



# Outline

- Preliminary remarks
- Existing literature on the topic and research gap
- TransVar – the corpus for study language variation and change in the Transcarpathian region
- Discovering language polymorphism through corpus material
- Ukrainian comparative material
- Comparative analysis between south-western Ukrainian dialects of the beginning of the twentieth century
- Discussion and Conclusion

# Outline

- Preliminary remarks
- Existing literature on the topic and research gap
- TransVar – the corpus for study language variation and change in the Transcarpathian region
- Discovering language polymorphism through corpus material
- Ukrainian comparative material
- Comparative analysis between south-western Ukrainian dialects of the beginning of the twentieth century
- Discussion and Conclusion

## Essential terms

- **Signal** – “anything that indicates meaning, whether it is based on phonological units or not” (Diver, 2012b: 490)
- **Meaning** – “the relatively small number of communicative units actually provided by an individual language, these being indicated by specific signals of that language” (Diver, 1975 [2012]: 48)
- **Distributional skewing** – irregularity that correlates with “certain characteristics of the messages that were communicated nearby” (Diver, 2012a: 298)

# Essential terms

- **Satellite relationship** – the relationship between a lexical morpheme and a grammatical morpheme, “in which the lexical morpheme stands as “the sun” and the grammatical morpheme as “the planet”” (Diver, 2012b: 484)
- **Attractor – occurrence** (“verb”) or **entity** (“noun”), the lexical morpheme in a satellite relationship, the term refers to the Great Attractor (Dressler, 1987)
- **Satellite** – the grammatical morpheme in a satellite relationship (Diver, 2012b: 484)
- **Satellite Clusters** (A and B): “The members of the one are used to create one kind of effect in the message, with the consequence that the lexical item is then classified as a “noun”; the members of the other, an effect that leads to classification as a “verb” (Diver, 2012b: 502)

# Essential terms

- **Grammatical system** – a division of semantic substance “into some number of parts, the totality of which exhaustively covers the entire domain of the substance” (Diver, 1975 [2012]: 61)

# Goal I: language polymorphism markers

- **Language polymorphism** – a result of on-going variation and/or change processes within a population that lead to appearance of significant differences in individual variants of speakers
- **Language polymorphism marker** – a difference in the distributional skewings of signals that indicate such processes

**RQ 1: How to find and consistently show the presence of language polymorphism markers in unstructured data?**

## Goal II: quantitative aposterioristic grammar

- **Grammar** – a description of all the grammatical systems of a certain variety or a group of varieties (with all the possible notes on the approximations, see Wiese (2023))
- **Aposterioristic** grammar – a grammar that derives the grammatical systems from the distributional skewings, witnessed in the empirical data
- **Quantitative** grammar – a grammar that relies on the corpus-driven (Divjak et al. 2017: 4) methodology, rather than corpus-based or corpus-assisted

**RQ 2: How feasible is it currently to build a quantitative aposterioristic grammar?**

# Outline

- Preliminary remarks
- Existing literature on the topic and research gap
- TransVar – the corpus for study language variation and change in the Transcarpathian region
- Discovering language polymorphism through corpus material
- Ukrainian comparative material
- Comparative analysis between south-western Ukrainian dialects of the beginning of the twentieth century
- Discussion and Conclusion

# Language polymorphism

- Language change: a bread and butter of historical linguistics (Campbell 2013)
- Language variation: a bread and butter of sociolinguistics (Labov 1966)
- Language variation and change and textual data
  - Grammatical Error Detection and Correction (GED/C) for language variation (Li et al. 2024)
  - Textual tradition and language variation and change
  - Pre-modern varieties and the notion of error (Bounatirou 2018; 2021)

# South-western Ukrainian dialectology

- Long tradition of study (Mikhal'chuk 1877; Mikhal'chuk and Tymchenko 1910; Hantsov 1923; Zilyns'kyi 1925, 1932, 1933; Kuraszkiewicz 1963; Dezső 1967; Karaś 1975; Shevelov 1979; Latta 1991; Xudaš, 1991; Myholynec' 2004; Del Gaudio 2017)
- Multiple material collections (Hnatjuk, 1904, 1912; Onyščuk, 1909; Nakonečna and Rudnyc'kyj 1940; Kuraszkiewicz 1963; Zheguc 2001; Myholynec' 2004) **not digitised**
- Rising standardization attempts (Magocsi 1996; Fontański and Chomiak 2000) and their digitization (Rabus and Scherrer 2017; 2025)

# Reference grammar

- Rarely a methodological introspection, heavy reliance on pre-established categories (Terhart 2024: 36–37)
- A very formulaic structure
  - Phonetics
  - Morphology
  - Syntax

# *A priori* and *a posteriori* in linguistics

- Slavic languages and corpus linguistics (Divjak et al., 2017)
- The discussion on *tertium comparationis* in typology (Haspelmath 2018: 97–101)
- Integrative vs. differential approach in dialectology (Kriuchkova 2007: 31, Goldin & Kryuchkova 2011; Hromko 2020)
- Columbia School of Linguistics: radical functionalism (Otheguy 2002; Diver 2012a, b)

# *A posteriori versus ex numeris*

- The notion of ideal type (Weber 1904)
- Tasks
  - Perplexity scoring (Jelinek et al. 1977)
  - Universal tagging (de Marneffe et al. 2021)
- Rule-based models: scholarly bias
- Large Language Models: inherent bias (Terry & de Lhoneux 2024; Guo et al. 2025)

Issue in computer-based, **not computer-driven** study

**Then how it is possible?**

**(Pause)**

# Outline

- Preliminary remarks
- Existing literature on the topic and research gap
- TransVar – the corpus for study language variation and change in the Transcarpathian region
- Discovering language polymorphism through corpus material
- Ukrainian comparative material
- Comparative analysis between south-western Ukrainian dialects of the beginning of the twentieth century
- Discussion and Conclusion

# Concept

- Morphosyntactically tagged corpus: CoNLL-U format
- Open-access
- Metatagging, where possible
- Additional layers:
  - Named entities
  - Basic vocabulary
  - Variation



TransVar as of 20.05.2026

# Materials

- Monographic-style description: Nakonečna and Rudnyc'kyj 1940
- Collections: Kuraszkiewicz 1963; Zheguc 2001; Myholynec' 2004
- Folklore tales: Hnatjuk, 1904, 1912; Onyščuk, 1909

