

## The list of main publications and talks by Ilia Afanasev (as of June 4<sup>th</sup> 2026)

### Main publications

---

- Journal article** Afanasev, Ilia. (forthc.) A corpus-driven analysis of linguistic variation in late XII – early XIV century Old East Slavic charters. *Journal of Historical Linguistics*. Preprint available at: <https://works.hcommons.org/records/557wk-p4hg7>.
- Conference paper** Afanasev, Ilia. 2026. TransVar – the Corpus for Variation and Change Study of the Historical Transcarpathian lects. Proceedings of Workshop on Dialects in NLP – A Resource Perspective (DialRes) @ LREC 2026, Palma de Mallorca, Spain, 2026. 195–208.
- Conference paper** Poritski, Vladislav et al. 2026. LLM as a Morphological Disambiguator for Belarusian: A Preliminary Study. *Proceedings of SIGUL 2026 Joint Workshop with ELE, EURALI, and DCLRL @ LREC 2026*, Palma de Mallorca, Spain, 2026. 42–48.
- Conference paper** Afanasev, Ilia. 2026. Quantitative Lect Description: A Case Study of Lemko from the Field Data of 1920s-1930s. *Proceedings of the Fifth Workshop on NLP Applications to Field Linguistics*, Rabat, Morocco, 2026. 46–59.
- Conference paper** Afanasev, Ilia. 2026. Evaluation Framework for Transfer Learning between Closely Related Lects: A Case Study of Lemko. *Proceedings of the 13th Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2026)*, Rabat, Morocco, 2026. 304–316.
- Journal article** Afanasev, Ilia. 2025. Computer-assisted Study of Historical Lemkian (Transcarpathian) Lects: Basic Vocabulary Approach. *Scripta and e-Scripta 25*, 2025. 11-24.
- Conference paper** Afanasev, Ilia et al. 2025. Rubic2: Ensemble Model for Russian Lemmatization. *Proceedings of the 10th Workshop on Slavic Natural Language Processing (Slavic NLP 2025)*, Vienna, Austria. 157–170.
- Conference paper** Afanasev, Ilia & Lyashevskaya, Olga. 2025. The Application of Corpus-Based Language Distance Measurement to the Diatopic Variation Study (on the Material of the Old Novgorodian Birchbark Letters). *Proceedings of the Third Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL-2025)*, Tallinn, Estonia. University of Tartu Library, Estonia. 153-164.
- Book chapter** Afanasev, Ilia. 2024. The Cross-Evaluation Crux for Computational Phylogenetic Linguistics. *Digital Geography. IMS 2023*. Springer Geography. Springer, Cham. 75-89.
- Book chapter** Afanasev, Ilia & Lyashevskaya, Olga. 2024. Chapter 2 String Similarity Measures for Evaluating the Lemmatisation in Old Church Slavonic". *Structuring Lexical Data and Digitising Dictionaries*. Leiden, The Netherlands: Brill. 13–35
- Journal article** Afanasev, Ilia. 2023. Cipher, transform, get lost: an anti-transparent system for distance measurement in East Slavic lects. *Journal of Language Relationship 21* (3/4). 159 – 177.
- Conference paper** Afanasev, Ilia. 2023. Multi-lect automatic detection of Swadesh list items from raw corpus data in East Slavic languages. *Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change*, Singapore. Association for Computational Linguistics, 76–86.
- Conference paper** Afanasev, Ilia et al. 2023. The Effect of (Historical) Language Variation on the East Slavic Lects Lemmatisers Performance. *Journal of Linguistics 74*(1). 225 – 233.
- Conference paper** Afanasev, Ilia & Lyashevskaya, Olga. 2023. From web to dialects: how to enhance non-standard Russian lects lemmatisation? *Proceedings of the 2023 CLASP Conference on Learning with Small Data*. 167–175.
- Book chapter** Afanasev, Ilia & Babanov, Andrey. 2023. The Old Church Slavonic Corpora and

- Their Use in Language Studies at the University. *Literature, Language and Computing: Russian Contribution*. 43 – 58.
- Conference paper** Lyashevskaya, Olga et al. 2023. Disambiguation in context in the Russian National Corpus: 20 years later. *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2023"*. 1-12.
- Conference paper** Afanasev, Iliia. 2023. The Use of Khislavichi Lect Morphological Tagging to Determine its Position in the East Slavic Group. *Proceedings of Tenth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2023)*, Dubrovnik, Croatia. 174-186.
- Journal article** Babanov, Andrey & Afanasev, Iliia. 2021. Opis struktur syntaktycznych we wczesnych pracach Zenona Klemensiewicza i Noama Chomskiego. *Językoznawstwo* 15. 73-80.
- Conference paper** Lyashevskaya, Olga & Afanasev, Iliia. 2021. An HMM-based PoS Tagger for Old Church Slavonic. *Journal of Linguistics* 72(2). 556-567.
- Preprint** Afanasev, Iliia, A Hybrid Lemmatiser For Old Church Slavonic (February 18, 2021). *Higher School of Economics Research Paper* No. WP BRP 106/LNG/2021, Available at SSRN: <https://ssrn.com/abstract=3787995> or <http://dx.doi.org/10.2139/ssrn.3787995>
- Journal article** Afanasev, Iliia. 2020. A corpus-based approach in archaeolinguistics. *Journal of Applied Linguistics and Lexicography* 2(2). 147-159.
- Journal article** Afanasev, Iliia & Kryuchkova, Olga. 2019. Automation In Cognitive Linguistics: The Case Of Old Slavonic And Old Church Slavonic Literary Texts. *Lingua Viva* 29. 48-57.

