovni A1; učebnice se orientují na výuku dospělého mluvčího; učebnice obsahují témata, která jsou později vyžadována u zkoušky k udělení občanství, ta je na úrovni B1; učebnice vyšly po roce 2000. Aktuálně jsou analyzovány učebnice *Krok za krokem* (Holá 2016), *Čeština Expres* (Holá, Bořilová 2016), *Czech it up* (Švarcová, Wenzel 2020), *Česky prosím* (Cvejnová 2019) a *Cestou necestou* (Svobodová, Muchová 2023), poté bude korpus obohacen o další učebnice odpovídající zvoleným kritériím. Excerptovaná slovní zásoba je v korpusu rozdělena do tematických oblastí, přičemž u každého slova je uvedeno, v jaké učebnici se vyskytuje. Vzniká tak přehledná tabulka, na jejímž základě je možné například charakterizovat jádro slovní zásoby pro úroveň A1, tedy slova považovaná všemi učebnicemi za důležitá, a periferii slovní zásoby, tj. slova vyskytující se pouze v jedné učebnici, jež pro danou úroveň nemusí být tak významná.

Jednu z možností práce s tímto korpusem ilustrujeme tématem rodina. To je poměrně jednoznačně omezené rodinnými příslušníky a vztahy mezi nimi. Celkový počet slov, která jsou zařazena k tématu rodina, je v naší databázi aktuálně 56. Nejvýrazněji tuto slovní zásobu pokrývá učebnice *Krok za krokem* s počtem 49 slov, další učebnice k tomuto tématu uvádějí slov méně, konkrétně *Cestou necestou* 30 slov ze základního slovníčku (učebnice je doplněna i slovníčkem rozšiřujícím, který do toho výčtu nezahrnujeme), *Česky prosím* 26 slov, *Česky expres* 25 a *Czech it up* 23 slov. Vychází však otázka, zda všechna slova, která jsou zařazena do učebnice *Krok za krokem*, budou mluvčímu na úrovni A1 užitečná, neboť základní slovní zásoba, která je uvedena shodně ve všech učebnicích, čítá pouze 12 slov, tedy pouze 21 % z celkového počtu. V rámci příspěvku budou proto podobným způsobem analyzovány i další okruhy slovní zásoby, například bydlení nebo člověk a zdraví.

Analýza základní slovní zásoby je doplněna o komparaci s dalšími jazykovými korpusy. V rámci tématu rodina byly využity např. Syn2020 a Oral v1, především pro srovnání slov, která jsou v učebnicích prezentována nejednotně: *matka*, *máma*, *maminka* a *otec*, *táta*, *tatínek*. Korpusová data poukazují na významné rozdíly mezi psanou a mluvenou formou. U psané převládají základní slovníkové tvary *matka* a *otec*, avšak při mluvené se nejčastěji užívá podoba *máma* a *táta*. Zdrobněliny *maminka* a *tatínek* nejsou nejčastější užívanou variací, ale počet výskytů je nezanedbatelný. Ideálně by tak ve slovníčcích měly být prezentovány všechny tři varianty a vzdělavatel by měl při výuce vysvětlit rozdíly v užití těchto slov.

Těžištěm práce je jazyková analýza slovní zásoby z diachronního pohledu, přiblížen bude její vývoj od praslovanštiny přes nářeční štěpení slovanských jazyků po současnou češtinu. Rozlišeny budou jevy všech jazykových rovin (hláskoslovné, morfologické, lexikální, sémantické), které ovlivnily formování základní slovní zásoby představené v korpusu. Toho bude následně využito při srovnání tvarů slov v jednotlivých slovanských jazycích. Cílem je pomocí začlenění diachronního pohledu podpořit osvojování českého jazyka pro nerodilé mluvčí se slovanským mateřským jazykem a omezit tak jejich transfery z rodného jazyka.



### Srovnání výskytů v korpusu Syn2020 a Oral v1

<i>slovo</i>	<i>syn2020; i.p.m.</i>	<i>oral v1, i.p.m.</i>
matka	<b>282,44</b>	<b>48,73</b>
máma	<b>136,38</b>	<b>303,85</b>
maminka	<b>95,46</b>	<b>151,85</b>
otec	<b>295,38</b>	<b>48,89</b>
táta	<b>120,61</b>	<b>282,79</b>
tatínek	<b>41,96</b>	<b>65,39</b>

### Bibliografie

Cvejnová, J. (2019). Česky, prosím I. Karolinum.

Holá, L. (2016). Česky krok za krokem 1. Akropolis.

Holá, L., Bořilová, P. (2016). Čeština expres 1: A1/1. Akropolis.

Kopřivová, M., Lukeš, D., Komrsková, Z., Poukarová, P., Waclawičová, M., Benešová, L., Křen, M. (2017) ORAL: korpus neformální mluvené češtiny, verze 1 z 2. 6. 2017. Ústav Českého národního korpusu FF UK, Praha. Dostupný z: <http://www.korpus.cz>

Kosek, P. (2014) Historická mluvnice češtiny – překlenovací seminář. 1. vyd. Brno: Masarykova univerzita.

Kosek, P. (2014) Historická mluvnice češtiny I. 1. vyd. Brno: Masarykova univerzita, 2014.

Křen, M., Cvrček, V., Henyš, J., Hnátková, M., Jelínek, T., Koček, J., Kovářiková, D., Křivan, J., Lamprecht, A., Šlosar, D., Bauer, J. (1986) Historická mluvnice češtiny. 1. vyd. Praha: Státní pedagogické nakladatelství, 1986.

Milička, J., Petkevič, V., Procházka, P., Skoumalová, H., Šindlerová, J., Škrabal, M. (2020) SYN2020: reprezentativní korpus psané češtiny. Ústav Českého národního korpusu FF UK, Praha. Dostupný z: <http://www.korpus.cz>

Svobodová, H., Muchová, Z. (2023) Cestou necestou A1+ (učebnice češtiny jako druhého/cizího jazyka). MUNI Pres, Brno.

Švarcová, T., Wenzel, J. (2020). Czech it up! 1: čeština pro cizince: úroveň A1 = Czech for foreigners: level A1. Univerzita Palackého v Olomouci.

## Using ultrasound in L2 Czech pronunciation teaching: /r/ and /l/ sounds

Jitka Veroňková, Tanja Kocjančič (Charles University)

Because of a high reliance on perception, L2 speech sound remediation is rather difficult for many L2 learners. A method that bypasses at least some of the challenges of traditional L2 pronunciation practice is ultrasound feedback via real-time tongue imaging, which provides information about the shape, position, and movement of the tongue during speech (Stone, 2005). By visualizing the tongue, it allows L2 learners to improve understanding of their articulatory goals, increase awareness of one's tongue movements, and practice by utilizing a visual articulatory control and not only an auditory one. The method has been attested in the remediation of different L2 vowels and consonants and has overall resulted in improved pronunciation even after a single short individual practice session (Kocjančič Antolík et al., 2019). Among others, ultrasound feedback has been shown to improve /r-l/ contrast for Japanese L2 English speakers, either by improving both targets (Gick et al., 2008; Bryfonski, 2023) or only /l/ (Tateishi & Winters, 2013). Less is known about the applicability of the method to the improvement in /r-l/ contrast in L2 Czech, where /r/ is categorized as an alveolar trill. A study exploring Spanish trill acquisition in L1 English speakers concluded that /r/ production depends more on the motoric learning of the speaker than on the pre-training articulatory position of the target sound (del Puy Ciriza, 2020).

In the L2 Czech of Chinese speakers (2 years study Czech in China), the canonical realization was around 40% for /r/ and 70% for /l/. The most common incorrect pronunciation of /r/ was the exchange for /l/ (30%) and vocalization as the realization of /l/ (19%). The data showed intra- and interspeaker variability based on the position of the target sound within a word and phonetic combinatorics. (Veroňková, 2022) In another experiment, the success rate of /r/ and /l/ in Chinese speakers was even lower and the ratio of incorrectly perceived items reached 12% in both /r/ and /l/; perception, similar to production, caused difficulties in different patterns and varied between subjects. (Poukarová & Veroňková, 2020)

The current study aims to explore the usability of the ultrasound feedback in the increase of L2 Czech /r-l/ contrast and to attest the method for trill acquisition. The method was incorporated into ongoing L2 Czech pronunciation lessons attended by 12 Korean and Chinese-speaking learners of Czech. The experiment includes the following steps:

- a perceptual test for discrimination of the phonemes /r/ and /l/ containing 72 items of minimal pair, e.g. *viry* – *vily*
- audio and ultrasound recordings of read texts with a length of about 50 isolated words at three time-points: pre-, mid-, and post-training.
- four 15-minute pronunciation training with ultrasound feedback per student pair.

Four learners with different starting levels of acquisition and manner of r-sound production are participating in the study.

Due to the incorporation of the method in the regular bachelor course, data collection took place continuously during the summer semester. A complete set of results will be available at the time of the conference. However, it is clear from ongoing analyses that the experiment accomplished our expectation in improved production of /r-l/ contrast in students. Additionally, the outcomes contribute to the growing knowledge of applying ultrasound feedback in L2 pronunciation teaching.



## Bibliography

- Bryfonski, L. (2023). Is seeing believing? The role of ultrasound tongue imaging and oral corrective feedback in L2 pronunciation development. *Journal of Second Language Pronunciation*, *9*(1), 103–129.
- del Puy Ciriza, M., & Rivera-Campos, A. (2020). Teaching the Spanish trill to L1 English speakers using ultrasound instruction: a preliminary study on pronunciation pedagogy. *Journal of Spanish Language Teaching*, *10*(2), 1–14.
- Gick, B., Bernhardt, B., Bacsfalvi, P., Wilson, I., & Zampini, M. (2008). Ultrasound imaging applications in second language acquisition. *Phonology and second language acquisition*, *36*(6), 309–322.
- Kocjančič Antolík, T., Pillot-Loiseau, C., & Kamiyama, T. (2019). The effectiveness of realtime ultrasound visual feedback on tongue movements in L2 pronunciation training: Japanese learners' progress on the French vowel contrast /y/-/u/. *Journal of Second Language Pronunciation*, *5*(1), 72–97.
- Stone, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical linguistics & phonetics*, *19*(6–7), 455–501.
- Poukarová, P. & Veroňková, J. (2020). Degree of Czech R/L acquisition in L1 Chinese speakers. In *ExLing 2020: Proceedings of 11th International Conference of Experimental Linguistics*, 12–14 October 2020, Athens, Greece.
- Tateishi, M., & Winters, S. (2013). Does ultrasound training lead to improved perception of a non-native sound contrast? Evidence from Japanese learners of English. In *Proceedings of the 2013 annual conference of the Canadian Linguistic Association* (pp. 1–15). Canadian Linguistic Association.
- Veroňková, J. (2022). Realizace českých souhlásek r a l u čínských mluvčích. *Casalc Review* *12*(1), 5–18.

21

*Metafora semper viva*

*Metaphor semper viva*

organizuje / organizer

Irena Vaňková

(CZE, ENG)

## Popis panelu / Workshop description

**(CZE)** Teorie konceptuální metafory G. Lakoffa a M. Johnsona byla svého času zjevením a inspirací, přestože metaforu a metonymii ukázal v širokých souvislostech lingvisticko-sémiotických jako „dva aspekty jazyka“ mnohem dřív už Roman Jakobson. Šíře však lingvisty zaujala až v pojetí lakoffovsko-johnsonovském, jako součást kognitivní lingvistiky. Od vydání *Metafor, kterými žijeme* (1980) byla ovšem jejich teorie i metodologie průběžně prověřována, revidována, obohacována a postupně také v různých ohledech překračována. Díky ní začala být metafora (resp. šíře vzato figurativnost vůbec) jako princip lidského myšlení a konání studována v dříve nezohledňovaných souvislostech, a zdaleka nejen v rámci jazykovědy.

Proponovaný panel by měl metaforu ukázat především v širším kontextu dnešní kulturně a antropologicky zaměřené lingvistiky a etnolingvistiky, opřené o kulturní sémiotiku. Zároveň však chceme v panelové debatě položit i obecnější otázky transdisciplinární, např. po možnostech studia metafory v rámci evoluční biologie, zoosémiotiky či etologie (srov. např. biosémiotickou metaforu). Vycházíme totiž z premisy, že jako lidé se máme vážně zabývat svým místem mezi ostatními živočišnými druhy, a vůbec svým zakotvením do prostředí, v němž se pohybujeme (řečeno v intencích ekolingvistiky). K tomu patří i metareflexe diskurzu, jímž – značnou měrou metaforicky – člověk a lidské společenství žije. To nás přivádí k pojmu *natureculture* (dnes už bez rigidního rozlišování mezi přírodou a kulturou) a také k pojmu *languaculture*, reflektujícímu fakt, že lidské prostředí je určeno jazykem a kulturou. Metaforu (jejíž pojmenování zde volíme jako zastřešující zkratku za nedoslovné myšlení vůbec) je třeba vidět v těchto souvislostech. Porozumění principu metaforizace v širších a obecnějších souvislostech (a v jiném modu, než jaký nabízejí jazyky mluveně-psané) se pak otevírá třeba při studiu znakových jazyků – a „viditelných metafor“, jímž bychom také rádi věnovali pozornost – podle možností (a přihlášených příspěvků) případně i v souvislosti s multimodálními podobami metafory.

V rámci panelu se budeme zabývat i otázkami, jež v souvislosti s metaforou (resp. doslovností a nedoslovností) pokládá antropologicky zaměřená lingvistika a etnolingvistika, zejm. s ohledem na lidovou kulturu a folklorní texty – a vůbec texty (v širokém smyslu slova), které se opírají o mytické myšlení a (s ním související) magickou semiózu.

Výzkumy metafory jako kognitivní strategie bývají zaměřeny i na její rétorické uplatnění a využití metafor pro různé sociální účely, srov. např. metaforická paradigmata ve vědě (zejm. v lingvistice), metafory v diskurzu politickém, pedagogickém či medicínském. Nechceme zapomenout ani na specifický diskurz umělecký; zvaným přednášejícím bude David S. Danaher (z Wisconsinské univerzity v Madisonu) s příspěvkem o metaforických aspektech konceptualizace lítosti (jakožto – jak se traduje – specificky české emoce) v Kunderově *Knize smíchu a zapomnění*

V panelu hodláme postavit do otázky, jak je patrné, jednak některé teoretické problémy studia metafory (a figurativnosti vůbec) v kontextu dnešní kulturní lingvistiky s vykročením směrem k transdisciplinárním přesahům, jednak způsoby uplatnění metafor v určitých konkrétních oblastech (folklor, literatura a umění vůbec, věda, politika, pedagogika, medicína ap.). Rádi bychom tak přispěli k promýšlení metafory a metaforizace z různých hledisek, v různých kontextech a v aktuálním dialogu nejen v lingvistice, ale i v transdisciplinárních souvislostech.

**(ENG)** In its time, G. Lakoff's and M. Johnson's Conceptual Metaphor Theory (CMT) was both revelation and inspiration—and this despite the fact that both metaphor and metonymy had already been recognized much earlier by Roman Jakobson as „two aspects of language“ within the broad context of a linguistic-semiotic approach to the study of language. Metaphor, however, truly began to interest linguists only when Lakoff and Johnson developed CMT as a central component of present-day

Cognitive Linguistics. Since the publication of their seminal book *Metaphors We Live By* (1980), both the theory and methodology of CMT have been refined, revised, enriched, and also gradually extended and embellished in a number of respects. Thanks to CMT, metaphor (or, more generally speaking, figurativity) began to be researched as a basic principle of human thinking and behavior, and this was true in ways that had earlier not been taken into consideration and with regard not only to the study of language.