# Tagging

- Universal Dependencies format (de Marneffe et al. 2021)
  - UPOS (Part-of-speech)
  - Feats (Morphological categories)
  - Lemmata (according to existing vocabularies)
  - Dependency parsing (syntactic relations)
- Variation and change
  - Manual
  - UA-GEC (Syvokon et al. 2023)

# Model

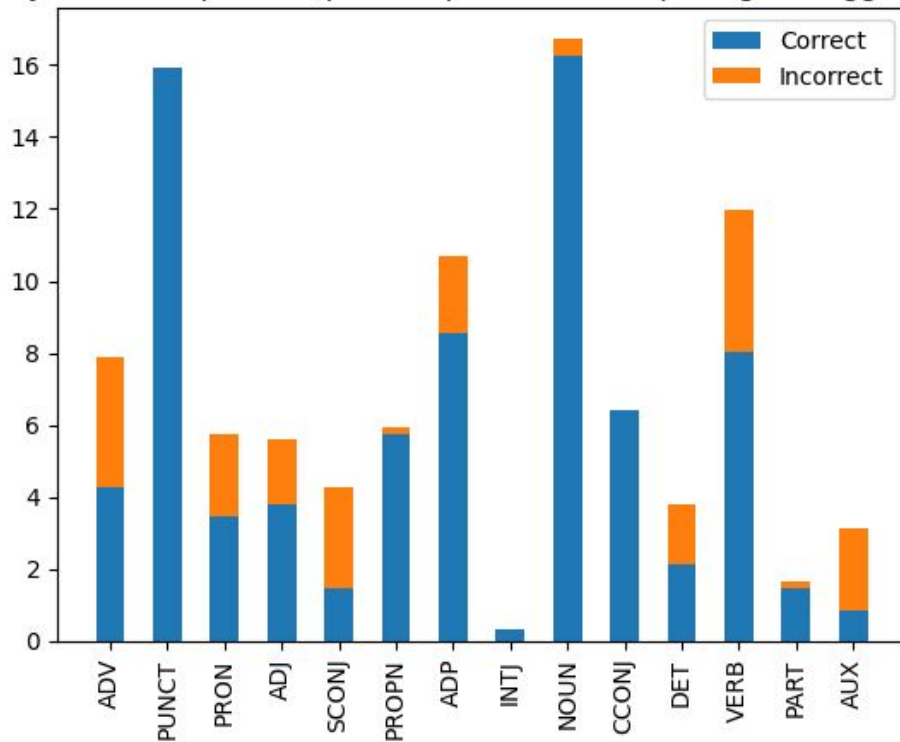
- Stanza (Qi et al. 2020)
- stanzatagger – an alternative way for morphological tagging (Scherrer 2021), trained on ParlaMint (Kopp et al. 2023)



Afanasev 2026

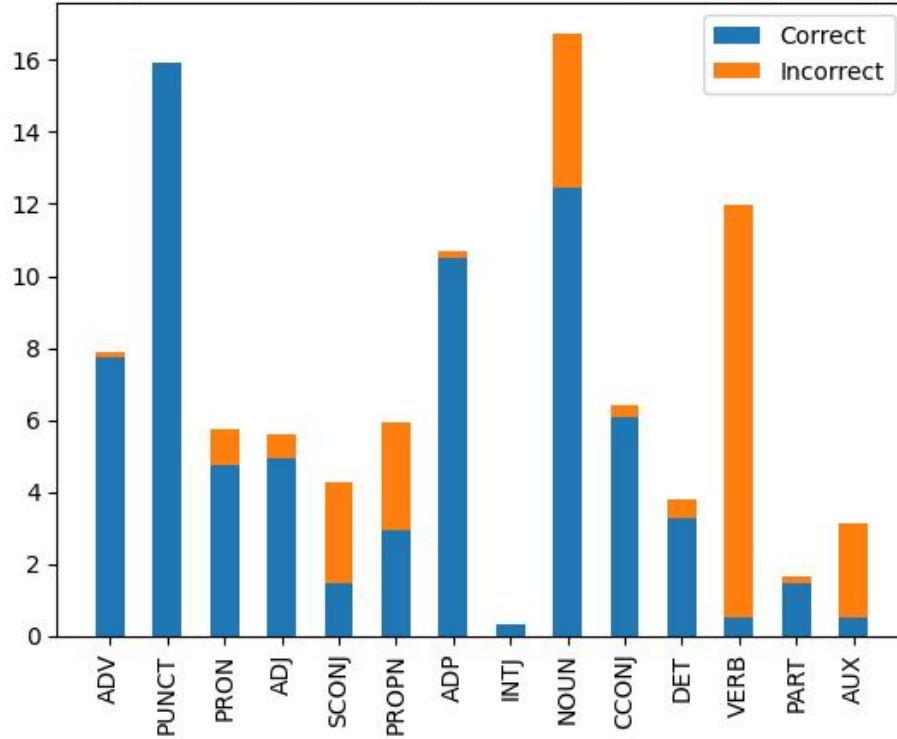
# The issue

Accuracy score comparison (part-of-speech and morphological tagging, Stanza)



# The issue

Accuracy score comparison (lemmatisation, Stanza, gold tagging)







## Is there a correlation?

- Visibly, it seems so
- Spearman's coefficient, p-value < 0.001
- Pearson's coefficient, p-value < 0.001

# Who is in charge?

Lemko (Nakonečna and Rudnyc'kyj, 1940,  
p. 31)

To		сут
<b>T-o</b>		<b>sut</b>
<b>DET-NEUT.NOM.SG</b>		<b>be.PRES.3PL</b>
вші́тки		лемкі́вски
všýtk-y		lemkývsk-y
all-NOM.PL		Lemko-NOM.PL
сéла	, де	по лемкі́вски
<b>sél-a</b>	, de	po lemkývsky
<b>village-NOM.PL</b>	, where	in Lemko
гва́рят	.	
fivár <sup>j</sup> -at	.	
speak-PRES.3PL	.	

'**These are** all Lemko **villages**, where one speaks Lemko.'

# Hypothesis

- Deep-level language polymorphism  
=> discrepancies in distributional skewings between TransVar and UD\_Ukrainian-IU  
=> multiple errors on all of the levels of morphosyntactic tagging
- A marker is a copular existential structure (Moro 2017)

# Outline

- Preliminary remarks
- Existing literature on the topic and research gap
- TransVar – the corpus for study language variation and change in the Transcarpathian region
- **Discovering language polymorphism through corpus material**
- Ukrainian comparative material
- Comparative analysis between south-western Ukrainian dialects of the beginning of the twentieth century
- Discussion and Conclusion

# Method

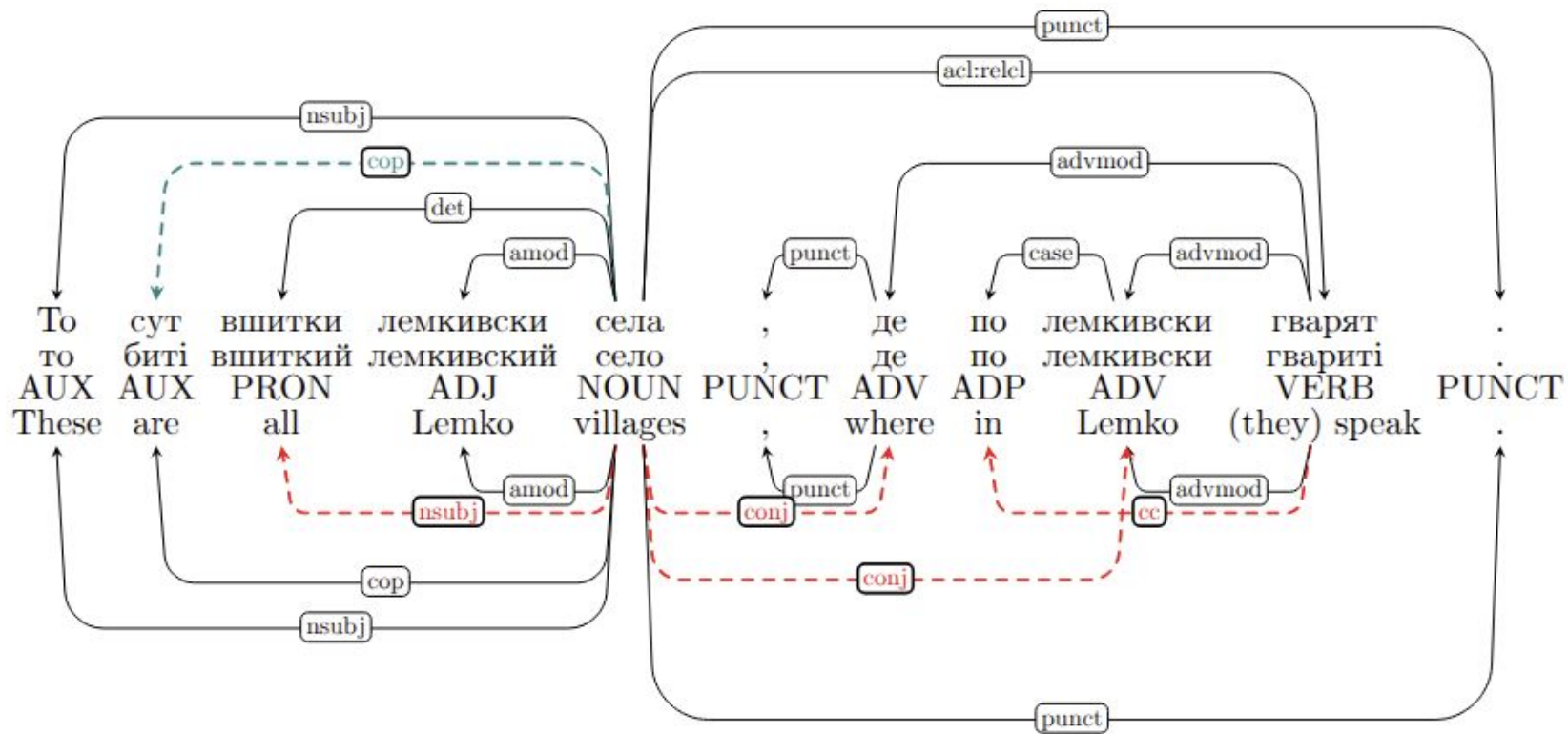
- Borrowed from biology (Brusatte et al. 2010)
- Multidimensional scaling of the lists of morphological features of individual specimens to trace macro-scale patterns
- Specimens: instances of *bymu* 'to be'

## Why not use qualitative analysis?

- Does not facilitate a new look on the data
- Less effective in highlighting trends (still easy to miss something)
- Does not enable a variationist perspective as much

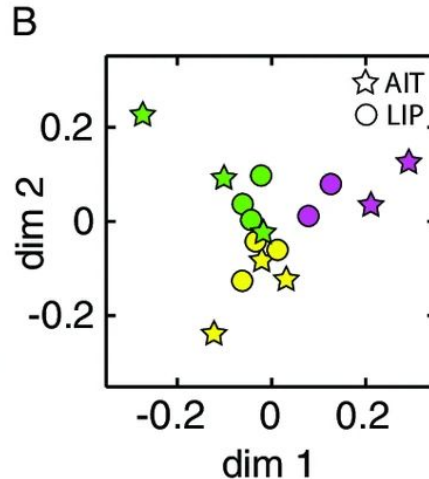
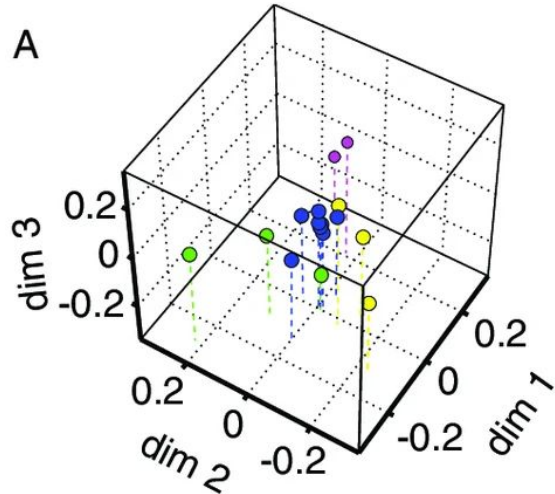
# Data gathering