The proposed panel is intended to explore metaphor primarily within the wider context of a contemporary cultural-anthropological linguistics and an ethnolinguistics grounded in cultural semiotics. At the same time, however, the panel will also pose more general transdisciplinary questions related, for example, to the possibility of studying metaphor within evolutionary biology and zoosemiotics (cf. biosemiotic metaphor). We start from the premise that, as human beings, we should reflect seriously on our relationship both to other animal species and also to our lived environment (a topic explored in the subfield of ecolinguistics). For this, however, we must necessarily engage in meta-reflection on human discourse given that discourse is central to how we as individuals living in a society understand ourselves: in discourse of this kind, metaphor always plays an essential role. This consideration brings us to the concepts of both *natureculture*, in which a rigid disjuncture between nature and culture is denied, and also *linguaculture*, which reflects the fact that the relationship to our lived environment is mediated by both language and culture. It is through these interrelated lenses that we must understand metaphor, which is word that may serve as an umbrella term for all non-literal or figurative forms of cognition. One example of how we might apply the principle of metaphorization more broadly and with regard to these kinds of general associations (and also in a way that goes beyond a focus on spoken-written language) is the study of „visual metaphors“ in sign language, and this is also a theme that the panel would indeed like to address, depending on proposal submissions, particularly in relationship to so-called multimodal metaphoricity.

In the proposed panel, presenters will also engage with topics that have been subjects of research in anthropological linguistics and ethnolinguistics and that directly concern metaphor (literalness and non-literalness). Presenters will analyze texts (in the broad sense of the word) related especially to folk culture and folklore as well as, more generally speaking, texts that rely on forms of mythical cognition and the magical semiosis that characterizes them.

Research conducted on metaphor as a cognitive strategy has focused on rhetorical applications related to the use of metaphor in various sociocultural domains, and these include: metaphorical paradigms in academic study (especially in linguistics) and also the use of metaphor in political, pedagogical, and medical discourse. We also should not leave aside the question of the role played by metaphor in aesthetic discourse, and David S. Danaher (University of Wisconsin-Madison), the panel's invited speaker, will present on metaphorical aspects of the conceptualization of the emotion concept *lítost* (which is, at least in theory, an emotion specific to Czech linguaculture) in the context of Milan Kundera's *The Book of Laughter and Forgetting*.

As made clear above, the questions framing the panel's intent concern theoretical problems in the study of metaphor (and figurativity more generally) that directly relate to specific questions in contemporary cultural linguistics with an awareness that these concerns are also transdisciplinary. Presenters will focus on applying metaphor theory to certain concrete domains: folklore, literature and art in general, science, politics, pedagogy, medicine, etc. They will thereby contribute to thinking through the notions of metaphor and metaphorization from a multiplicity of interrelated perspectives—in various specific domains as well as in ongoing dialogue not only with the field of linguistics proper, but rather in a broader transdisciplinary sense.

## Literatura / References

- Bartmiński, Jerzy. 1988. „Niebo się wstydzi”. Wokół ludowego pojmowania ładu świata. In: *Kultura – literatura – folklor*. Ed. Marek Graszewicz, Jacek Kolbuszewski. Warszawa: UMCS, 96-106.
- Bernárdez, Enrique. 2008. *El lenguaje como cultura*. Madrid: Alianza.
- Blumenberg, Hans. 2010. *Paradigms for a Metaphorology*. Ithaca – New York: Cornell University Press.
- Bohát, Róbert. 2021. Metaphors to Survive by: Mimicry as Biometaphors, Embodiment of Sign and Cognitive Tools (not only) in Animals? *Linguistic Frontiers*, 4 (1), 31-43.
- Głaz, Adam, Danaher David S., Łozowski Przemysław (eds.). 2013. *The Linguistic Worldview. Ethnolinguistics, Cognition, and Culture*. London: Versita.
- Jakobson, Roman. 1995. Dva aspekty jazyka a dva typy afatických poruch (1956). In: *Poetická funkce*. Ed. Miroslav Červenka. Jinočany: Nakladatelství H&H
- Lakoff, George – Johnson, Mark. 1980. *Metaphors We Live By*. Chicago: The University of Chicago Press.
- Kövecses, Zoltán. 2005. *Metaphor in Culture. Universality and Variation*. Cambridge: Cambridge University Press.
- Kövecses, Zoltan. 2020. *Extended Conceptual Metaphor Theory*. Cambridge: Cambridge University Press.
- Malone, Nicholas and Kathryn Ovenden. 2017. Natureculture. In: *The International Encyclopedia of Primatology*. Ed. Augustin Fuentes. 1–2. John Wiley & Sons, Ltd.  
<https://doi.org/10.1002/9781119179313.wbprim0135>
- Markoš, Anton (ed.). 2010. *Jazyková metafora živého*. Praha: Pavel Mervart.
- Niebrzegowska-Bartmińska, Stanisława. 1986. Słońce raduje się – metafora czy mit? *Akcent*, 4 s. 37-39.
- Nöth, Winfried. 1990. *Handbook of Semiotics*. Bloomington / Indianapolis: Indiana University Press.
- Sapir, David J. – J. Christopher Crocker (eds.). 1977. *The Social Use of Metaphor: Essays on the Anthropology of Rhetoric*. University of Pennsylvania Press.
- Sharifian, Farzad. 2017. *Cultural Linguistics*. Amsterdam / Philadelphia: John Benjamins.
- SSiSL – *Słownik stereotypów i symboli ludowych, t. 1. Kosmos. cz. 1. Niebo, światła niebieskie, ogień, kamienie*, Ed. Jerzy Bartmiński. Lublin: Wydawnictwo UMCS 1996.
- Stibbe, Arran. 2021. *Ecolinguistics: Language, Ecology and the Stories We Live By*. 2nd ed. New York and London: Routledge.
- Underhill, James. 2009. *Humboldt, Worldview, and Language*. Edinburgh: Edinburgh University Press.
- Völkel, Svenja and Nico Nassenstein (eds.). 2022. *Approaches to Language and Culture*. [Anthropological Linguistics 1.] Berlin/Boston: De Gruyter Mouton.
- Wilcox, Phylis P. 2000. *Metaphor in American Sign language*. Washington, D. C.: Gallaudet University Press.
- Yu, Ning. 2008. Metaphor from body and culture. In: *The Cambridge Handbook of Metaphor and Thought*, ed. Raymond W. Gibbs, Jr. 247-261. Cambridge: Cambridge University Press.  
<https://doi.org/10.1017/CBO97805411816802.016>

## Obsah / Table of Contents

*Mimikry jako biometafory, kterými přežíváme, aneb: vtělesnění znaku a kognitivní nástroje (nejen) u zvířat*

Róbert Bohát (Univerzita Karlova)

*Metaphors We Learn By – Or Don't? Corpus Comenius As a Probe Into the Students' Metaphors*

Róbert Bohát (Charles University)

*Metaphor matters*

Adam Głaz (Maria Curie-Skłodowska University in Lublin, Poland)

*Fluid litost, "blocked" litost: Metaphorical puzzles in Milan Kundera's The Book of Laughter and Forgetting*

David S. Danaher (University of Wisconsin-Madison)

*Jazyk buňky: metafora jako vlastnost živého*

Anton Markoš (Univerzita Karlova v Praze, Přírodovědecká fakulta)

*Recept na politický guláš: konceptuální sféra JÍDLO v metaforách českého politického diskurzu*

Ekaterina Rycheva (Univerzita Karlova)

*Strojová a jazyková metafora živého*

Jana Švorcová (Charles University)

*Metafora a metamorfóza. Kontext magického obrazu světa a jeho odraz v moderní poezii*

Irena Vaňková (Univerzita Karlova)

*Metafora a části těla. Kaskády významových přenosů v tradičním českém folkloru*

Veronika Vodrážková (Univerzita Karlova)

### Mimikry jako biometafory, kterými přezíváme, aneb: vtělesnění znaku a kognitivní nástroje (nejen) u zvířat

Róbert Bohát (Univerzita Karlova)

Je možné produktivně aplikovat Teorii kognitivní metafory (Cognitive Metaphor Theory, CMT) na studium mimikry? Na základě několika případových studií navrhuji pracovat s biologickým napodobováním (mimikry) jako s druhem biosémiotické metafory. Mezi kognitivní metaforou v lidské kognici a mimikry u zvířat totiž existují nejméně dvě významné paralely. První paralelou je skutečnost, že jak metafora, tak i mimikry jsou příkladem sémiotického *přenosu* (etymologicky: *meta-fora*) totožnosti či znaku ze zdroje (nebo modelu) na vnímanou totožnost či znak cíle (mimika). Jedním z prvních kroků k ověření hypotézy je srovnání definic a jejich vyjádření v matematických symbolech. Definice metafory jako procesu identifikace nebo reprezentace znaku A jako znaku B by se v matematických symbolech dala shrnout následovně:  $Me = \{\exists A = B \Leftrightarrow A \neq B\}$ , t.j. metafora (Me): existuje takové A, které je B tehdy a jen tehdy, když A není B. Podobně mimikry – jako proces identifikace nebo reprezentace biologického druhu A jako druhu B – je v matematických symbolech:  $Mi = \{\exists A = B \Leftrightarrow A \neq B\}$ , t.j. mimikry (Mi): existuje takový druh A, který “je” druhem B tehdy a jen tehdy, když druh A není druhem B. Ze srovnání těchto definic potom vyplyne izomorfie: jestli platí, že  $Me = \{\exists A = B \Leftrightarrow A \neq B\}$  a  $Mi = \{\exists A = B \Leftrightarrow A \neq B\}$ , potom  $Me = Mi$ , t. j. metafora a mimikry jsou *procesuálně izomorfní*. Jak víme, že k takovému *přenosu* (řecky: *metafora*) vnímané totožnosti dochází? Jedním důkazem je odpovídající změna v reakcích (chování) lidských nebo zvířecích interpretů (interpretátorů) – u člověka např. obranná reakce před “vosou” při setkání s pestřenkou, nebo u zvířecích predátorů – např. mravenců – místo obvyklé predace housenky (modráska) s ní “jednají jako s vlastními larvami”, dokonce i jako s larvami královen (Pilátová 2014). Jinými slovy: nejen lidé, ale i zvířata svými reakcemi dokazují, že *vnímají* (“*chápu*”) *druh X jako druh Y* díky tomu, že se tělo i chování mimika stali vtělesněným znakem modelového druhu (srv. Maran 2011, 2017). Kognitivní metafora je proces, při němž “*chápeme pojmy jednoho druhu na základě pojmů jiného druhu*”; oba tyto sémiotické procesy tedy zahrnují “*nazírání jednoho druhu věci na základě jiného druhu věci*” a to je v rámci CMT nazýváno “*myšlením metaforickým*” (Lakoff a Johnson 2002 [1980], str. 128, 210). Samozřejmě, to neznamená, že by živočišný interpret byl schopen “*pracovat s metaforou*” reflektivně či analyticky tak, jako člověk; jednou její limitací je, že podobně jako jiné sémioticko-kognitivní paralely se i biometafora (mimikry) od lidské metafory liší mírou složitosti, reflexe, intence a stupněm uvědomování si metaforičnosti.

Druhá, ekologická paralela spočívá v tom, že druh obývající jednu niku (např. mol: nejedovatý herbivorní primární konzument) je vnímán jako druh obývající jinou niku (např. sršeň: jedovatý omnivorní predátor). Tak je ovlivněný *umwelt* (sensu Uexküll) potenciálního predátora tím, že vnímá např. mola (nesytku sršňovitou) jako “sršeň” a jeho reakce na tento podnět nebude predace ale vyhnutí se. Jelikož jsou tyto změny v chování systematické, nejde jen o povrchní podobnost a tak bychom tento proces mohli nazvat biosémiotickou metaforou (biometaforou), tj. MOL JE SRŠEŇ nebo KOŘIST JE PREDÁTOR (srv. Coletta 1993). Přirozeně, lidský pozorovatel si může uvědomit, že identifikace “sršně” je pouze v přeneseném smyslu, zatímco u zvířecích interpretů nám taková reflexe není známá.

Dalším společným jmenovatelem mezi mimikry a metaforou je, že tato (bio)metaforická identifikace nevyžaduje “dokonalou” či úplnou podobnost mezi zdrojovým a cílovým znakem (nebo druhem). To koresponduje s kategorizací na základě prototypů v CMT, tedy s “otevřenými” kategoriemi, při nichž pro metaforickou identifikaci stačí pouze částečná podobnost (srv. Lakoff, Johnson 1980; Rosch 1983). Tento paralelismus mezi mimikry a metaforou může být jazykovědně využit při hlubší analýze vtělesněnosti lidské i zvířecí kognice (embodied cognition) a zároveň by mohl poskytnout podporu jedné ze základních tezí CMT - jestli totiž zvířata i rostliny zažívají tento druh metaforické semiózy bez lidského jazyka, potom by metaforizace opravdu nebyla “povrchní” záležitostí jazyka, ale hlubokou

záležitostí vtělesněné mysli, jak argumentovali Lakoff a Johnson (1980; srv. Eco 2018: 350; Kull 2018, 352—364).

### **Bibliografie**

Coletta, W. John. (1993) The Semiotics of Nature: Towards an Ecology of Metaphor and a Biology of Mathematics. *The American Journal of Semiotics*, 10(3), 223–244.  
<https://doi.org/10.5840/ajs1993103/410>

Eco, Umberto, 2018. Giorgio Prodi and the lower threshold of semiotics. *Sign Systems Studies* 46(2/3), 343—351. DOI: 10.12697/SSS.2018.46.2-3.07

Kull, Kalevi, 2018. Umberto Eco on the biosemiotics of Giorgio Prodi. *Sign Systems Studies*, 46(2/3), 352—364. <https://doi.org/10.12697/SSS.2018.46.2-3.08>

Lakoff, George, Johnson, Mark, 1980. *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lakoff, George, Johnson, Mark, 2002 [1980]. *Metafory, kterými žijeme*. Brno. Host – vydavatelství, s.r.o.

Maran, Timo (2011). Becoming a sign: The mimic's activity in biological mimicry. *Biosemiotics*, 4, 243–257.

Maran, Timo (2017). *Mimicry and Meaning: Structure and Semiotics of Biological Mimicry*. (Biosemiotics 16). Dordrecht: Springer.

Pilátová, Jana. (2014). Jak si modrásek povídá s mravenci. *Vesmír* 93, 657, 2014/12.

Rosch, Eleanor, 1983. Prototype Classification and Logical Classification: The Two Systems. *New Trends in Conceptual Representation: Challenges To Piaget's Theory*, 73—86. Edited by Ellin Kofsky Scholnick.

Uexküll, Jakob von. (2010) *A Foray into the Worlds of Animals and Humans: With A Theory of Meaning*. Translated by Joseph D. O'Neill. Minneapolis: University of Minnesota Press.



**Metaphors We Learn By – Or Don't?**  
**Corpus Comenius As a Probe Into the Students' Metaphors**

Róbert Bohát (Charles University)

**Background:** What cognitive metaphors influence high school students' learning? The Cognitive (Conceptual) Metaphor Theory (CMT) established a theoretical framework for the study of metaphors as a deep, conceptual aspect of human cognition (Lakoff & Johnson 1980; Lakoff 1998). Corpus linguistics adds quantification as well as attested, authentic and usage-based data to this research (Partington 1995, Stefanowitsch & Gries 2008, Semino et al. 2015, Deignan 2009). Steen (2007) and the Praggeljaz group (2007) have proposed “an explicit and rigorous procedure for metaphor analysis” – the *Metaphor Identification Process of VU University of Amsterdam* (aka MIPVU method). Semino et al. (2015) applied this method in a corpus based analysis of VIOLENCE and JOURNEY metaphors in cancer patient discourse; the corpus data analysis showed that metaphors could empower or disempower cancer sufferers psychosomatically in coping with their disease and therapy.