- Collect all the instances of *бѣ* 'to be'
- Collect their heads and immediate dependencies
- Collect all the UD morphological features (parts-of-speech and morphological categories)
  - For the instances of *бѣ* 'to be'
  - For their heads and immediate dependencies



# Multidimensional scaling (MDS)

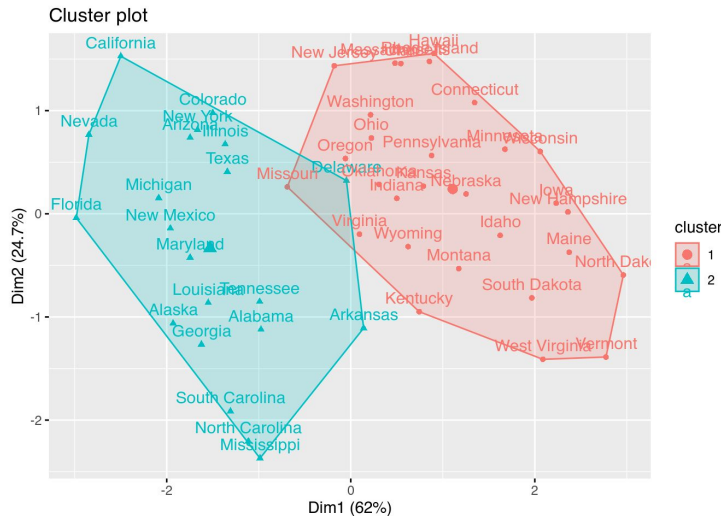
- Shortens a number of dimensions in data
- Applied to dialect differences (Mansfield et al. 2023)
- Useful for both humans and machines



(Lehky and Sereno,  
2007)

# K-means clusterisation

- The most basic clustering mechanism (R Core Team 2026)
- Splits data into k clusters by grouping of k randomly chosen points



# Optimal cluster number

- Frequently defined based on a preliminary analysis
- Otherwise, a handful of automatic methods (Kassambara and Mundt 2026)
  - Elbow
  - Silhouette
  - Gap statistics

# Principal Component Analysis (PCA)

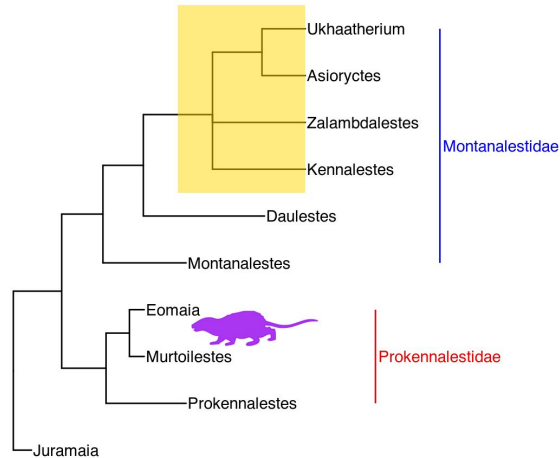
- Operates on a logic, similar to MDS
- Preserves key dimensions in the data (the ones that explain the most of the variation)
- Facilitates better explanatory analysis

# Probabilistic clustering

- R package pvclust (Suzuki and Terada 2019)
- Hierarchical ascending clustering
- Assesses probability that a certain cluster indeed exists

# Visualisation of the results

- Basic R plot functions (R Core Team 2026)
- Visualisation of trees with ggtree (Yu et al. 2017)



# Qualitative analysis

- Do instances of small territorial lects go together?
- If so, what are the groupings within them?

# Outline

- Preliminary remarks
- Existing literature on the topic and research gap
- TransVar – the corpus for study language variation and change in the Transcarpathian region
- Discovering language polymorphism through corpus material
- **Ukrainian comparative material**
- Comparative analysis between south-western Ukrainian dialects of the beginning of the twentieth century
- Discussion and Conclusion

## Available Ukrainian corpora

- GRAC (Shvedova et al.)
- UD\_Ukrainian-IU
- UD\_ParlaMint (Kopp et al. 2023)

## Why ParlaMint?

- GRAC is not represented in UD
- Results of Stanza hint that behaviour of *бѣму* 'to be' in IU heavily differs between TransVar and this corpus