**Methodology:** Inspired by Semino et al. (2015), I built the pedagogical *Corpus Comenius* (*CorCo*; in English, with ca. 1.2 million tokens) which offers an insight into the teachers' and students' learning metaphors. The pedagogical subcorpus contains administrative and professional texts, the student subcorpus contains more than 250 high school students' responses (at the International School of Prague) to the *Learning Metaphor Survey* (LMS; 47,000 tokens) with 3 prompts on learning and metaphors where students explain what metaphors empower or disempower them and why. The corpus is annotated for metaphors using the MIPVU method.

**Preliminary results:** In the student subcorpus of *CorCo*, a rather “technical” definition of learning in *Prompt 1* (“What is learning/teaching? How does it work?”) contains 78% “mechanical” metaphors (i.e. metaphors identifying learning as a machine-like process, e.g. LEARNING IS CONDUIT TRANSFER, MACHINE WORK, etc.); 13% of the metaphors are LEARNING IS A JOURNEY, while LEARNING IS A DISCOVERY made up 4% of the metaphors. *Prompt 2* solicits metaphors indirectly by asking: “What is learning like?” Mechanical metaphors made up ca. 39% of the total; “organic” metaphors (i.e. describing learning as an organic process, such as LEARNING IS GROWTH, SYMBIOSIS, PHOTOSYNTHESIS, etc.) constituted 29.5% and the LEARNING IS A JOURNEY metaphor ca. 21%. *Prompt 3* asks for a direct metaphor of learning that empowers them and the reasons why; the most frequent empowering metaphor is LEARNING IS A JOURNEY with 38%. For example, a Grade 9 student (aged 15) wrote: “I like the metaphor ‘Roadmap to a concept’ because learning is not something negative but something positive and helpful and a roadmap on a journey is very helpful. And by being able to understand the roadmap makes you reach your destiny.” The second most empowering category are organic metaphors with 18 %, followed by the LEARNING IS WAR metaphor with 15% and LEARNING IS PLAY(ING) with almost 13% (e.g. “It insinuates that you must continue working or “battling” for an extended period of time. It also gives a slight atmosphere of *competition* and for me, being a reasonably competitive person, this helps to motivate me to finish the problem.” Or: “Learning is like gameplay because *playing games motivates me* and inspires me to work, it takes us to higher levels.”).

**Implications:** The preliminary analysis of the data indicates that while the *status quo* of the 21st century learning (tested for by Prompt 1) is dominated by mechanical metaphors like LEARNING IS TRANSFER or MACHINE WORK, the two other prompts (testing for student preferences and empowerment) show that students prefer metaphors like LEARNING IS A JOURNEY and ORGANIC PROCESS, WAR & PLAY because they make them feel “positive, not negative” about learning, they “motivate” them, and strengthen them to reach “higher levels”, etc. Once the full frequency analysis of the student subcorpus learning metaphors is completed, the results will serve as a basis for the creation

of the **Learning Metaphor Toolkit** for students and teachers - similar to the medical *Metaphor Menu* that was recommended for cancer patients and their health care staff (Semino 2014).

## Bibliography

- Cameron, L. (2011). *Metaphor in educational discourse*. London: Continuum.
- Deignan, A. (2005) *Metaphor and Corpus Linguistics*. John Benjamins Publishing Company.
- Hanks, P. (2010) Nine issues in metaphor theory and analysis. *International Journal of Corpus Linguistics*, 15 (1). pp. 133-150. ISSN 1384-6655 Available from: <http://eprints.uwe.ac.uk/13008>
- Lakoff, G. (1998). The Contemporary Theory of Metaphor. In *Metaphor and Thought* (pp. 202-251). Cambridge, UK: Cambridge University Press.
- Pragglejaz Group. (2007). MIP: A Method for Identifying Metaphorically Used Words in Discourse. *Metaphor and Symbol*, 22(1), 1-39. doi:10.1080/10926480709336752
- Semino, E., Deignan, A., & Littlemore, J. (2013). Metaphor, Genre, and Recontextualization. *Metaphor and Symbol*, 28(1), 41-59. doi:10.1080/10926488.2013.742842
- Semino, E. (2014, December 17). A 'metaphor menu' for cancer patients. Retrieved April 22, 2017, from <http://www.ehospice.com/uk/ArticleView/tabid/10697/ArticleId/13414/language/en-GB/View.aspx>
- Semino, E., Demjén, Z., Demmen, J., Koller, V., Payne, S., Hardie, A., & Rayson, P. (2015). The online use of Violence and Journey metaphors by patients with cancer, as compared with health professionals: a mixed methods study. *BMJ Supportive & Palliative Care*, 7(1), 60-66. doi:10.1136/bmjspcare-2014-000785
- Steen, G. (2007). Finding metaphor in discourse: Pragglejaz and beyond. *Cultural Studies Journal of Universitat Jaume I*, V(2007), 9-25. Retrieved from <http://www.e-revistas.uji.es/index.php/clar/article/view/1348>
- Svoboda, K. (1947). Tropy a figury. [Naše řeč, ročník 31 \(1947\), číslo 1](#), s. 4-12

## Metaphor matters

Adam Głaz (Maria Curie-Skłodowska University in Lublin, Poland)

In his science-fiction novel *Embassytown*, China Miéville explores the idea of a language without counterfactuals and so without metaphor. As the author admits in an interview with Joshua Miller (2019), this linguistically impossible situation makes him more and more convinced that in human language the opposite is true: metaphor is everything.

Metaphors require departure from factuality to perspecivization. Some authors have called them lies that tell the truth (e.g. Pinkard 2000) – but what kind of truth and whose truth do they tell? Metaphors are not about reality, they are about us, about what we see, how we see it, how we want to see it, and what stories we want to tell. However, there are many “we’s”, living in different metaphorical worlds, sometimes very subtly different: where Polish speakers are likely to say *w Polsce* (‘in Poland’, objectified and detached), English speakers would often say *back home* (a more intimate message grounded in the HOME metaphor for one’s country), while Ukrainians are heard as talking about *наша Україна* (‘our Ukraine’, dear to us and ours for life).

Our metaphors also change through time, in response to how we change in tone or axiology: consider the case of *welfare state*, a profoundly positive notion that originally referred to the social progress made by many European countries after WWII, but one that has recently acquired negative connotations through re-contextualization and re-framing (Grzegza 2016).

On a more general level, metaphors are being employed as a survival strategy in the face of climate catastrophe (here’s one!). Those can range from eco-conscious but anthropocentric discourse of the Earth as our house and home, and of humans as a family (see Głaz 2022), to a life-centred story of our planet, or even a broader story of the entire universe, where humans need to relinquish their privileged position (Stibbe 2021). New metaphors are being examined, explored, and experimented with, such as the notion of *Great Jurisprudence*, i.e. “the fundamental laws or principles that govern how the universe functions” (Cullinan 2007: 72): it is rather remarkable that in attempts to move beyond anthropocentrism, we are profiling the Universe as a system of laws that does best when governed with knowledge, wisdom, frugality, and care.

This shows that metaphors are hard to deal with, and yet they allow us to create worlds and construct realities (see Grace 1987) in the form of symbolic spaces in which we live, make sense of our experience and try (more or less successfully) to communicate that experience to others.

## Bibliography

- Cullinan, Cormac. 2007. Sowing wild law. *Environmental Law & Management* 19: 71-76.
- Głaz, Adam. 2022. *Linguistic Worldview(s): Approaches and Applications*. New York and London: Routledge.
- Grace, George W. 1987. *The Linguistic Construction of Reality*. London–New York–Sydney: Croom Helm.
- Grzega, Joachim. 2016. Alternative European values in European headlines: Competitiveness, privatization, solidarity, socialization, welfare state. *Journal for EuroLinguistiX* 13: 18–27. <https://www1.ku.de/SLF/EngluVglSW/ELiX/grzega-161.pdf>
- Miéville, China. 2011. *Embassytown*. Pan Macmillan & Del Rey Books.
- Miller, Joshua R. 2019. Metaphor is everything: A conversation with China Miéville. *Michigan Quarterly Review* 58(4): 595–604.
- Pinkard, Carolyn. 2000. *Lies that Tell the Truth. A Book of and about Metaphor*. San Jose, New York, Lincoln and Shanghai: Writer's Showcase.
- Stibbe, Aran. 2021. *Ecolinguistics: Language, Ecology and the Stories We Live By*. 2nd ed. London and New York: Routledge.

**Fluid *lítost*, “blocked” *lítost*:  
Metaphorical puzzles in Milan Kundera’s *The Book of Laughter and Forgetting***

David S. Danaher (University of Wisconsin-Madison)

**Topic.** I examine the key word *lítost* in Milan Kundera’s *The Book of Laughter and Forgetting*, and I do so across a number of textual variants in three languages [1, 2, 3, 4, 5]. *Lítost* is a key word in the novel that Kundera insisted remain both untranslated and unglossed in translations of the book, and it features as both the title of and a frequently repeated word in the fifth part of the novel. Eschewing a global reading of this emotion concept as a key word in the book [see 6], I will focus instead on two puzzles that present themselves to a serious reader of the work across textual variants and that both turn, each in its own way, on metaphorical matters.

**First puzzle.** The first of these concerns the aesthetic embedding of *lítost* in a major metaphorical subtext in the novel that is related to fluid—more specifically to water, tears, and semen. While Kundera does take care to anchor the emotion concept in this metaphorical subtext through the language of the work itself, it is also the case that emotion concepts across languages have more or less stable aspects of meaning associated with them [7], two of which are physiological embodiment (a happy person may smile) and conventional metaphorical representation (anger may be understood as a heated fluid that puts pressure on the body-as-container). Using the Czech National Corpus to analyze *lítost* from this perspective, I argue that it is an emotion associated with fluid in terms of both embodiment and metaphor, and that this facilitates Kundera’s aesthetic embedding.

**Second puzzle.** The second puzzle relates to the phrase *vzpříčení lítosti* [2] that Heim translates as “impacted *lítost*” [1] but that Kundera consistently altered in later textual variants: Asher has “blocked *lítost*” [3], Kérel has “un blocage de la *lítost*” [4], and the updated Czech-language version has “zablokovaná *lítost*” [5] Using the Czech National Corpus once again, I argue that the shift from *vzpříčení* (which in English can be translated literally as “jamming”) to “blockage” is motivated in large part by metaphorical considerations that are, in turn, tied to more general considerations related to translatability.

**Conclusion.** This contribution is, then, a focused application of Contemporary Metaphor Theory (CMT) with help from corpus linguistics analysis to the aesthetics of one particularly complex text. The solutions I present to these two puzzles both illustrate the targeted analytical power of CMT and, at the same time, enhance our appreciation of the text’s literariness.

## Bibliography

- [1] Kundera, Milan. 1980. *The Book of Laughter and Forgetting*, trans. Michael Henry Heim. New York: Knopf.
- [2] Kundera, Milan. 1981. *Kniha smíchu a zapomnění*. Toronto: Sixty-Eight Publishers.
- [3] Kundera, Milan. 1994. *The Book of Laughter and Forgetting*, trans. Aaron Asher. New York: HarperCollins.
- [4] Kundera, Milan. 2011. *Le livre du rire et de l'oubli*, trans. François Kérel, in Milan Kundera, *Oeuvres I (édition définitive)*, 927-1136. Paris: Bibliothèque de la Pléiade.
- [5] Kundera, Milan. 2017. *Kniha smíchu a zapomnění*. Brno: Atlantis.
- [6] Danaher, David S. Forthcoming. *Lítost* as category of existence in *The Book of Laughter and Forgetting*, in *Milan Kundera Known and Unknown: Multidimensional Analysis of Selected Works*, ed. Karen von Kunes. New York: Bloomsbury.
- [7] Kövecses, Zoltán. 1990. *Emotion Concepts*. New York: Springer.

### Jazyk buňky: metafora jako vlastnost živého

Anton Markoš (Univerzita Karlova v Praze, Přírodovědecká fakulta)

„Opravdová, imaginativní metafora, [...] vyjadřuje účastné poznání [participant knowledge] a je schopna je zprostředkovat“, říká O. Barfield (2015, 40) a myslí tím lidské vlastnosti. M. Ajvaz (2023, 307-8) naopak představuje hylozoický model kosmu s „jazykem“, „vědomím“ – a tudíž i schopností analogizovat a tvořit metafory. Budu zde držet střední cestu a diskutovat metaforické schopnosti života, definovaného jako entitu založenou na buňkách; podklady čerpám z prací, jichž jsem byl spoluúčasten (Markoš, ed. 2010; Markoš Švorcová 2019; Horský & Markoš 2023).

Lidský jazyk (promluvy) lze redukovat na lineární ne-monotonní zápis s pomocí konečného souboru značek (abecedy). Podobně DNA, RNA a proteiny (níže jen „DNA“ jako pars pro toto pro tyto biopolymery) jsou lineárními nemonotonními zápisy s použitím konečného souboru značek. Jde v druhém případě o zápis – redukcí jazyka, anebo jde o pouhý strojový „jazyk“, tj. program?

Budeme-li pracovat s hypotézou, že lineární biopolymery jsou zápisy jazyka analogickými s jazykem (tj. představují jeho redukcí do skriptu), pak jako „mluvčí“ vystoupí buňky a mnohobuněčné útvary a jejich „jazyk těla“. Zmíněné biopolymery jsou pak redukcí této řeči.

Zpětný převod zápisu přirozeného jazyka *nutně* vyžaduje interpretaci zápisu; z toho plyne, že „čtenáři“ budou ovlivněni podle výsledku svého interpretačního úsilí. Navíc může být zápis opatřen diakritikou (chemické i strukturní značky), kterou lze přidávat, umazávat, anebo uchovávat přes řadu generací. Dospíváme tak k poznání, že biozápisy mohou být současně „kanonické“ i proměnlivé, a v každé generaci buněčné linie jim může být rozuměno různě.

Ještě zajímavější je přenos zápisů mezi různými liniemi buněk, což vede k nutnosti porozumění různým dialektům (např. při splynutí pohlavních buněk) i abdukci při přijetí polymerů neznámých (např. virové aj. formy tzv. horizontálního přenosu.).

Otázka zní: dokážeme takový interpretační proces odhalit a popsat? V té chvíli dojde i na metafory. V návaznosti na příspěvky J. Švorcové a R. Bohata se pokusím také o některé příklady, které by se takto daly interpretovat.

## **Bibliografie**

Ajvaz M (2023) *Dynamický logos. O významové stavbě světa*. Praha, Filosofia.

Barfield O (2015) *Básnická řeč*. Praha, Malvern.

Deely J (2010) *medieval philosophy revisited*. University of Scranton Press.

Horský J & Markoš A (2023) *Pohyby semen a tvarů. Výprávění v dějepsectví a evoluční biologii*.  
Červený Kostelec, Pavel Mervart.

Markoš A ed. (2010) *Jazyková metafora živého*. Červený Kostelec, Pavel Mervart.

Markoš A & Švorcová J (2019) *Epigenetic processes and the evolution of life*. CRC Press.



**Recept na politický guláš: konceptuální sféra JÍDLO v metaforách českého politického diskurzu**

Ekaterina Rycheva (Univerzita Karlova)

Konceptuální metafory hrají v politickém diskurzu důležitou roli, protože usnadňují komunikaci politiků s veřejností, pomáhají vysvětlit složité politické problémy a představit je ve srozumitelnější podobě, vytvářejí emocionální spojení mezi politikem a publikem a posilují atraktivitu sdělení, protože použití silné a výstižné metafory může přesvědčit posluchače o správnosti určitého politického postoje nebo opatření.

Jednou z nejproduktivnějších konceptuálních sfér v politickém diskurzu je sféra JÍDLO. Národní kuchyně je nejstarší součástí kultury lidí a jejich jazykového obrazu světa. V moderním českém politickém diskurzu se hojně používají názvy typických produktů (houby, mák, švestka, jablko atd.) a pokrmů tradiční české kuchyně (svíčková, guláš, koláč, knedlík, rohlík, omáčka atd.) jako konceptuální metafory (např. POLITIKA JE JÍDLO, POLITIK JE KUCHARĚ, POLITICKÝ PROGRAM JE GULÁŠ, MOC JE KOLÁČ). Konceptuální metafory se zdrojovou sférou JÍDLO se používají jak k vysvětlení složitých politických jevů, tak k diskreditaci politických oponentů (*byrokratický guláš, předvolební guláš, mediální guláš, levicový guláš* atd.). Politici jsou prezentováni jako kuchaři (*U prezidenta Miloše Zemana byla vždy patrná touha být režisérem či šéfkuchařem povolebního dění.*) a chování politika je srovnáváno s kuchařem, který připravuje jídlo (*vařit rozpočet, recept na pečení totality*).

V projevech politiků se pro zvýšení emocionality a expresivity řeči často používají frazeologické jednotky a ustálené spojení s „gastronomickými“ komponenty (*ani za mák, opít rohlíkem, potřebovat jako sůl, chodit kolem horké kaše, pečení holubi nelitají do huby* atd.). Politici používají frazeologické jednotky a přísloví nejen v užití, ale i v transformované podobě pro aktualizaci ustálené jednotky v určitém kontextu (*dostat státní houby s octem, ukrojit z volebního koláče, nechat v náruči horký brambor, hlad po moci je špatný volební kuchař*). Důležitým rysem používání konceptuálních metafor se zdrojovou sférou JÍDLO v českém politickém diskurzu je snaha mluvčího ukázat svůj negativní postoj k situaci nebo jevu (*Vidina velkých drobků z velkého rozpočtového koláče bledne před zadlužením.*).

## **Bibliografie**

- [1] Gibbs, Raymond W. Jr. (2019). *Metaphor Wars: Conceptual Metaphors in Human Life*. Cambridge University Press.
- [2] Kövecses, Zoltán (2010). *Metaphor — a practical introduction*. Oxford: Oxford University Press.
- [3] Lakoff, George (2009). *The Political Mind: A Cognitive Scientist's Guide to Your Brain and Its Politics*. Penguin Books, New York.
- [4] Lakoff, George (2016). *Moral Politics. How Liberals and Conservatives Think*. The University of Chicago Press.
- [5] Lakoff, George, Johnson, Mark (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- [6] Nekula, Marek (2000). *Metafora a poznání*. *Estetika* 36, 15–25.
- [7] Teun A. van Dijk (1977). *Text and context: explorations in the semantics and pragmatics of discourse*. Longman Linguistics Library, 21. London: Longman.
- [8] Teun A. van Dijk (1995). *Discourse semantics and ideology* Teun A. van Dijk *Discourse & Society* Vol. 6, No. 2, pp. 243-289.
- [9] Vaňková, Irena (2002). *Lingvistika mysli a těla*. *Vesmír* 81, č. 11, 627–629.
- [10] Vaňková, Irena (2007). *Nádoba plná řeči. Člověk, řeč a přirozený svět*.
- [11] Vaňková, Irena, Nebeská, Iva, Saicová Římalová, Lucie, Šlédrová, Jasňa (2005). *Co na srdci, to na jazyku*. Karolinum.

**Strojová a jazyková metafora živého**

Jana Švorcová (Charles University)

Strojová metafora trvale, nekriticky a invazivně proniká naše uvažování o přírodě a živém již od novověku. Tato metafora plynule sedimentovala nejen v rámci naší konceptualizace živých organismů, ale prostupuje i texty vysokoškolských učebnic či uznávaných biologických časopisů věnovaných buněčným, neuronálním či evolučním jevům (Alberts a kol. 2002). Historicky se analogie mezi organismy a stroji datuje již od Descartese, přičemž zvolený archetyp se dále vyvíjel ruku v ruce se společenským pokrokem. V 17. století byly organismy přirovnávány k hodinám, v 18. století k parním strojům, v 19. století k továrnám a od 20. století se stalo rutinou přirovnání lidské mysli k počítačům. V molekulární biologii se často používá pojem továrna ve zřetelné snaze vidět složité procesy v buňce, jako jsou vzájemné interakce proteinů, jako procesy, které mají efektivitu, výrobní kapacitu a vnitřní mechanismy. V tomto rámci slouží jaderná DNA jako forma programu a organismy pak jako kybernetické entity řídící se tímto předem stanoveným průběhem kódování a dekodování (Dawkins 1976, 1995; Alberts 1998; Horwitz and Johnson 2017). Podle tohoto pohledu mají geny kauzální vliv na fenotyp a ontogenezi organismů, přičemž každý aspekt organismálního fenotypu je pevně určen a řízen programem (Noble 2006; Laland 2013). V důsledku toho je DNA vnímána pouze jako instruktivní informace, nikoli jako informace, kterou je třeba interpretovat podobně třeba jako lidské texty. Příspěvek se zaměří na analýzu dostupné odborné literatury (především recenzovaných zahraničních článků v kvalitních časopisech), ale i na analýzu učebnic či populárně-vědeckých časopisů, na nichž se ukáže míra i samozřejmost používání strojové metafory i její pevná usazenost v našem vědeckém jazyce i uvažování.

Příspěvek se pokusí zodpovědět i další otázky: v čem spočívají koncepční nedostatky metafory stroje a proč na ni věda stále spoléhá? Existuje životaschopná alternativa k programové metafoře života a měli bychom ji hledat? Naší snahou je navrhnout alternativní, nemechanistickou metaforu živého, která odráží podstatu živých organismů v jejich zkušenosti, historicitě a svébytnosti. Život a jeho procesy jsou pak analogické používání psaného jazyka a jeho neustálé interpretaci (living as languaging, Markoš et al. 2018). Příspěvek by tak měl posloužit i jako jakýsi prequel příspěvků Antona Markoše a Roberta Boháta.

## Bibliografie

- Alberts, B (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell*, 92(3): 291-4
- Alberts, B, Johnson, A, Lewis, J, Raff, M, Roberts, K, Walter, P (2002) *Molecular Biology of the Cell*. Garland Science
- Dawkins, R (1976) *The Selfish Gene*. Best Books
- Dawkins, R (1995) *The River out of Eden*. Basic Books
- Horwitz, R, Johnson, GT (2017) Whole cell maps chart a course for 21st-century cell biology. *Science*, 356: 806-807
- Laland KN, Odling-Smee J, Hoppitt W, Uller T (2013) More on how and why: a response to commentaries. *Biology and Philosophy* 28(5): 793-810
- Markoš, A, Švorcová, J, Lhotský, J (2017) Living as Linguaging: Distributed Knowledge in Living Beings. In: Cowley, S, Vallée-Tourangeau, F (eds) *Cognition Beyond the Brain*. Springer, Cham
- Noble, D (2006) *Music of Life. Biology beyond Genes*. Oxford Univeristy Press

## Metafora a metamorfóza. Kontext magického obrazu světa a jeho odraz v moderní poezii

Irena Vaňková (Univerzita Karlova)

Co všechno může znamenat, že metaforami žijeme, formulovali nejen Lakoff a Johnson (2002) a pak řada jejich pokračovatelů (srov. např. Semino – Demjén, eds., 2016). Už mnohem dříve poukázal k metafoře a metonymii jako ke dvěma aspektům (resp. „pólům“) jazyka Roman Jakobson (1995a), a to ve velmi širokém kontextu a mj. i s ohledem na kognitivní a kulturní aspekty lidského sémiotického chování včetně umění (srov. i Nöth, 1990). Kromě dvou typů afatických poruch připomněl i další polohy metaforičnosti a metonymičnosti: Saussurův zřetel ke kódu nebo ke kontextu, pierceovské ikony a indexy, ale i Freudův snový symbolismus a ve vazbě na magický vztah člověka k realitě také poznatek z Frazerovy *Zlaté ratolesti*, že i principy magických obřadů mají dvojí charakter: metafoře odpovídá magie imitativní a metonymii magie kontaktní (Frazer 1994). Referát míří právě do této sféry – k principům ustrojení magického obrazu světa, zejm. k metafoře, která ho významně spoluutváří, a též k metamorfóze, již lze v jistém ohledu pokládat za její specifický typ (Arutjunova 1998, Valček 2006).

Moc metafory konstruovat žitou realitu (Kövecses 2020: 6) je vždy jevem vázaným na určitý kontext, na kulturu daného společenství i specifický náhled na svět spojený s jistým modelem lidského prožívání, přemýšlení a komunikace, s jistým typem chápání reality i místa člověka v ní. Magický (resp. mytický) obraz světa, jehož principy se realizují ve folklorních textech (v pohádkách, pověstech, baladách a třeba i v pranostikách), představuje dnes kontext dosti specifický. Jerzy Bartmiński tu mluví o mýtu oživeného kosmu, který „uvnitř sebe a sám se sebou komunikuje“ a jehož integrální součástí je i člověk (Bartmiński 2016). Svědectví najdeme v heslech lublinského *Slovníku lidových stereotypů a symbolů* (SSiSL): slunce se směje, člověk s ním může rozprávět, hvězdy se na nás shora dívají: v poezii by šlo o antropomorfační metaforu, ale dávný venkovský člověk tu rozuměl doslovně (Niebrzegowska 1986, o vývoji metaforizačních procesů v kontextu lidového obrazu světa srov. i Łozowski 2023). V takto žitém kosmu se vše vzájemně zrcadlí a každá část světa je zároveň zrcadlovým odrazem celého univerza (Kajfosz 2011). Běh světa přitom zajišťují zákony imitativní a kontaktní magie (eo ipso metafory a metonymie) a nic není náhodné. V magickém obrazu světa mají tyto principy ontologickou platnost: řídí běh světa (Kajfosz 2009, 2011) a s jejich pomocí lze běh světa někdy dokonce ovlivnit, např. (ale nejen) slovem (Vaňková 2000).

Dílem, na jehož základě lze podobu magického obrazu světa (a světa mýtu, srov. Jakobson 1995b) dobře rekonstruovat, je Erbenova *Kytice* (Erben 2003). V baladách se pregnantně realizují principy magické sémiotiky (Kajfosz 2011, Vaňková 2011) a setkáváme tu s příklady („žitých“) metafor řídicích životaběh aktérů: se snovými symboly (*Vodník*), s obrazy na vodní hladině (*Štědrý den*), ale i s vizemi národní budoucnosti (*Věštkyně*). Tyto obrazy jsou vlastně zdrojovými oblastmi metafor, jež odkazují k budoucím, v perspektivě textu často ještě neznámým oblastem cílovým. Jsou tu i příklady metamorfóz: proměna zemřelé ženy v mateřidoušku, jež náhle vykvétá na jejím hrobě a v níž děti po vůni poznávají matku (a v dalším posunu analogicky – díky pověstem – vnímají takto důvěrně i „matku vlast“), proměna zavražděného manžela v holoubka či magická podvojnost ženy-lilie či ženy-vrby.

Takto budované kontexty magického obrazu světa, resp. svěbytného světa mýtu (srov. i Otruba 1994) s jeho specifickou ontologií i etikou jsou též inspirací pro moderní literaturu. Příkladů lze uvést mnoho, zde se soustředíme na poezii Jiřího Wolkeru (2021). Je třeba aspoň zmínit, že i v uměleckém textu, resp. v poetické vizi světa má metafora mnohá specifika, resp. že se tu „konvenční metafora užívá nekonvenčně“, že podléhá extenzím, elaboracím, kombinacím a různě explicitním problematizacím (srov. Lakoff – Turner 1987, Vaňková 2010, Vaňková 2022) a to platí i pro Wolkerovu poezii. Způsobů realizace metafory je v ní mnoho, např. poetické definice a jejich modifikace, např. *Slunce je veliký básník* (17), *Já nejsem žena, / já jsem hrob* (66). Obraz skutečnosti má u Wolkeru mytický charakter (rozhovory s věcmi, mystické propojení dějů a jevů, zástupnost osudů) a výsostné místo v konstruování tohoto obrazu patří metamorfóze. Oči se mění ve světlo, ruce v chléb; kamení rozkvétá; kavárna se stává hřbitovem. I smrt je proměnou: *snad stanu se stromem, snad děckem, snad hromadou kamení* (139).

Mezi metamorfózou ve folkloru (a v *Kytici*) a metamorfózou ve Wolkerově poezii (či v jiných literárních textech) je jistě rozdíl: jde o různé kontexty a obrazy světa, z nichž se ten druhý tak či onak inspiruje prvním. Snad tu lze rozlišit dvojí fenomén: „metaforickou metamorfózu“ (utvářenou v rámci

magického obrazu světa především ve folkloru a také u Erbeny) a „metamorfickou metaforu“, již při stylizaci (či v připomínce) magického obrazu světa využívá v různých modifikacích moderní literatura.