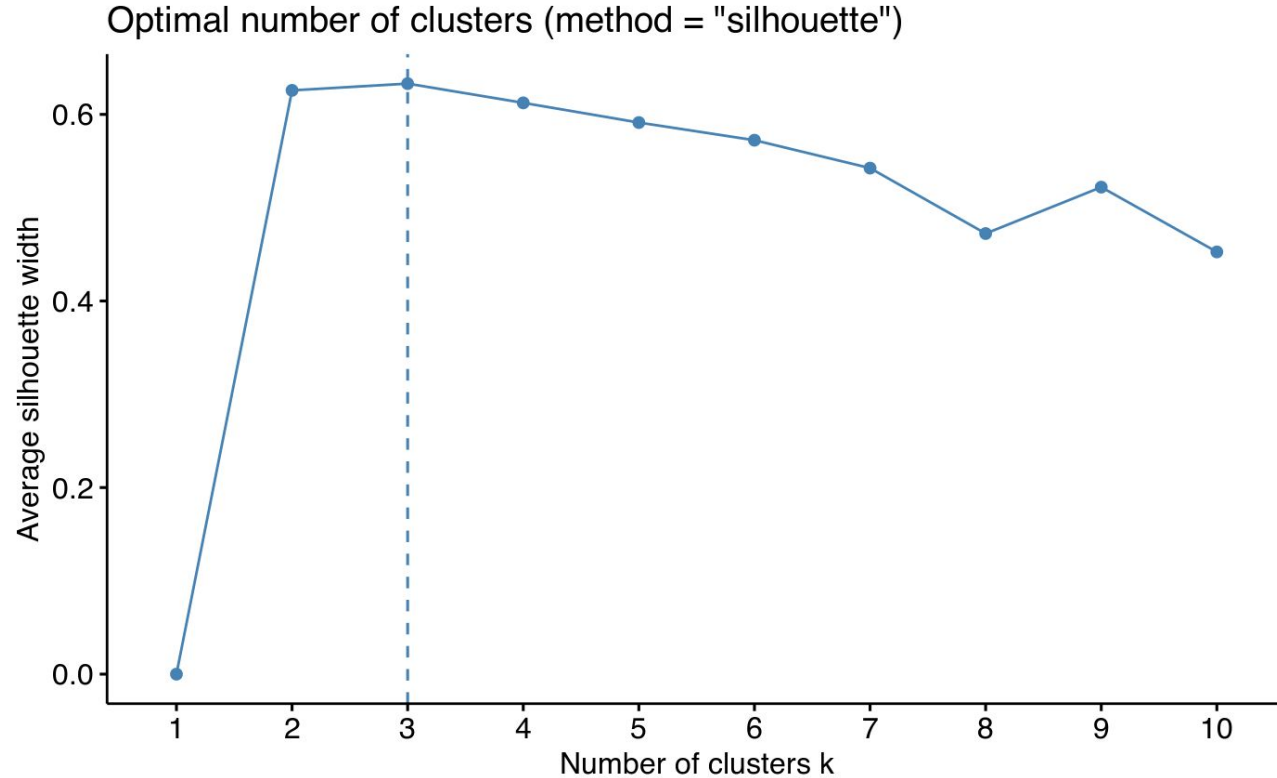
## Possible issues

- Time discrepancies
- Tagging discrepancies
- Genre discrepancies
- Size discrepancies

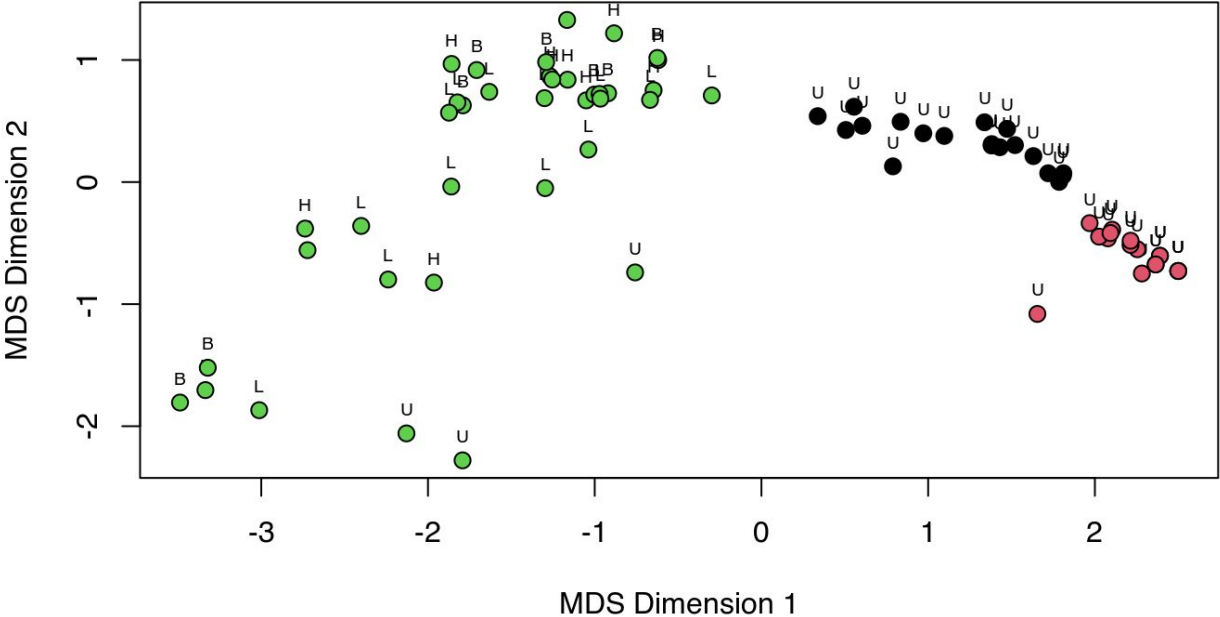
# Outline

- Preliminary remarks
- Existing literature on the topic and research gap
- TransVar – the corpus for study language variation and change in the Transcarpathian region
- Discovering language polymorphism through corpus material
- Ukrainian comparative material
- Comparative analysis between south-western Ukrainian dialects of the beginning of the twentieth century
- Discussion and Conclusion

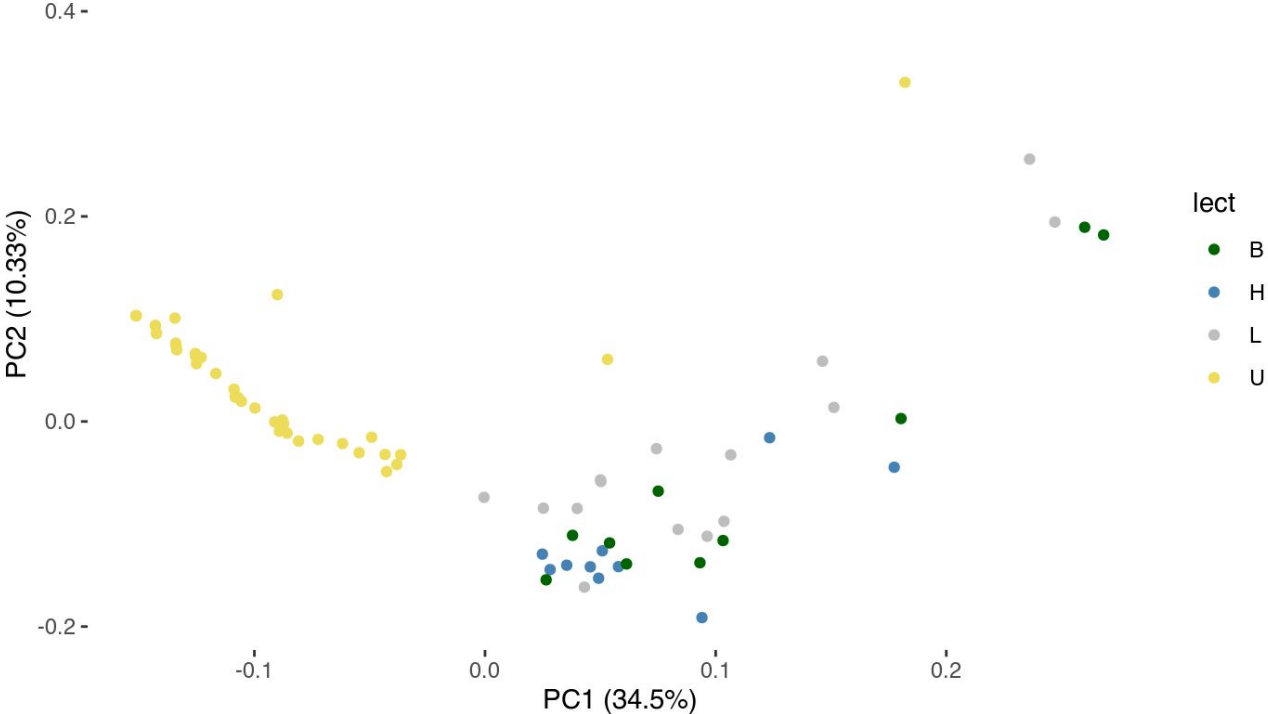
# Silhouette



# MDS + K-means



# PCA





# Qualitative analysis: clusters

Modern Standard Ukrainian (Kopp et al., 2023)

(...)	де	є
(...)	<b>de</b>	<b>je</b>
(...)	<b>where</b>	<b>be.PRES. 3SG</b>

правовий	нігілізм	i
prɛw-ɔw-ij	nɪgɪlɪzm-∅	i
law-ADJ-MASC.NOM.PL	<b>nihilism-NOM. SG</b>	and

хамство	.
<b>hámstv-o</b>	.
<b>rudeness-NOM. SG</b>	.

‘(...) **Where there is judicial nihilism and rudeness**’

Hutsul (Nakonečna and Rudnyc’kyj, 1940, p. 75)

(...)	бо	то
(...)	bɔ	<b>t-ó</b>
(...)	because	<b>DET-NEUT.NOM.SG</b>

волóки	не	бу́ли
vɔlókʲ-ĭ	ně	<b>bú-l-ĭ</b>
<b>ribbon-NOM.PL</b>	not	<b>be-PST-3PL</b>

клочені	(...)
<b>klotʲ-ĕnʲ-ĭ</b>	(...)
<b>flax-PTCP.PST.PASS-3PL</b>	(...)

‘(...) because **that the ribbons were not flax** (...)’

# Qualitative analysis: insights from GRAC

- The construction is present
- Auxiliary pronoun differs in its phonetic form: *mo* vs. *ɥe*
- Correlation between lects with *mo* and lects with *ɥe*  
*mo cym* vs. *ɥe ε* ‘these are’

# Outline

- Preliminary remarks
- Existing literature on the topic and research gap
- TransVar – the corpus for study language variation and change in the Transcarpathian region
- Discovering language polymorphism through corpus material
- Ukrainian comparative material
- Comparative analysis between south-western Ukrainian dialects of the beginning of the twentieth century
- Discussion and Conclusion

# Discussion

- Differences in the behaviour of 'to be' between these groups are substantial
- Within each dataset the behaviour of 'to be' is more consistent.
- Semantic similarities vs. crucial morphosyntactic differences

# Limitations

- Neither Transcarpathian lects, nor Modern Standard Ukrainian well-represented
- Subcorpora reflect idiolects

# TODO

- Perform UD tagging of GRAC and use GRAC constructions for comparison
- Study more closely the differences between Modern Standard Ukrainian and Transcarpathian
- Perform an ablation study and/or permutation test
- Score a Jaccard over tagsets and/or use a control verb

# Conclusion

- Responsible pattern replication from biology may provide with additional insights
- The combination of morphosyntactic tagging and clustering techniques facilitates a fully quantitative description of language polymorphism
- The creation of a quantitative aposterioristic grammar is feasible, yet currently achieving it is very hard