## Bibliografie

- Arutjunova, Nina D.: Metafora i metamorfoza. In: *Jazyk i mir čeloveka*. Moskva: Jazyki ruskogo kul'tury, 1998, s. 356–357.
- Bartmiński, Jerzy: In: Formy přítomnosti *sacrum* ve folkloru. In: *Jazyk v kontextu kultury. Dvanáct statí z lublinské kognitivní etnolingvistiky*. Praha: Karolinum, 2016, s. 49–58.
- Erben, Karel Jaromír: *Kytice – České pohádky*. Praha: Nakladatelství Lidové noviny, 2003.
- Frazer, James George: *Zlatá ratolest. Magie, mýty, náboženství*. Praha: Mladá fronta 1994.
- Jakobson, Roman: Dva aspekty jazyka a dva typy afatických poruch. In: *Poetická funkce*. Jinočany: H+H, 1995a, s. 55–73.
- Jakobson, Roman: Poznámky k dílu Erbenovu. In: *Poetická funkce*. Jinočany: H+H, 1995b, s. 500–530).
- Kajfosz, Jan: *Magia w potocznej narracji*. Katowice: Wydawnictwo Uniwersytetu Śląskiego, 2009.
- Kajfosz, Jan: Magie a řád světa v Erbenově Kytici. In: *Karel Jaromír Erben a úloha paměťových institucí v historických proměnách. 2. díl. Z Českého ráje a Podkrkonoší. Supplementum 14*. Semily – Turnov 2011, s. 34–45.
- Kövecses, Zoltán: *Extended Conceptual Metaphor Theory*. Cambridge: Cambridge University Press, 2020.
- Lakoff, George – Johnson, Mark: *Metafory, kterými žijeme*. Brno: Host, 2002.
- Lakoff, George – Turner, Mark: *More than Cool Reason. A Field Guide to Poetic Metaphor*. Chicago – London 1987.
- Łozowski, Przemysław: Wschód słońca w kontekście dyskusji o niedosłowności znaczeń w metaforze i metonimii. *Etnolingwistyka*, 2023, 35, s. 169–184.
- Niebrzegowska Stanisława: Słońce raduje się – metafora czy mit? *Akcent*, 1986, 4, s. 37–39.
- Nöth, Winfried: *Handbook of Semiotics*. Bloomington – Indianapolis: Indiana University Press, 1990.
- Otruba, Mojmir: Kreační mýtu poezii (Erbenova balada Vrba). In: *Týž, Znaky a hodnoty*. Praha: Československý spisovatel 1994, s. 125–147.
- Semino, Elena – Demjén, Zsófia (eds.): *The Routledge Handbook of Metaphor and Language*. London: Routledge, 2016.
- SSiSL: *Słownik stereotypów i symboli ludowych I (1-4), II (1-7)*. J. Bartmiński – S. Niebrzegowska Bartmińska (eds.). Lublin: UMCS, 1996–2022.
- Valček, Pavel: *Slovník literárnej teórie. A-Ž*. Bratislava: Literárne informačné centrum, 2006. Heslo „Metamorfóza“, s. 225.
- Vaňková, Irena: „Mocné slovo“, řeč a komunikace v Erbenově Kytici. *Česká literatura*, 48, 2000, č. 6, s. 630–635.
- Vaňková, Irena: Slovo v poezii. Inspirace kognitivnělingvistické. In: S. Fedrová (ed.): *Česká literatura v intermedialní perspektivě. IV. kongres světové literárněvědné bohemistiky*. Praha: Akropolis – Ústav pro českou literaturu AV ČR, 2010, s. 425–437.
- Vaňková, Irena: Přichyl ucho k zemi blíž. Erbenova Kytice a sémiotické principy magického obrazu světa. In: *Karel Jaromír Erben a úloha paměťových institucí v historických proměnách. 2. díl. Z Českého ráje a Podkrkonoší. Supplementum 14*. Semily – Turnov, 2011, s. 46–65.
- Vaňková, Irena: „A teď mi to řekněte jinými slovy.“ Umělecké texty v kognitivnělingvistické a kognitivněpoetické perspektivě. Několik poznámek. *Jazykovědné aktuality*, LIX, č. 3–4, 2022, s. 73–86. Dostupné na: [https://www.jazykovednesdruzeni.cz/wp-content/uploads/2023/01/JA-3\\_4\\_2022-K-TISKU.pdf](https://www.jazykovednesdruzeni.cz/wp-content/uploads/2023/01/JA-3_4_2022-K-TISKU.pdf)

Wolker, Jiří: *Básně*. Praha – Brno: Host, 2021.

## Metafora a části těla. Kaskády významových přenosů v tradičním českém folkloru

Veronika Vodrážková (Univerzita Karlova)

Teorie konceptuální metafory (Lakoff – Johnson, 1980) vychází z obecného předpokladu, že jedna pojmová oblast (cílová) je pojmána jako oblast jiná (zdrojová), tradičně konkrétnější, názornější než oblast první. Jedná se o základní model metaforického přenosu. V případě metonymie dochází k posunu významu pouze v jedné oblasti. Naopak s více dimenzemi, resp. mentálními prostory, které se mísí ve zdroji, pracuje teorie mísení, blendingu (Fauconnier – Turner, 2003; Turner, 2005). V jazykových dokladech můžeme nalézt příklady, které poukazují na uplatnění možná ještě složitěji kombinovaných významových přenosů. Z teoretického hlediska se tedy v našem příspěvku ptáme, k jakým kaskádám významových přenosů dochází, jak se uplatňují metafory, metonymie, případně blendy v kombinaci.

Příkladem jedné z významových kaskád může být frazém „vykročit pravou nohou“ – noha je metonymickým zástupcem celkového scénáře pohybu, chůze; pravosti je přisuzováno štěstí; průběh celého dne nebo života je pak připodobňován metaforicky k pohybu po cestě, resp. k chůzi. Tato pověra je zachycena nejen ve frazeologii, ale také v popisu některých realizovaných zvyklostí, např. „vstávati se má pravou nohou napřed“ (Bartoš, 1892, s. 311). Spojitost štěstí a pravosti je specifickým vztahem, jedním z mnoha vztahů, které tvoří zákonitosti svébytného (magického) světa zrcadleného ve folkloru (např. Kajfosz, 2006; Vaňková, 2011). Ptáme se tedy mimo jiné i na to, do jaké míry se tyto vztahy promítají do kaskád významových přenosů.

V představovaném výzkumu je analyzován folklor verbální (písňe, rčení, říkadla, pranostiky apod.) a částečně i neverbální (rituály, pověry, dětské hry, léčebné postupy apod.), a to zejména ze sbírek *Naše děti* (Bartoš, 1888), *Moravský lid* (Bartoš, 1892), *Prostonárodní české písňe a říkadla* (Erben, 1864), *Moravské národní písňe s nápěvy do textu vřaděnými* (Sušil, 1860) aj.

Ve verbální i neverbální sféře folkloru jsou vyhledávány části lidského těla. Právě u tělesných jevů se totiž předpokládá, že budou v rámci konceptuálních metafor tvořit primárně zdrojovou oblast, oblast zakoušenou smysly a zažitou tělem (a tvořící tzv. představová schémata, Johnson, 1987; k tělesné bázi percepce, kognice i jazyka také Gibbs, 2005). Kromě toho je zapojení tělesných částí typické pro úkony doprovázející hry nebo říkanky (např. pot'ukávání patičky dítěte při „kovej, kovej, kovářičku, okovej mi mou nožičku“).

Výsledky analýzy budou mimo jiné součástí komplexního výzkumu českých somatických pojmů, který bude zaměřen nejen na folklor.



## Bibliografie

Bartoš, František. 1888. *Naše děti. Jejich život v rodině, mezi sebou, v obci, jejich poesii, zábavy, hry i práce společné.* Brno: Mor. akc. knih. tiskárny. Dostupné z <<https://kramerius5.nkp.cz>>.

Bartoš, František. 1892. *Moravský lid. Františka Bartoše sebrané rozpravy z oboru moravské lidovědy.* Telč: Emil Šolc. Dostupné z <<https://kramerius5.nkp.cz>>.

Brouček, Stanislav – Jeřábek, Richard (eds.). 2007. *Lidová kultura. Národopisná encyklopedie Čech, Moravy a Slezska.* Praha: Mladá Fronta 2007.

Erben, Karel Jaromír. 1864. *Prostonárodní české písně a říkadla.* Praha: J. Pospíšil. Dostupné z <<https://kramerius5.nkp.cz>>.

Gibbs, Raymond W. 2005. *Embodiment and Cognitive Science.* Cambridge: Cambridge University Press.

Johnson, Mark. 1987. *The Body in the Mind. The Bodily Basis of Meaning, Imagination, and Reason.* Chicago: The University of Chicago Press.

Fauconnier, Gilles – Turner, Mark. 2003. *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities.* Chicago – London: Basic Books.

Kajfosz, Jan. 2006. Pranostika jako relikv magického obrazu světa. *Kuděj: Časopis pro kulturní dějiny* 2, s. 22–34.

Lakoff, George – Johnson, Mark. 1980. *Metaphors We Live By.* Chicago: The University of Chicago Press.

Sušil, František. 1860. *Moravské národní písně s nápěvy do textu vřaděnými* Brno: tiskem a nákladem Karla Winikera Dostupné z <<https://kramerius5.nkp.cz>>.

Vaňková, Irena. 2011. „Přichyl ucho k zemi blíž.“ *Erbenova Kytice a sémiotické principy magického obrazu světa.* In: Piorecká Kateřina, Navrátil Ivo (eds.): Karel Jaromír Erben a úloha paměťových institucí v historických proměnách. Referáty z vědecké konference konané ve dnech 15. – 16. dubna 2011 na Malé Skále. 2. díl. (Z Českého ráje a Podkrkonoší. Supplementum 14.) Semily – Turnov 2011

Turner, Mark. 2005. *Literární mysl. O původu myšlení a jazyka.* Brno: Host.

# Poster session

Bienále české lingvistiky 2024  
Biennial of Czech Linguistics 2024

## Obsah / Table of contents

*Elektronická lexikální databáze indo-iránských jazyků: Persko-český slovník*

Vladimír Benko (Slovenská akadémia vied a Univerzita Komenského v Bratislavě), Zuzana Kříhová (Filozofická fakulta, Univerzita Karlova), Boris Lehečka (Moravská zemská knihovna v Brně), Darina Vystrčilová (Sociologický ústav Akademie věd ČR)

*Morphosyntax of Czech binary gender derivation: A nanosyntactic perspective*

Daniel Boucník (MU)

*Stylometric features and readability of Czech legal texts*

Silvie Cínková (Charles University)

*Jak změřit český pravopis*

Václav Horký (ÚČNK FF UK)

*Pravopisná a hláskoslovná analýza Bible cisterciácké (1456)*

Anna Michalcová (Charles University)

*Piš, jak slyš: analýza automatických a manuálních přepisů nahrávek nerodilých mluvčích češtiny*

Michal Novák, Peter Polák, Kateřina Rysová, Magdaléna Rysová, Ondřej Bojar (Univerzita Karlova, Matematicko-fyzikální fakulta)

*Vývoj českého přechodníku přítomného ve srovnání s jeho chorvatským a slovinským protějškem neboli byl český přechodník předurčen k zániku?*

Aleksandra Wójcik (Jagellonská univerzita)

## Elektronická lexikální databáze indo-íránských jazyků: Persko-český slovník

Vladimír Benko (Slovenská akadémia vied a Univerzita Komenského v Bratislavě),  
Zuzana Kříhová (Filozofická fakulta, Univerzita Karlova),

Boris Lehečka (Moravská zemská knihovna v Brně),

Darina Vystrčilová (Sociologický ústav Akademie věd ČR)

Příspěvek představí projekt zaměřený na převod lexikálních dat dříve využitých pro publikaci tří česko-perských, resp. persko-českých slovníků (Vystrčilová, 2002; 2014; 2017) do nově vytvořené databáze v rámci projektu ELDI (Elektronická lexikální databáze indoíránských jazyků. Pilotní modul perština). Data byla nejprve převedena z Excelu do formátu FLEx (FieldWorks Language Explorer, SIL International, 2018)<sup>24</sup>. Projekt využívá FLEx pro import, sdílení, sběr a další zpracování lexikálních dat, s češtinou a angličtinou jakožto analytickými jazyky. Ověřená data z FLExu jsou pravidelně transformována pomocí aplikace v programovacím jazyce XProc 3.0<sup>25</sup> z formátu LIFT<sup>26</sup> do formátu TEI Lex-0<sup>27</sup> pro zveřejnění na webových stránkách projektu. Webová aplikace využívá XML databázi eXist-db a framework TEI Publisher, upravený pro práci s lexikografickými daty ve formátu TEI Lex-0. Mezi její stěžejní funkce patří vyhledávání, třídění výsledků podle frekvence lemmat v korpusech, filtrování podle faset, propojení s novým perským korpusem vytvořeným pro účely projektu a hypertextové odkazy na externí zdroje.

Struktura slovníkových hesel vychází z tradičních lexikografických přístupů. Zahrnuje zápis lemmatu v perském písmu, jeho fonetickou transkripci, event. pravopisné a gramatické varianty, frekvenci výskytu lemmatu v jazykových korpusech, gramatické informace, překlad, příkladové věty, křížové odkazy a nadřazené sémantické domény (Moe, 2012).

V počátečních fázích byla data čerpána ze tří perských textových korpusů (Uppsala Persian Corpus, Seraji, Megyesi & Nivre, 2012), perské verze databáze Open Subtitles (Tiedemann, 2016) a Persian Talkbank dostupného z webu Sketch Engine<sup>28</sup>. Později začala být využívána data z nového perského korpusu Araneum Persicum (Benko, 2022), vytvořeného přímo pro potřeby projektu, využívajícího nástroje FLOSS (Baroni a Bernardini, 2004; Pomikálek, 2011; Suchomel a Pomikálek, 2012; Michelfeit et al., 2014). Dostupné jsou také nástroje pro morfologicko-syntaktickou anotaci perských textů, jako jsou TreeTagger (Schmid; 1994), UDPipe (Straka et al., 2016) a CSTlemma (Jongejan a Dalianis, 2009), natrénované na Seraji UD Treebank (Seraji 2016).

Perská lexikální databáze i nový korpus jsou integrovány do vzdělávacích programů výuky perského jazyka na Filozofické fakultě UK a využívány v překladatelské praxi.

### Poděkování

Prezentovaný výzkum byl částečně finančně podpořen agenturami TAČR (Projekt č. TL0300369) a VEGA (Projekt č. 2/0016/21).

<sup>24</sup> <https://software.sil.org/fieldworks/>

<sup>25</sup> <https://xproc.org>

<sup>26</sup> <https://github.com/sillsdev/lift-standard>


<sup>27</sup> <https://bit.ly/tei-lex-0>

<sup>28</sup> <https://www.sketchengine.eu/talkbank-persian-corpus/>

Nalezeno položek: 78

1 2 3 4 > >|

▶ هوا [havá] 1. subst *vzduch ; ovzduší*. 

▶ آب و هوا [áb-o-havá] subst *počasí ; podnebí ; klima*. 

▶ تهویه هوا [tahvije(-je havá)] subst *ventilace ; větrání ; klimatizace*. 

▶ هواپیما [havá-pejmá] subst *letadlo ; letoun*. 

Dotaz pro: základní pole  
Zadejte dotaz  
Q havá

Slovník

Persko-český slovník 78

Výchozí jazyk

perština 78

Cílový jazyk

čeština 76

Slovní druh  Zobrazit vše (7)

## Bibliografie

- Benko, V.: Aranea Go Middle East: Persicum. In Aleš Horák, Pavel Rychlý, Adam Rambousek (eds.): Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2022, pp. 1-9. Brno: Tribun (2022).
- Baroni, M., & Bernardini, S. BootCaT: Bootstrapping Corpora and Terms from the Web. In Proceedings of LREC 2004, pp. 1313-1316 (2004).
- Jongejan, B, and Dalianis, H.: Automatic training of lemmatization rules that handle morphological changes in pre-, in- and suffixes alike. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. Suntec, Singapore : Association for Computational Linguistics, pp. 145-153 (2009).
- Michelfeit, J., Pomikálek, J., Suchomel, V.: Text Tokenisation Using unitok. In 8th Workshop on Recent Advances in Slavonic Natural Language Processing, Brno, Tribun EU, pp. 71–75 (2014).
- Moe, R., Semantic Domains. <https://semdom.org/> (2012).
- Pomikálek, Jan. Removing boilerplate and duplicate content from web corpora. PhD thesis, Masaryk university, Faculty of informatics, Brno, Czech republic (2011).
- Schmid, H.: Probabilistic Part-of-Speech Tagging Using Decision Trees. Proceedings of International Conference on New Methods in Language Processing, Manchester (1994).
- Seraji, M., Ginter, F., Nivre, J.: Universal Dependencies for Persian. In Proceedings of the 10th International Conference on Language Resources and Evaluation (LRE), pp. 2361–2365 (2016).
- Seraji, M., Megyesi, B. & Nivre, J., A Basic Language Resource Kit for Persian. In Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12) (pp. 2245–2252). European Language Resources Association (2012). [http://www.lrec-conf.org/proceedings/lrec2012/pdf/338\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/338_Paper.pdf).
- Straka, M., Hajič J., Straková J.: UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC), Portorož, Slovenia (2016)
- Suchomel, V., Pomikálek, J. Efficient Web Crawling for Large Text Corpora. In Adam Kilgarriff, Serge Sharoff. Proceedings of the seventh Web as Corpus Workshop (WAC7). Lyon, pp. 39–43 (2012).
- Tiedemann, J.: Finding Alternative Translations in a Large Corpus of Movie Subtitle. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 3518–3522) (2016). <https://aclanthology.org/L16-1559.pdf>.
- Vystrčilová, D.: Příruční slovník persko-český. Farhang-e hamráh-e fársí-čekí (Persian-Czech Desk Dictionary). PCHE (2002).
- Vystrčilová, D.: Goftegu.cz. Elektronický tematický česko-perský a persko-český slovník. Vlastním nákladem, (2014). <https://goftegu.cz/>.
- Vystrčilová, D.: Česko-perský slovník. Farhang-e čekí-fársí. Vlastním nákladem (2017).

## Morphosyntax of Czech Binary Gender Derivation: A Nanosyntactic Perspective

Daniel Boucník (MU)

The gender shift between Masculine and Feminine nouns is considered to be the most well explored area of grammatical gender in Czech linguistics (Straková 1980, Čmejrková 2002, Ziková 2003), yet most of the existing literature covers lexicalist approaches, leaving morphological analyses underrepresented. In the Czech binary gender morphology, a prototypical profession noun ‘učitel’ (1) is grammatically Masculine (e.g. requiring Masculine agreement in modifying adjectives, as shown in (2)). The Feminine form of the noun can be derived from it by adding the suffix *-ka* (same example). Moreover, in Czech, Masculine grammatical gender is often used for the broad meaning, encompassing all individuals within a group irrespective of the social notion of gender (e.g. ‘all teachers’ as in (3)). The derived form ‘učitelka’ then carries a more specialized meaning for a subset of all teachers – those who identify as female.

- (1) učitel → učitel -ka  
teacher.M.SG.NOM teacher -F.SG.NOM  
base form – *broad meaning* derived form – *specialized meaning*
- (2) šikovný učitel  
skilled.M.SG teacher.M.SG
- (3) všichni učitelé na této škole  
all.M.PL teacher.M.PL at this school  
‘all the (male and female) teachers at this school’
- (4) žáb -a → žab -ák  
frog -F.SG.NOM frog -M.SG.NOM  
base form – *broad meaning* derived form – *specialized meaning*

This paper aims to fill the gap in the morphosyntactic literature on grammatical gender of Czech by proposing an account couched in the framework of nanosyntax (cf. Starke 2018, Caha 2009, 2021, Caha, Taraldsen-Medová 2022, Blix 2022). Nanosyntax is a theory of morphology within which syntax builds a structure first, following the building instruction in the form of the universal functional sequence of features (fseq), an example of a feature-geometry approach (Harley, Ritter 2002), which can take the form of e.g. NP < Masculine < Feminine < Singular < Nominative. Within nanosyntactic approach, this structure is only then matched with the lexicon to compile the structure itself, the meaning and phonology, and by that the theory creates fertile soil for purely formal descriptions of the morphology. In the footsteps of Janků (2022) and Starke (2024), this paper adopts a purely formal approach by rebranding the Masculine and Feminine features as Gender and Markedness respectively, enabling descriptions of base and derived forms of binary gender shifts as the shifts between the purely grammatical gender (without interpretation) and the specialized and restricted gender which only allows gender interpretation. Such a rebranding serves as a step towards a purely grammatical representation of morphology. A detailed modelling of the how to derive both the base forms and the derived forms is proposed, utilizing the lexicalization algorithm of Starke (2018). This work contributes to the understanding of Czech grammatical gender by analyzing how morphemes create masculine and feminine forms. It demonstrates how the nanosyntax framework, using the lexicalization algorithm, can derive both masculine-to-feminine (“učitel” -> “učitelka”) and feminine-to-masculine (“žába” ->

“žabák”) shifts. This paves the way for analyzing other parts of the nominal domain such as profession and animal nouns in nanosyntax.

## Bibliography

Blix, Hagen. “Interface Legibility and Nominal Classification: A Nanosyntactic Account of Kipsigis Singulatives.” *Glossa: A Journal of General Linguistics* 7 (September 22, 2022).

<https://doi.org/10.16995/glossa.5825>.

Caha, Pavel. “Modeling Declensions without Declension Features. The Case of Russian in: Acta Linguistica Academica Volume 68 Issue 4 (2021).” *Acta Linguistica Academica*, June 3, 2021, 385–425. <https://doi.org/10.1556/2062.2021.00433>.

———. “The Nanosyntax of Case.” Dissertation, University of Tromsø, 2009.

<https://munin.uit.no/bitstream/handle/10037/2203/thesis.pdf>.

Caha, Pavel, and Lucie Taraldsen Medová. “On the feminine singular declension of the Russian demonstrative,” 2022. <https://www.muni.cz/vyzkum/publikace/2222757>.

Čmejrková, Světlá. “Rod v jazyce a komunikaci: specifika češtiny.” *Slovo a slovesnost* 63, no. 4 (November 2002): 263–87.

Harley, Heidi, and Elizabeth Ritter. “Person and Number in Pronouns: A Feature-Geometric Analysis.” *Language* 78, no. 3 (2002): 482–526.

Janků, Lucie. “The Nanosyntax of Czech Nominal Declensions.” Doctoral dissertation, LingBuzz, 2022. <https://lingbuzz.net/lingbuzz/006705>.

Starke, Michal. “Complex Left Branches, Spellout, and Prefixes.” In *Exploring Nanosyntax*, edited by Lena Baunaz, Liliane Haegeman, Karen De Clercq, and Eric Lander. Oxford University Press, 2018.

<https://doi.org/10.1093/oso/9780190876746.003.0009>.

Starke, Michal. “Nanosyntax” – Lectures at Masaryk University (2023-2024)

Straková, Vlasta. “Slovotvorný význam a sémantické rysy.” *Slovo a Slovesnost* 41, no. 2 (1980).

<http://sas.ujc.cas.cz/archiv.php?art=2667>.

Ziková, Markéta. “K podstatě slovotvorného procesu přechylování.” *Sborník Prací Filozofické Fakulty Brněnské Univerzity, Řada jazykovědná*, 52, no. A51 (2003): 125–32.

## Stylometric features and readability of Czech legal texts

Silvie Cinková (Charles University)

Democratic and transparent governments must keep citizens informed about their rights and obligations. The Czech government requires clear language in law bills and has committed to implementing clear communication between the administration and the public until 2030 (Mlsna / Sláma 2019). Also, language practitioners in law and pedagogy promote and teach plain, client-oriented writing in their professional communities (Vučka 2019; Šamánková / Kubíková 2022; Šváb 2021). At the same time, recent academic research on Czech clear writing has also been predominantly conducted by lawyers (Dvořáková 2021) and pedagogists (Halfarová 2018; Pluskal 1996).

We present findings from our experiment with a small corpus of paraphrased Czech legal texts (Anonymous). The corpus contains six legal documents. Each document has been provided with two paraphrases. Paraphrases to all originals were written by the same two lawyers committed to plain legal writing. Each of these six document triples was subject to a randomized on-line reading-comprehension survey (Arslan / Walther / Tata 2020), with at least 30 responses for each document.

In a preliminary analysis of the results, the version author turned out to be the strongest reading-comprehension predictor, and therefore this study focuses on the stylometric features of the authors. Fig.1 presents the cumulative probability density of reading comprehension by author, across all documents. The maximum score per document was 7, with the scale granularity 0.25. The curve corresponding to *jasa* has a longer tail and a gentler slope up to approximately the score of 6, compared to the curves representing the other two authors. This shows that weak readers confronted with *jasa*'s text versions achieve higher scores than weak readers of other's versions (shorter tail, steeper slope). This effect gradually disappears in strong readers (approximately 6.5 – 7 points).

In three of the six documents, *jasa*'s readers outperformed the readers of other versions with high statistical significance and considerable effect size (multiple comparison Dunn's test, epsilon squared). In the other three, the difference was not statistically significant. This nevertheless corresponds to what we see in Fig.1, since Dunn's test compares the differences in medians (which were not much different indeed), while the cumulative probability density differs in weaker-to-mediocre readers but keeps almost equal to the others at the extremes.

We compared the reading comprehension on all six text version triples with the following stylometric metrics, using QuitaUp (Cvrček / Čech / Kubát 2020) and PMLTQ (Janssen 2021; Pajas / Štěpánek 2009) with UDPipe (Straka 2018): Shannon Entropy, Hapaxes, Thematic Concentration (Čech 2016), Lexical Diversity (zTTR), Verb Distance, Activity /Descriptivity, Frequency of passive constructions, Sentence length distribution, Noun phrase length distribution, and Syntactic complexity (Jagaiah / Olinghouse / Kearns 2020). The texts written by *jasa* turned out significantly different from both other authors' texts in Hapaxes and Shannon Entropy. Another significant difference appeared between *jasa* and *orig* (original legalese) in Passive Constructions (*jasa* avoided them). Given that the metrics cover both syntactic and lexical aspects of the texts, it is somewhat surprising that the linguistic aspects do reflect the differences in reading comprehension between the authors to such a humble extent. On the other hand, with the content and genre being controlled for, there was possibly not enough leeway for linguistic variation. Lawyers suggest that the main difference is in the conceptual coherence of the content, especially in argumentative texts (where *jasa*'s readers also performed best): a good text supports each conclusion with a reference to the relevant law and to the story of the case, which ought to match well enough to make the conclusion superfluous for its obviousness. The lessons learned from this experiment with controlled content and varying style are: the observation of clear writing tips alone does not guarantee a comprehensible text. And a hypotheses-generating assumption arises at the same time: a clear line of thoughts is the main principle of clear writing, and it both suppresses and compensates for deviations from the overt syntactic and lexical plain language guidelines.



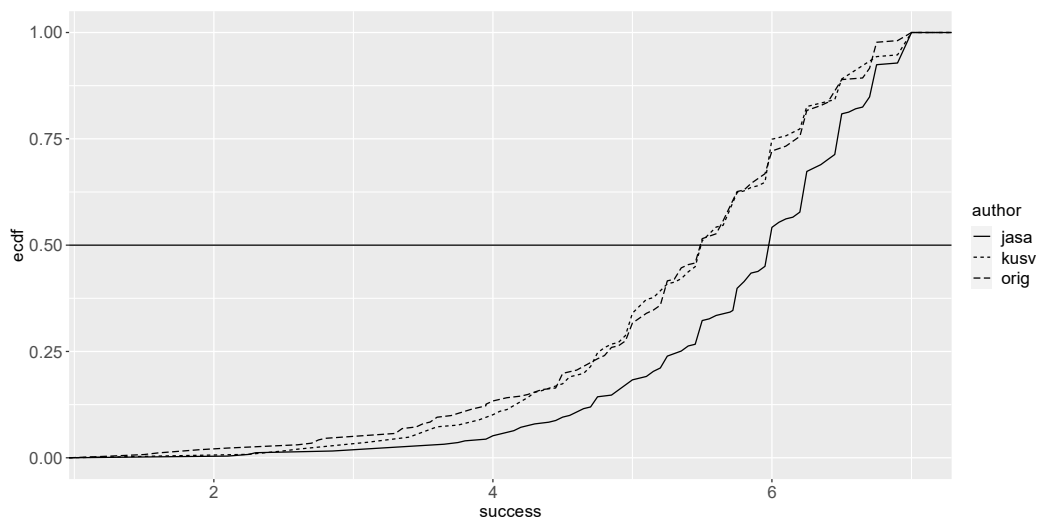


Figure 6 Cumulative probability density of reading comprehension by author, across documents.

## Bibliography

- Arslan, Ruben C. / Walther, Matthias P. / Tata, Cyril S.** (2020): "formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R", in: *Behavior Research Methods* 52 (1): 376–387. 10.3758/s13428-019-01236-y.
- Čech, Radek** (2016): *Tematická koncentrace textu v češtině*. (= Studies in computational and theoretical linguistics) Praha: Ústav formální a aplikované lingvistiky.
- Cvrček, Václav / Čech, Radek / Kubát, Miroslav** (2020): *QuitaUp – a tool for quantitative stylometric analysis*. Czech National Corpus and University of Ostrava. <https://korpus.cz/quitaup/> [letzter Zugriff 18. November 2022].
- Dvořáková, Eva** (2021): "Plain Language in the Czech Republic, the UK, and the US: An Analysis of Tax Return Instructions", in: *Studie z aplikované lingvistiky* 1: 7–19.
- Halfarová, Libuše** (2018): *Comprehensibility of Legal Text and Its Understanding by People with Specific Needs*. in: *Proceedings of the International Scientific Conference ECONOMIC AND SOCIAL POLICY*. Čeladná, Czech Republic: Vysoká škola PRIGO. 81–90. <https://www.narodacek.cz/history-of-conference-proceedings/>.
- Jagaiah, Thilagha / Olinghouse, Natalie G. / Kearns, Devin M.** (2020): "Syntactic complexity measures: variation by genre, grade-level, students' writing abilities, and writing quality", in: *Reading and Writing* 33 (10): 2577–2638. 10.1007/s11145-020-10057-x.
- Janssen, Maarten** (2021): *Integrating TEITOK and KonText/PMLTQ at LINDAT*. in: *Selected Papers from the CLARIN Annual Conference 2020* (= Linköping Electronic Conference Proceedings). Linköping, Sweden: Linköping University Electronic Press, Linköpings universitet. 104–110. (= Linköping Electronic Conference Proceedings).
- Mlsna, Petr / Sláma, David** (2019): *Klientsky orientovaná veřejná správa 2030. Koncepce - verze pro vnitřní i vnější připomínkové řízení*. Ministerstvo vnitra ČR.
- Pajas, Petr / Štěpánek, Jan** (2009): *System for Querying Syntactically Annotated Corpora*. in: *Proceedings of the ACL-IJCNLP 2009 Software Demonstrations*. 33–36.
- Pluska, Miroslav** (1996): "Zdokonalení metody pro měření obtížnosti didaktických textů", in: *Pedagogika* XLVI: 63–78.
- Šamánková, Jana / Kubíková, Barbora** (eds.) (2022): *Jak psát srozumitelné úřední texty. Příručka srozumitelného psaní pro úředníky*. 1. Aufl. Brno: Kancelář veřejného ochránce práv.
- Straka, Milan** (2018): *UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task*. in: *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*. Brussels, Belgium: Association for Computational Linguistics. 197–207. <https://www.aclweb.org/anthology/K18-2020>.
- Šváb, Jakub** (2021): *Jak psát, aby se to dalo číst : příručka přístupného psaní*. (= Praktik) Vydání první Praha: Leges.
- Vučka, Jan** (2019): *Právní psaní*. C.H. Beck.

## Jak změřit český pravopis

Václav Horký (ÚČNK FF UK)

V tomto příspěvku ukážu, jak je možné vyčíslit míru, jakou se uplatňuje v pravopise hláskové písemné soustavy fonologický a morfematický princip, tzn. princip vzájemně jednoznačného vztahu mezi fonémy a grafémy, resp. morfémy a jejich písemnými obrazy, a představím předběžné výsledky pro fonologický princip v českém pravopise.

**Dosavadní práce o českém pravopise.** Teoreticky se už věnovala českému pravopisu a jeho principům řada prací (Kunstovný 1914, Vachek 1933, Sgall & Novák 1964, Starý 1992, Sgall 1994, Uličný 2015 ad.). Neshodují se v tom, jak vyložit některé pravopisné jevy, ani v tom, jestli je důležitější princip fonologický, morfematický, nebo ještě jiný – např. podle P. Sgalla (1994) je to princip fonologický, podle O. Uličného (2018) morfematický, podle J. Vachka provádí český pravopis jejich „obdivuhodnou syntézu“ (1942) a podle A. Frinty je český pravopis historický (1909).