**Thank you!**



Project site

# References I (Ukrainian)

- Dezsb 1967 – Л. Деже. Очерки по истории закарпатских говоров. Будапешт: Akadémia Kiadó.
- Henryk Fontański and Mirosława Chomiak. 2000. Gramatyka jezyka lemковского. Śląsk, Katowice.
- Salvatore Del Gaudio. 2017. An introduction to Ukrainian dialectology. Wiener slavistischer Almanach. Linguistische Reihe Sonderband 94. Peter Lang, Frankfurt am Main Bern Wien.
- Hantsov, V. 'Dialektolohichna klasyfikatsiia ukrains'kykh hovoriv,' ZIFV, 4 (1923) (repr, Köln 1975)
- Hnatjuk, V. 1904. Znadobi do galic'ko-rus'koï demonolohii. Zibrav Volodimir Gnatuk [Materials for Galician-Ruthenian Demonology. Collected by Volodymyr Hnatjuk] // Etnografichnyi žbirnik / Vidaè Etnografichna komisiia Naukovogo tovaristva imeni Ševčenka. L'viv, 1904. Tom H. [3] + 272 s.
- Hnatjuk, V. 1912. Znadobi do ukrains'koï demonolohii. Zibrav Volodimir Gnatuk. Tom II. Vipusk 1 [Materials for Ukrainian Demonology. Collected by Volodymyr Hnatjuk. Vol. II. Issue 1] // Etnografichnyi žbirnik / Vidaè Etnografichna komisiia Naukovogo tovaristva imeni Ševčenka. L'viv, 1912. Tom HHH.
- Kuraszkiewicz, W. Zarys dialektologii wschodnio-slowiańskiej z wyborem tekstów gwarowych (2nd ed., Warsaw 1963)
- Kassambara A, Mundt F (2026). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 2.0.0. With contributions from Laszlo Erdey (Faculty of Economics and Business, University of Debrecen, Hungary), <https://CRAN.R-project.org/package=factoextra>.
- Karaš, M. (ed). Studia nad dialektologią ukraińską (Cracow 1975) [materials of I. Zilyns'kyi]
- Kopp, Matyáš; Kryvenko, Anna and Rii, Andriana, 2023, Ukrainian parliamentary corpus ParlaMint-UA 4.0.1, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1900>.
- Latta, Vasyl. Atlas ukrains'kykh hovoriv Skhidnoi Slovacchyny (Bratislava 1991)
- Magocsi, R. P. [A History of Ukraine](#). University of Toronto Press, 1996
- Ol'ha F. Myholynec'. 2004. Ukrains'ki zakarpats'ki hoviryky : teksty. Lira, Užhorod.
- Mikhal'chuk, K. Narechiiia, podnarechiiia i govory luzhnoi Rossii v sviazi s narechiiiami Galichiny. Trudy Etnografichesko-statisticheskoï ekspeditsii v Zapadno-Russkii kraï, 7 (Saint Petersburg 1877)
- Mykhal'chuk, K.; Tymchenko, Ie. Prohrama do zbyrannia dialektichnykh odmin ukrains'koï movy (Kyiv 1910)
- Hanna Nakonečna and Jaroslav Bohdan Rudny'kyj. 1940. Ukrainische Mundarten : Südkarpatoukrainisch ; (Lemkisch, Bojkisch und Huzulisch) [Ukrainian dialects: South Carpathian Ukrainian; Lemkian, Bojkian and Huzulian]. Arbeiten aus dem Institut für Lautforschung an der Universität Berlin ; 9. Otto Harrassowitz, Berlin.
- Onyščuk, Antin. 1909. Materiali do gucul's'koï demonolohii. Zapisav v Zelenici, Nadvirnians'kogo povita, 1907—1908 rr. — L'viv, 1909. — T. XI. — Č. 2. — S. 1—139. [Onyshchuk, Antin. Materials for Hutsul Demonology. Recorded in Zelenytsia, Nadvirna district, 1907–1908 [...] folk teacher. Materials for Ukrainian Ethnology. — Lviv, 1909. — Vol. XI. — Part 2. — pp. 1–139.]
- Achim Rabus and Yves Scherrer. 2025. [Dialects, Topic Models, and Border Effects: The Rusyn Case](#). In Proceedings of the 10th Workshop on Slavic Natural Language Processing (Slavic NLP 2025), pages 38–43, Vienna, Austria. Association for Computational Linguistics.
- Yves Scherrer and Achim Rabus. 2017. Multi-source morphosyntactic tagging for spoken Rusyn. In Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial), pages 84–92, Valencia, Spain. Association for Computational Linguistics.
- Shvedova et al. – Генеральний регіонально анотований корпус української мови (ГРАК) / М. Шведова, П. фон Вальденфельс, С. Яригін, А. Рісін, В. Старко, Т. Ніколаєнко та ін. – Київ, Львів, Єна, 2017-2022: uacorporus.org.
- Shevelov, G. A Historical Phonology of the Ukrainian Language (Heidelberg 1979)
- Xudaš 2001 – Худаш М., Демчук М. Походження українських карпатських і прикарпатських назв населених пунктів (відантропонімі утворення). Київ, 1991.
- Ivan Zheguc. 2001. Vybrani teksty z hucul's'koho hovoru v Zakarpatti [Selected texts from the Hutsul dialect in Transcarpathia]. I. Zheguc, Munich.
- Zilyns'kyi, I. 'Do pytannia pro dialektolohichnu klasyfikatsiiu ukrains'kykh hovoriv,' ZNTSh, 141-3 (1925)
- Zilyński, J. Opis fonetyczny języka ukraińskiego (Cracow 1932)
- Zilyns'kyi, I. Karta ukrains'kykh hovoriv z poiasnenniamy (Warsaw 1933)

# References II (TL & CL)

- Brusatte, Stephen L., Michael J. Benton, Graeme T. Lloyd, Marcello Ruta & Steve C. Wang. 2010. Macroevolutionary patterns in the evolutionary radiation of archosaurs (tetrapoda: diapsida). *Earth and Environmental Science Transactions of the Royal Society of Edinburgh* 101, 367–382. <https://doi.org/10.1017/S1755691011020056>.
- Bounatirou, Elias. 2018. Zum Konzept 'Fehler' am Beispiel der Analyse des „Novyj Margarit“ des A. M. Kurbskij. Die Frage der Bewertung sprachlicher Richtigkeit in historischen slavischen Texten, in: Vittorio Springfield TOMELLERI; Inna Veniaminovna VERNER (Hrsgg.), *Latinitas in the Slavonic World. Nine Case Studies*, München – Berlin – Washington D. C.: Peter Lang, 237–280. (= *Specimina philologiae Slavicae* 192)
- Bounatirou, Elias. 2021. Neponjatij jazyk Kurbskogo: novye podchody k interpretaciji „anomalij“ v istoričeskich tekstach [‘Die unverständliche Sprache Kurbskij: neue Forschungsansätze zur Interpretation von ‘Anomalien’ in historischen Texten’]. *Vestnik SPbGU, Serija Jazyk i literatura* 18 (1), 37–55.
- Lyle Campbell. 2013. *Historical Linguistics: An Introduction*, new edition, 3 edition. Edinburgh University Press.
- Diver, W. (2012a). "PHONOLOGY AS HUMAN BEHAVIOR". In *Language: Communication and Human Behavior*. Leiden, The Netherlands: Brill. [https://doi-org.uaccess.univie.ac.at/10.1163/9789004209107\\_015](https://doi-org.uaccess.univie.ac.at/10.1163/9789004209107_015)
- Diver, W. (2012b). "THEORY". In *Language: Communication and Human Behavior*. Leiden, The Netherlands: Brill. [https://doi-org.uaccess.univie.ac.at/10.1163/9789004209107\\_019](https://doi-org.uaccess.univie.ac.at/10.1163/9789004209107_019)
- Diver, W. (2012c). "THE HISTORY OF LINGUISTICS IN THE WEST: HOW THE STUDY OF LANGUAGE WENT WRONG IN THE WESTERN TRADITION". In *Language: Communication and Human Behavior*. Leiden, The Netherlands: Brill. [https://doi-org.uaccess.univie.ac.at/10.1163/9789004209107\\_020](https://doi-org.uaccess.univie.ac.at/10.1163/9789004209107_020)
- Divjak, Dagmar, Janda, Laura A. and Kocahańska, Agata. "Why cognitive linguists should care about the Slavic languages and vice versa". *Cognitive Paths into the Slavic Domain*, edited by Dagmar Divjak and A Dressler, Alan (1987). "[The Large-Scale Streaming of Galaxies](#)". *Scientific American*. 257 (3): 46–55.
- Diver, W. (1975 [2012]). "The Nature of Linguistic Meaning." In Huffman and Davis (eds.) *Language: Communication and human behavior: The linguistic essays of William Diver*. Leiden/Boston: Brill, pp. 47–63.
- Goldin, V. E. & O. Yu. Kryuchkova. 2011. Korpus ruskoj dialektnoj reči: kontsepsiia i parametry otsenki [Corpus of Russian Dialectal Speech: Concept and Evaluation Parameters]. In *Komp'uternaia lingvistika i intellektual'nye tekhnologii : materialy ezhegodnoi mezhdunarodnoi konferentsii, bekasovo, 25–29 maia 2011 goda* [computational linguistics and intellectual technologies: proceedings of the annual international conference, bekasovo, may 25–29, 2011], vol. 10, 359–367. Moscow: Russian State University for the Humanities.
- Guo, Yanzhu, Simone Conia, Zelin Zhou, Min Li, Saloni Potdar & Henry Xiao. 2025. Do large language models have an English accent? evaluating and improving the naturalness of multilingual LLMs. In *Wanxiang Che, Joyce Nabende, Ekaterina Shutova & Mohammad Taher Pilehvar*(eds.), *Proceedings of the 63rd annual meeting of the association for computational linguistics (volume 1: long papers)*, 3823–3838. Vienna, Austria: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2025.acl-long.193>. <https://aclanthology.org/2025.acl-long.193/>.
- Haspelmath, Martin. 2018. How comparative concepts and descriptive linguistic categories are different [published version]. In *Aspects of linguistic variation*, ed. by van olmen, daniël & mortelmans, tanja & brisard, frank, 83–114. N. L.: De Gruyter Mouton.
- Hromko, T. 2020. Stanovlennia monohovirkovoi deskriptsii u vitczyznianomu movoznavstvi (kinets' xix – 40-vi roky xx st.) [formation of monographic description in domestic linguistics (late 19th – 1940s)]. *Aktual'ni pytannia humanitarnykh nauk* [Current Issues in Humanities] 34(2), 118–123.
- Jelinek, F., R. L. Mercer, L. R. Bahl & J. K. Baker. 1977. Perplexity—a measure of the difficulty of speech recognition tasks. *The Journal of the Acoustical Society of America* 62(51). 563. <https://doi.org/10.1121/1.2016299>
- Labov, W. (1966). *The social stratification of English in New York City*. Washington DC Center for Applied Linguistics.
- Lehky SR, Sereno AB. Comparison of shape encoding in primate dorsal and ventral visual pathways. *J Neurophysiol* 97: 307–319, 2007.
- Nguyen, Li, Taslimipoor, Shiva and Yuan, Zheng. "What can NLP do for linguistics? Towards using grammatical error analysis to document non-standard English features" *Linguistics Vanguard*, vol. 10, no. 1, 2024, pp. 465-477. <https://doi-org.uaccess.univie.ac.at/10.1515/lingvan-2024-0001>
- Mansfield, J., Leslie O'Neill, H., & Li, H. (2023). Dialect differences and linguistic divergence: A crosslinguistic survey of grammatical variation. *Language Dynamics and Change*, 13(2), 232-276.
- Marie-Catherine de Marneffe, Christopher Manning, Joakim Nivre, and Daniel Zeman (2021). [Universal Dependencies](#). *Computational Linguistics* 47(2): 255–308.
- Andrea Moro. 2017. *Copular Sentences*. The Wiley Blackwell Companion to Syntax, Second Edition (eds M. Everaert and H.C. Riemsdijk).
- Otheguy, Ricardo. 2002. Saussurean anti-nomenclaturism in grammatical analysis: a comparative theoretical perspective. In *Signal, meaning, and message: perspectives on sign-based linguistics*, 373–403. N. L.: John Benjamins Publishing Company. <https://doi.org/10.1075/sfl.48.18oth>. <https://doi.org/10.1075/sfl.48.18oth>.
- Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton & Christopher D. Manning. 2020. Stanza: a python natural language processing toolkit for many human languages. In *Asli Celikyilmaz & Tsung-Hsien Wen* (eds.), *Proceedings of the 58th annual meeting of the association for computational linguistics: system demonstrations*, 101–108. Online: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-demos.14>. <https://aclanthology.org/2020.acl-demos.14/>.
- R Core Team (2026). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. doi:10.32614/R.manuals <<https://doi.org/10.32614/R.manuals>>. <<https://www.R-project.org/>>.
- Scherrer, Y 2021, *Adaptation of morphosyntactic taggers : Cross-lectal and multilectal approaches . in M Zampieri & P Nakov* (eds) , *Similar languages, varieties, and dialects : A computational perspective . Studies in Natural Language Processing* , Cambridge University Press , pp. 138-166 . <https://doi.org/10.1017/9781108565080.011>
- Suzuki R, Terada Y, Shimodaira H (2019). pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. R package version 2.2-0, <https://CRAN.R-project.org/package=pvclust>.
- Oleksiy Syvokon, Olena Nahorna, Pavlo Kuchmiichuk, and Nastasiia Osidach. 2023. [UA-GEC: Grammatical Error Correction and Fluency Corpus for the Ukrainian Language](#). In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, pages 96–102, Dubrovnik, Croatia. Association for Computational Linguistics.
- Terhart, Lena. 2024. *A grammar of Paunaka* [Comprehensive Grammar Library 7]. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.10517532>.
- Rigouts Terryn, Aylla & Miryam de Lhoneux. 2024. Exploratory study on the impact of English bias of generative large language models in Dutch and French. In *Simone Balloccu, Anya Belz, Rudali Hudrom, Ehud Reiter, Joao Sedoc & Craig Thomson* (eds.), *Proceedings of the fourth workshop on human evaluation of nlp systems (humeval) @ Irec-colling 2024*, 12–27. Torino, Italia: ELRA & ICCL. <https://aclanthology.org/2024.humeval-1.2/>.
- Yu, G., Smith, D.K., Zhu, H., Guan, Y. and Lam, T.T.-Y. (2017). gtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol*, 8: 28-36. <https://doi.org/10.1111/2041-210X.12628>
- Weber, M. 1904. Die „Objektivität“ sozialwissenschaftlicher und sozialpolitischer erkenntnis. *Archiv für Sozialwissenschaft und Sozialpolitik* 19(1), 22–87. <https://nbn-resolving.org/urn:nbn:de:0168-ssaar-50770-88>.
- Wiese, Heike. 2023. *Grammatical systems without language borders: Lessons from free-range language*. (Conceptual Foundations of Language Science 9). Berlin: Language Science Press. DOI: 10.5281/zenodo.10276182