Naproti tomu kvantitativních prací, které by mohly pomoci s odpovědí na některé sporné otázky (srv. Uličný 2013), je málo a vycházejí z malých vzorků nebo nejdou do hloubky (Kučera 1998, Caravolasová a kol. 2012, Uličný a kol. 2016) nebo se týkají jen dílčích otázek (Těšitelová 1964, Sgall 1978/79).

**Možnosti měření principů** K tomu, abychom změřili rozsah fonologického a morfematického principu, potřebujeme reprezentativní korpus, který má pro každý slovní tvar zarovnanou grafémovou a fonémovou podobu, jejich členění na morfy a přiřazení morfů k morfémům (viz příklad na druhé straně). Z toho je pak možné odvodit přechody – dvojice odpovídajících si vstupních a výstupních jednotek – mezi písemnými jednotkami a fonémy a morfémy: 1) foném – řetěz grafémů, 2) grafém – řetěz fonémů, 3) morfém – písemný (nebo také zvukový) morf, 4) písemný (nebo také zvukový) morf – morfém.

Složitost jednotlivých polí přechodů – skupin přechodů se stejnou vstupní jednotkou –, která odráží, jak se u nich uplatňuje fonologický a morfematický princip, je možné měřit několika mírami. Nejjemnější z nich je Simpsonův index – průměr poměrných četností jednotlivých přechodů vážený jimi samými. Ten odráží jak počet přechodů, tak poměrné četnosti všech z nich; čím vyšší je jeho hodnota, tím méně je pole přechodů složité. (Této míry používá Caravolasová a kol. 2012; stejně jemná je také entropie nebo perplexita, srv. Borleffsová a kol. 2017, Milička 2022.) Hrubší míra je četnost nejčetnějšího přechodu, nejhrubší a skoro bezcenná je prostá jednoznačnost (tzn. jestli obsahuje pole jediný přechod). Průměr výsledků pro jednotlivá pole (vážený jejich četnostmi) potom odráží, jak se uplatňuje fonologický a morfematický princip v pravopise jako celku.

**Předběžné výsledky pro český pravopis.** Na desíti tisících slovních tvarů s nejvyšší průměrnou redukovanou četností (ARF, viz Křen 2017) v korpusu SYN2020 (tedy na třech čtvrtinách jeho tokenů, ale méně než setině typů) jsem zjišťoval, jak se uplatňuje fonologický princip v českém pravopise. K tvarům jsem vytvořil zarovnané fonologické přepisy a z toho odvodil pole přechodů mezi fonémy a grafémy (bez ohledu na velikost písmen), a to jednak ničím nepodmíněná, jednak podmíněná sousedními jednotkami. Pak jsem pro ně spočítal různé míry složitosti a jejich průměry – viz tabulku na druhé straně. Hodnoty měř v ní uvádím z hlediska tokenů, ale i z hlediska typů vypadají podobně; z hlediska typů by však možná vypadaly významně jinak, kdyby se zahrnuly i méně četné tvary.

I když se pokládá obecně čtení za jednodušší než psaní (Sgall & Novák 1964/65; Kučera 1998), průměrný Simpsonův index nepodmíněných polí přechodů od fonémů k řetězům grafémů je vyšší (0,87) než polí od grafémů k řetězům fonémů (0,84). (Ke stejnému vztahu dospívá i Caravolasová a kol. 2012.) Při podmínění oběma sousedními jednotkami je však index polí od fonémů nižší (0,97) než index polí od grafémů – ten je skoro čistá jedna, tzn. čtení je s ohledem na sousední jednotky skoro jednoznačné. Stejně jako vztahy mezi průměrnými Simpsonovými indexy jsou i vztahy mezi průměrnými četnostmi nejčetnějšího přechodu. Jiné vztahy a podstatně nižší hodnoty jsou u jednoznačnosti – skoro žádné pole přechodů není naprosto jednoznačné; srovnání s předchozími mírami však také ukazuje, jak málo tato míra vypovídá.

Příklad slovního tvaru:

Grafémy a členění na morfy	s   c h ů z   k   y
Fonémy a členění na morfy	s   x . ú s   k   i
Morfémy	s   chod   k   y[NFP1]

Průměrné míry polí přechodů – Simpsonův index (S), četnost nejčastějšího přechodu (M) a jednoznačnost (J):

Levý, pravý kontext	Od fonémů k řetězům grafémů			Od grafémů k řetězům fonémů		
	S	M	J	S	M	J
0, 0	0,87	0,91	0,08	0,84	0,89	0,19
1, 0	0,94	0,96	0,60	0,90	0,92	0,43
0, 1	0,90	0,93	0,53	0,97	0,98	0,72
1, 1	0,97	0,98	0,86	1,00	1,00	0,95

## Bibliografie

- Borleffsová, E. a kol. (2017): Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: a narrative review. *Reading and Writing*.
- Caravolasová, M. a kol. (2012): Common Patterns of Prediction of Literacy Development in Different Alphabetic Orthographies. *Psychological Science*, s. 678–686.
- Frinta, A. (1909). *Novočeská výslovnost*. Česká akademie císaře Františka Josefa pro vědy, slovesnost a umění.
- Křen, M. (2017): ARF. In: P. Karlík – M. Nekula – J. Pleskalová (eds.), *CzechEncy – Nový encyklopedický slovník češtiny*. Online: <https://www.czechency.org/slovník/ARF>.
- Kučera, K. (1998). Vývoj účinnosti a složitosti českého pravopisu od konce 13. do konce 20. století. *Slovo a slovesnost*, s. 178–199.
- Kunstovný, O. (1914). O českém pravopise a jeho opravě. Dědictví Komenského.
- Milička, J. (2022): *Lexikální diverzita*. FF UK, habilitační práce.
- Sgall, P. (1978/79). K některým problémům pravopisu a kodifikace spisovné češtiny. *Český jazyk a literatura*, s. 302–308.
- Sgall, P. (1994). Lingvistický pohled na český pravopis. *Slovo a slovesnost*, s. 168–177, 270–286.
- Sgall, P. – P. Novák (1964). Diskuse o pravopise. K racionalizaci českého pravopisu. *Slovo a slovesnost*, s. 126–135.
- Sgall, P., – P. Novák (1964/65). Ke způsobu uvažování (o pravopisu a jiných „historických danostech“). *Český jazyk a literatura*, s. 315–320.
- Starý, Z. (1992). *Psací soustavy a český pravopis*. Univerzita Karlova.
- Těšitelová, M. (1964). Diskuse o pravopise. Otázky homonymie slov a slovních tvarů rozlišených jen psaním i a y. *Slovo a slovesnost*, s. 201–212.
- Uličný, O. (2013). K fonologickým rysům českého pravopisu. In *Studie k moderní mluvnici češtiny. 5, K české fonetice a pravopisu* (s. 139–143). Univerzita Palackého v Olomouci.
- Uličný, O. (2015). K teorii českého pravopisu. In *Preliminária k moderní mluvnici češtiny* (s. 267–272). Univerzita Palackého v Olomouci. Online: <https://sites.google.com/site/mluvnicecestiny/publikacni-vystupy>
- Uličný, O. (2018). Subsystem českých grafémů a principy českého pravopisu. *Slavia*, s. 373–381.
- Uličný, O. a kol. (2016). Ke zkoumání struktury a funkcí českého pravopisu. *Bohemica Olomucensia*, s. 108–120.
- Vachek, J. (1933). Český pravopis a struktura češtiny. *Listy filologické*, s. 287–319.
- Vachek, J. (1942). Psaný jazyk a pravopis. In *Čtení o jazyce a poesii* (s. 229–306). Družstevní práce.

Děkuji prof. Cvrčkovi a třem anonymním recenzentům. Tuto práci podpořil program Univerzitní výzkumná centra UK č. 24/SSH/009.

## Pravopisná a hláskoslovná analýza Bible cisterciácké (1456)

Anna Michalcová (Charles University)

Samohlásková kvantita představuje jednu z nejméně prozkoumaných oblastí ve vývoji češtiny. Zvláště omezující pro její zkoumání je její nepravidelné označování ve starších vývojových fázích. Představa o situaci staročeské kvantity není spolehlivá; tato problematika je ještě více komplikována nejistotami ohledně původu pračeské kvantity a faktem, že mnoho jevů se objevilo v různých časových obdobích a ovlivnilo pouze určité oblasti jazykového regionu. Přestože v spisovném jazyce prošla většina dlouhých samohlásek odlišným vývojem než jejich krátké protějšky, čímž je umožněna rekonstrukce některých starších kvantit (změny  $ó > uo > ů$ ,  $é > í$ ,  $ú > ou$ ,  $y > ej$ ,  $ie > í$ ), samohlásky [á] a [í], které zůstávaly beze změn, a latinská tradice záznamu, která nevyznačovala délku samohlásek, představují významné překážky pro spolehlivou analýzu zápisu kvantity v českých středověkých rukopisech. Tyto faktory činí rekonstrukci staročeské kvantity značně obtížnou, což ji staví na pozici jednoho z posledních velkých nevyřešených úkolů diachronní strukturální fonologie češtiny (Vintr 2005: 192; Kosek 2014: 66).

V tomto ohledu je zcela výjimečný pramen tzv. *Bible cisterciácké* z roku 1456 (New York City, The Morgan Library & Museum, sign. MS M.752, více viz Kvas 1997: 106). *Bible cisterciácká* poskytuje vysoce reprezentativní zápis diakritického pravopisu, jenž může s vysokou mírou spolehlivosti odrážet tehdejší výslovnost staré češtiny. V *Bibli cisterciácké* najdeme v naprosté většině případů odpovídajícím způsobem zapsanou jak kvalitu souhlásek, tak kvantitu samohlásek (např. *Ptáky nebejské a ryby mořské. kterěz przecházegí stezky mořské* Ps 8,9). V tomto příspěvku představíme předběžnou analýzu kvantity samohlásek v tomto rukopisu, přičemž se zaměříme na oblasti, které byly již dříve identifikovány jako problematické v odborné literatuře, viz např. Komárek 1962, Gebauer 1963, Pisaro 2002, Nedvěďová (interní dokument ÚJČ AV ČR), Synková 2017.

Konkrétně se zaměříme na dva klíčové typy: 1) dvojslabičné a víceslabičné tvary přivlastňovacích zájmen *náš* a *váš*, u kterých se v rukopisu objevují zaznamenané délky samohlásek, což koresponduje s předpoklady historických mluvnic (např. Gebauer 1960: 486–491; Vážný 1963: 127–128); 2) *nt-ová* participia, u kterých od 14. století podle odborné literatury (Gebauer 1958: 57; Dvořák 1970: 71–73) docházelo ke krácení samohlásky v sufixu (*-uc-*, *-ic-*, *-ěc > -ec-*), což potvrzují i naše předběžné analýzy rukopisu. Tyto nálezy ukazují na významnost tohoto pramene pro diachronní lingvistiku a jeho přínos pro hlubší pochopení kvantity v historii češtiny.

V rámci příspěvku bude rovněž podrobně představena metodologie použitá při analýze rukopisu, která se opírá o využití umělé inteligence a HTR (Handwritten Text Recognition), tj. technologii automatického rozpoznávání rukopisného textu. Pro přípravu automatické transkripce je využívána platforma Transkribus (více např. Muehlberger a kol. 2019, Romein a kol. 2024). Tyto postupy reflektují rostoucí zájem o digital humanities a představují v kontextu české lingvistiky relativně nový přístup.

## Bibliografie

- Dvořák, E. (1970). *Vývoj přechodníkových konstrukcí ve starší češtině*. Praha: Univerzita Karlova.
- Gebauer, J. (1958). *Historická mluvnice jazyka českého, Díl III, Tvarosloví, II. Časování*. Praha.
- Gebauer, J. (1960). *Historická mluvnice jazyka českého, Díl III, Tvarosloví, I. Skloňování*. Praha.
- Gebauer, J. (1963). *Historická mluvnice jazyka českého, Díl I, Hláskosloví*. Praha.
- Komárek, M. (1962). *Historická mluvnice česká I, Hláskosloví*. 2., upr. vyd. Praha: SPN.
- Kosek, P. (2014). *Historická mluvnice češtiny – překlenovací seminář*. Brno: Masarykova univerzita.
- Kyas, V. (1997). *Česká bible v dějinách národního písemnictví*. Praha – Řím: Vyšehrad – Křesťanská akademie.
- Muehlberger, G. – Seaward, L. – Terras, M. – Ares Oliveira, S. – Bosch, V. – Bryan, M. – Colutto, S. – Déjean, H. – Diem, M. – Fiel, S. – Gatos, B. – Greinoecker, A. – Grüning, T. – Hackl, G. – Haukko-vaara, V. – Heyer, G. – Hirvonen, L. – Hodel, T. – Jokinen, M. – Kahle, P. – Kallio, M. – Kaplan, F. – Kleber, F. – Labahn, R. – Lang, E.M. – Laube, S. – Leifert, G. – Louloudis, G. – McNicholl, R. – Meunier, J.-L. – Michael, J. – Mühlbauer, E. – Philipp, N. – Pratikakis, I. – Puigcerver Pérez, J. – Putz, H. – Retsinas, G. – Romero, V. – Sablatnig, R. – Sánchez, J.A. – Schofield, P. – Sfikas, G. – Sieber, C. – Stamatopoulos, N. – Strauß, T. – Terbul, T. – Toselli, A.H. – Ulreich, B. – Villegas, M. – Vidal, E. – Walcher, J. – Weidemann, M. – Wurster, H. – Zagoris, K. (2019). Transforming scholarship in the archives through handwritten text recognition: Transkribus as a case study [online], *Journal of Documentation*, 75(5), 954–976, [cit. 20. 3. 2024]. Dostupné z: <https://doi.org/10.1108/JD-07-2018-0114>.
- Nedvěďová, M. *Kvantita samohlásek ve starší češtině*. Interní dokument ÚJČ AV ČR.
- Pisaro, M. J. (2002). *Quantity in Czech: A Dialectal and Historical Analysis*. Disertační práce. Illinois: Chicago, The Faculty of the Division of the Humanities.
- Romein, C. A. – Hodel, T. – Gordijn, F. – Zundert, J. J. van – Chagué, A. – Lange, M. van – Jensen, H. S. – Stauder, A. – Purcell, J. – Terras, M. M. – Heuvel, P. van den – Keijzer, C. – Rabus, A. – Sitaram, C. – Bhatia, A. – Depuydt, K. – Afolabi-Adeolu, M. A. – Anikina, A. – Bastianello, E. – ... – Zweistra, R. (2024). Exploring Data Provenance in Handwritten Text Recognition Infrastructure: Sharing and Reusing Ground Truth Data, Referencing Models, and Acknowledging Contributions. Starting the Conversation on How We Could Get It Done. *Journal of Data Mining and Digital Humanities*, [cit. 20. 3. 2024]. Dostupné z: <https://doi.org/10.5281/zenodo.10804745>.
- Synková, P. (2017). *Popis staročeské apelatvní deklinace (se zřetelem k automatické morfologické analýze textů Staročeské textové banky)*. Praha: Filozofická fakulta Univerzity Karlovy.
- Vážený, V. (1963). *Historická mluvnice česká II: Tvarosloví, I. část, Skloňování*. Praha: Státní pedagogické nakladatelství.
- Vintr, J. (2005). *Das Tschechische. Hauptzüge seiner Sprachstruktur in Gegenwart und Geschichte*. München: Otto Sagner.

**Piš, jak slyš: analýza automatických a manuálních přepisů nahrávek  
nerodilých mluvčích češtiny**

Michal Novák, Peter Polák, Kateřina Rysová, Magdaléna Rysová, Ondřej Bojar  
(Univerzita Karlova, Matematicko-fyzikální fakulta)

V příspěvku se věnujeme hodnocení efektivity systémů automatického rozpoznávání řeči (ASR) pro účely přepisu mluvených částí zkoušek z českého jazyka pro nerodilé mluvčí. Zkoumáme potenciál technologie ASR jako klíčového prvku v následném procesu vývoje automatizovaného nástroje pro hodnocení mluveného projevu cizinců. Zaměřujeme se na porovnání různých metodologií přepisu mluveného slova do psaného textu – zkoumáme výhody a nevýhody přímého ručního přepisu i ASR systémů. Cílem je posoudit, jak efektivní jsou tyto systémy ve srovnání s lidským přepisem a jaký rozsah manuálních oprav je nutný pro dosažení přesného přepisu.

V oblasti ASR se zaměřením na češtinu již byl vytvořen automatický transkripční systém orientovaný přímo na nerodilé mluvčí [2]. Tento model dosáhl na testovacích datech, jimiž byly nahrávky izolovaných slov a slovních taktů vyslovených nerodilými mluvčími, úspěšnosti 77 % v transkripci hlásek. V rámci daného projektu byl vyvinut také anotační nástroj Anophone [3]. Tato aplikace slouží jednak jako databáze nahrávek nerodilých mluvčích češtiny a zároveň umožňuje manuální anotaci těchto nahrávek. Tyto nástroje, spolu s dalšími ASR aplikacemi orientovanými na češtinu rodilých mluvčích, jako je například UWebASR [7], představují kontext, do kterého náš výzkum vstupuje.

Našimi daty jsou autentické nahrávky mluvených projevů nerodilých mluvčích, které byly pořízeny v rámci jazykových zkoušek pro cizince organizovaných Ústavem jazykové a odborné přípravy Univerzity Karlovy (ujop.cuni.cz) a Národním pedagogickým institutem ČR (npi.cz). Bezmála 300 minut nahrávek obsahuje mluvené projevy 39 nerodilých a několika rodilých mluvčích (v roli zkoušejících).

Nahrávky jsme nechali ručně přepsat rodilými mluvčími podle předem daných pravidel s důrazem na to, aby přepisy zachycovaly co nejautentičtější, co mluvčí na nahrávce skutečně řekli (anotátoři tak například zachycovali mj. přeřeky, opakování slov nebo slova nedořečená). Tyto referenční přepisy slouží jako základ pro srovnání s výstupy z existujících ASR systémů [5, 6]. Zaměřujeme se na kvalitativní srovnání manuálních přepisů, automatických přepisů a přepisů vzniklých manuální post-editací automatických přepisů s cílem identifikovat specifické aspekty mluveného projevu, které současné ASR systémy nezachycují nebo záměrně v rámci snahy o dobrý přepis eliminují [1, 4]. V rámci kvantitativní analýzy dále zjišťujeme, které ASR systémy poskytují výstupy, jež jsou pro post-editaci nejvhodnější. Věnujeme se také otázce, zda je manuálně post-editovaný přepis zatížený zkreslením vůči ASR systému, na jehož základě byl vytvořen. Výsledky těchto analýz nám umožní lépe porozumět limitům stávajících ASR systémů a poskytnou cenné informace pro následný vývoj našeho hodnotícího nástroje.

Výzkum prezentovaný v příspěvku byl financován z prostředků Programu na podporu aplikovaného výzkumu v oblasti národní a kulturní identity na léta 2023 až 2030 (NAKI III) Ministerstva kultury ČR v rámci projektu Automatické hodnocení mluveného projevu v češtině (DH23P03OVV037).

## Bibliografie

- [1] Al-Ghezi, R., Getman, Y., Rouhe, A., Hildén, R., Kurimo, M. Self-Supervised End-to-End ASR for Low Resource L2 Swedish. *Proc. Interspeech 2021*, 1429–1433, 2021.
- [2] Holaj, R. *Nástroj pro automatický přepis řeči nerodilých mluvčí českého jazyka*. Disertační práce, Masarykova univerzita, Filozofická fakulta, 2023. URL <https://is.muni.cz/th/io67a/>
- [3] Holaj, R., Pořízka, P. ANOPHONE: An Annotation Tool for Phonemes and L2 Annotation Systems for Czech. *Journal of Linguistics/Jazykovedný časopis*, 74(1): 333–344, 2023.
- [4] Park, S., Culnan, J. A comparison between native and non-native speech for automatic speech recognition. *The Journal of the Acoustical Society of America*, 145(3\_Supplement), 1827–1827, 2019.
- [5] Pratap, V. et al. Scaling speech technology to 1,000+ languages. arXiv preprint arXiv:2305.13516, 2023.
- [6] Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, 28492–28518, 2023. PMLR.
- [7] Švec, J., Bulín, M., Pražák, A., Ircing, P. UWebASR – Web-based ASR engine for Czech and Slovak. In *Proceedings of CLARIN Annual Conference 2018*, 190–193, Pisa, Italy, 2018. CLARIN.

## Vývoj českého přechodníku přítomného ve srovnání s jeho chorvatským a slovinským protějškem neboli byl český přechodník předurčen k zániku?

Aleksandra Wójcik (Jagellonská univerzita)

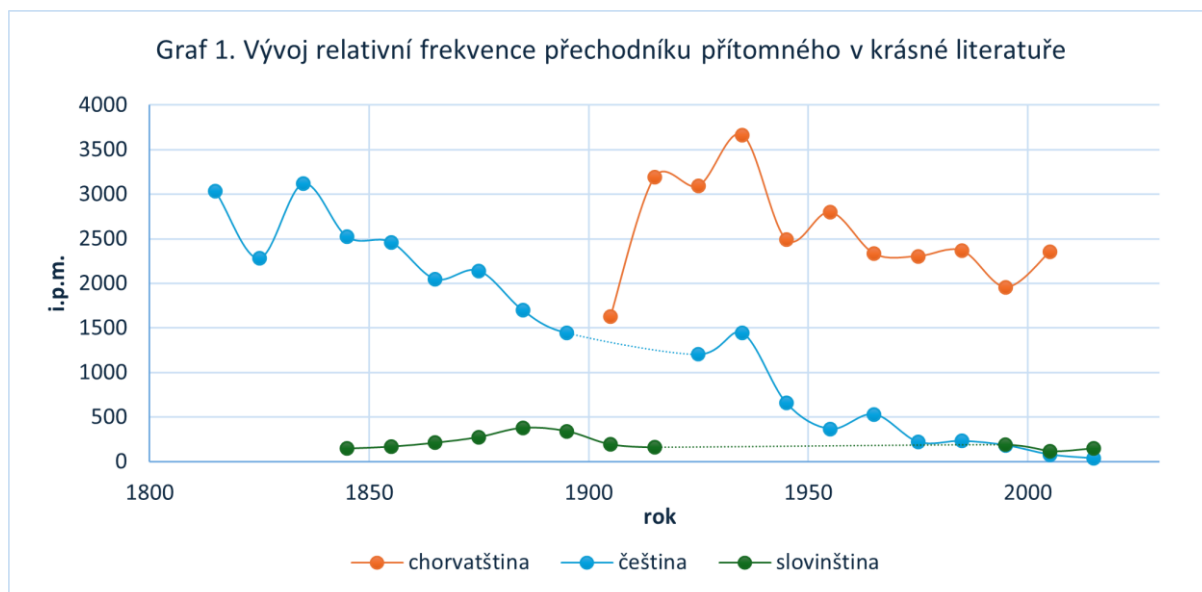
Příspěvek se zaměřuje na změny ve frekvenci českého přechodníku přítomného a jeho chorvatského a slovinského protějšku (chorv. *glagolski prilog sadašnji*, slovin. *deležje na -č, -aje, -e*) v krásné literatuře v období 1810–2019. Výzkum se opírá o metody korpusové lingvistiky a vychází z manuálně zpracovaných dat z korpusů: Archiv 19. století a SYN v9 pro češtinu (v rozmezí let 1810–2019), Riznica pro chorvatštinu (období 1900–2009), IMP (Erjavec 2014) a Gigafida 2.0 pro slovinštinu (v rozmezí let 1840–2018); časová a stylová omezení jsou dána dostupností textů. Cílem příspěvku je porovnat vývojové tendence přechodníku přítomného ve třech zmíněných jazycích a poukázat na příčiny rozdílů mezi nimi. Zvláštní pozornost je věnována českému přechodníku, jehož mimořádně rychlý ústup nemá obdoby ani v chorvatštině, ani ve slovinštině.

Poster navazuje na dosavadní diachronní výzkum frekvence českých přechodníků, jednak na klasické monografie Emila Dvořáka (1970, 1983), jednak na nejnovější práce Olgy Nádvorníkové (2021, v tisku) a Aleksandry Wójcik (2023). Frekvence přechodníků v chorvatštině nebyla zatím zkoumána ani synchronně, natož diachronně. Naopak ve slovinštině se ve středu pozornosti nachází hlavně vývoj přechodníků v 19. století (srov. Jesenšek 1996, 1998), zatímco chybí údaje pro 20. a 21. století; v případě dvou článků (Mikolič Južnič 2012, 2013), jež se pokoušejí zachytit situaci po roce 1990, je třeba konstatovat, že s ohledem na určité metodické nepřesnosti neposkytují spolehlivá data. Tímto vyvstává potřeba doplnit dosavadní poznatky s použitím jednotné metodologie, což by umožnilo přímé srovnání průběhu změn ve frekvenci českých, chorvatských a slovinských přechodníků (viz graf 1).

Kontrastivní přístup se totiž ukazuje jako vhodný způsob odhalení činitelů ovlivňujících frekvenci těchto slovesných tvarů. Je zřejmé, že podobnosti a rozdíly ve frekvenci přechodníků nesouvisí s mírou příbuznosti jazyků: chorvatský přechodník přítomný je v krásné literatuře dvacetkrát častější než jeho slovinský protějšek (2355,35 i.p.m. vs. 117,32 i.p.m. v desetiletí 2000–2009), přestože se jedná o dva velmi blízké jazyky, zatímco frekvence přechodníku přítomného v češtině, jež se ještě na přelomu 19. a 20. století blížila frekvenci v chorvatštině, je současně na čtyřikrát nižší úrovni než ve slovinštině (35,50 i.p.m. vs. 149,84 i.p.m. v desetiletí 2010–2019). Zdá se tedy, že míru (ne)používání těchto slovesných tvarů nelze pokládat za imanentní danost, nýbrž za vlastnost, již podstatně ovlivňují úsudky a představy uživatelů jazyka o postavení přechodníků, jejich stylovém příznaku apod. Nasvědčují tomu rozdíly ve frekvenci přechodníků mezi funkčními styly: české a slovinské přechodníky se objevují zásadně častěji v krásné literatuře než v publicistice, zatímco v chorvatštině – vzhledem k nepřítomnosti silného knižního příznaku přechodníku přítomného – je tento stylový rozdíl výrazně menší.

Na frekvenci přechodníků má vliv také jazyková politika: kodifikace (srov. Karlík 2022), způsob výkladu o přechodnících ve středních školách a na vysokých školách, doporučení jazykových autorit, a dokonce přímé zásahy jazykových redaktorů do textů (srov. dotazníkové setření mezi překladateli a redaktory z knihy Nádvorníkové, 2023). Význam tohoto faktoru dokládají zejména dva protichůdné příklady: pokus o umělé oživení slovinských přechodníků ve druhé polovině 19. století (srov. Jesenšek 1998) a mimořádně rychlý ústup českých přechodníků po druhé světové válce, který lze spojit s dřívějšími odmítavými posudky Jiřího Hallera, Václava Ertla a Josefa Zubatého (srov. Wójcik 2023). Takové příklady potvrzují fakt, že je třeba na vývoj přechodníků nahlížet nejen podle morfologických kritérií, nýbrž především z hlediska vztahů jazyka a společnosti.





## Bibliografie

- Archiv 19. století: nerevidovaný interní diachronní korpus Ústavu Českého národního korpusu FF UK. Cit. 29. 2. 2024. <<http://www.korpus.cz>>.
- DVOŘÁK, Emil (1970): *Vývoj přechodníkových konstrukcí ve starší češtině*. Praha.
- DVOŘÁK, Emil (1983): *Přechodníkové konstrukce v nové češtině*. Praha.
- ERJAVEC, Tomaž (2014): *Digital library and corpus of historical Slovene IMP 1.1*. Slovenian language resource repository CLARIN.SI. Cit. 29. 2. 2024. <<http://hdl.handle.net/11356/1031>>.
- Gigafida 2.0: Corpus of Written Standard Slovene. Cit. 29. 2. 2024. <<https://viri.cjvt.si/gigafida>>.
- JESEŇSEK, Marko (1996): Poskus periodizacije razvoja oblik na -č in -ši v slovenskem knjižnem jeziku 19. stoletja. *Jezik in slovstvo* 61, s. 171–189.
- JESEŇSEK, Marko (1998): *Deležniki in deležja na -č in -ši. Razširjenost oblik v slovenskem knjižnem jeziku 19. stoletja*. Maribor.
- KARLÍK, Petr (2022): Struktura prezentních participií v češtině. *Korpus – gramatika – axiologie* 26, s. 39–55.
- MIKOLIČ JUŽNIČ, Tamara (2012): Korpusna analiza slovenskega deležja v različnih besedilnih tipih. In: Tomaž Erjavec – Jerneja Žganec Gros (eds.), *Zbornik 8. konference Jezikovne tehnologije, 8. do 12. oktober 2012*. Ljubljana, s. 147–152.
- MIKOLIČ JUŽNIČ, Tamara (2013): Neosebne glagolske oblike v prevodni in izvorni slovenščini: primer deležja. In: Špela Vintar (eds.), *Slovenski prevodi skozi korpusno prizmo*. Ljubljana, s. 70–93.
- NÁDVORNÍKOVÁ, Olga (2021): Stylistic normalisation, convergence and cross-linguistic interference in translation: The case of the Czech transgressives. In: Mario Bisiada (eds.), *Empirical studies in translation and discourse*. Berlin, s. 53–91.
- NÁDVORNÍKOVÁ, Olga (2023): *Český přechodník jako konverbum: Korpusová analýza překladových a nepřekladových textů*. Praha.
- NÁDVORNÍKOVÁ, Olga (v tisku): The Czech converb confronted with its French and Polish counterparts: Investigation of diachronic factors shaping the properties of converbs. In: Paola Coticelli – Eystein Dahl – Jelena Zivojinovic (eds.), *Diachronic, Typological, and Areal Aspects of Converbs*. Berlin.
- Riznica: Croatian language corpus Riznica 0.1 (2018). Institut za hrvatski jezik i jezikoslovlje, Slovenian language resource repository CLARIN.SI, Cit. 29. 2. 2024. <<http://hdl.handle.net/11356/1180>>.
- SYN v9: Korpus SYN [verze 9 z 5. 12. 2021] (2021). Praha: Ústav Českého národního korpusu FF UK. Cit. 29. 2. 2024. <<https://www.korpus.cz>>.

WÓJCIK, Aleksandra (2023): The decline of Czech transgressives – a *memento mori* for the Polish language?. *Polonica* 43, s. 193–206.