## Main talks

- 
- Conference talk** *Language polymorphism markers: a new way to predict language variation and change* (03/06/2026, ACDH-CH Research Day 8, Vienna, Austria)
- Seminar talk** *BEFORE as meaning of -ŕ-in A. Kuśniewicz's Król Obojga Sycylii 'King of the two Sicilies'* ( 22/05/2026, Columbia School of Linguistics seminar, Online/New York, USA)
- Seminar talk** *Corpus-level comparison of eastern Transcarpathian Slavic varieties* (20/05/2026, Slavistisch-linguistisches Kolloquium, Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Germany)
- Conference talk** *TransVar – the Corpus for Variation and Change Study of the Historical Transcarpathian lects* (16/05/2026, Workshop on Dialects in NLP — A Resource Perspective (DialRes) @ LREC 2026, Palma de Mallorca, Spain)
- Conference talk** *LLM as a Morphological Disambiguator for Belarusian: A Preliminary Study*. (with Vladislav Poritski and Oksana Volchek, 11/05/2026, SIGUL 2026 Joint Workshop with ELE, EURALI, and DCLRL @ LREC 2026, Palma de Mallorca, Spain)
- Conference talk** *Exploring textual tradition through LLMs: an analysis of parallel Old and Middle Bulgarian and Middle Greek corpus* (with Ivan P. Petrov, 08/05/2026, Colloquia Ceranea VIII, Łódź, Poland)
- Conference talk** *Quantitative Lect Description: A Case Study of Lemko from the Field Data of 1920s-1930s* (29/03/2026, the 5th Workshop on NLP Applications to Field Linguistics, Rabat, Morocco)
- Poster talk** *Evaluation Framework for Transfer Learning between Closely Related Lects: A*

- Seminar talk** *Case Study of Lemko* (29/03/2026, the 13th Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2026), Rabat, Morocco)
- Seminar talk** *Značenie -Ń v pol'skom jazyke: Predvaritel'nye zamečanija* [Meaning of -Ń in Polish: Preliminary Notes] (11/03/2026, Seminar po istoričeskoj antropologii, istoričeskoj semantike i istorii ponjatij [Seminar on historical anthropology, historical semantics and history of concepts], online)
- Seminar talk** *A study of the diatopic variation in Old East Slavic charters by means of a small unannotated corpus-based language comparison* (09/03/2026, Data in Historical Linguistics, online)
- Guest talk** *Wie die Sprachwissenschaft die Literaturwissenschaft bereichern kann* [How linguistics can assist literature studies] (26/02/2026, Dokumentationsstelle für ost- und mitteleuropäische Literatur, Vienna, Austria)
- Conference talk** *Heatmap-based visualisation of the linguistic variation (on the material of East Slavic small territorial lects)* (07/12/2025, Digital Linguistics panel @ 49. Austrian Linguistics Conference, Klagenfurt, Austria)
- Conference talk** *Закарпатські діалекти за матеріалами початку ХХ століття: спроба кількісного опису* [Transcarpathian dialects based on materials from the early 20th century: an attempt at a quantitative description] (02/12/2025, Тимченківські читання.9 [Тymchenko readings.9], Lviv, Ukraine)
- Conference talk** *Transkarpatische Materialien in der Bibliothek der Universität Wien: ein Überblick* [Transcarpathian materials in the library of the University of Vienna: a short overview] (14/11/2025, Die Jahrestagung und Generalversammlung der Österreichischen Gesellschaft für Slawistik, Klagenfurt, Austria)
- Conference talk** *Ispol'zovanie korpusnogo materiala v klasterizacii blizkorodstvennyh lektov (na primere Saratovskogo dialektologičeskogo korpusa)* [Using corpus material to cluster closely related lects (on the material of Saratov dialectological corpus)] (14/11/2025, Jazyk i obshhestvo v sinhronii i diahronii [Language and society in synchronic and diachronic studies], Saratov, Russia)
- Conference talk** *Stvareńnie korpusa dla daśledavańniaŭ varyjatyŭnaści na materyjale biełaruskaj movy* [Creating a corpus for studying language variation in Belarusian] (27/09/2025, XII International Congress of Belarusian Studies, Berlin, Germany)
- Conference talk** *Korpusbasierte Erforschung der altostslawischen Rechtsurkunden des 12.–14. Jahrhunderts* [Corpus-based research of the 12th–14th century Old East Slavic legal charters] (19/09/2025, 33. Treffen der Jungslavist\*innen, Freiburg im Breisgau, Germany)
- Conference talk** *We do not frequently borrow words that we do not borrow frequently: the search for quantifiable characteristics of swadeshness* (17/09/2025, Protolang 9, Vienna, Austria)
- Conference talk** *Rubic2: Ensemble Model for Russian Lemmatization* (with O. Lyashevskaya et al., 31/07/2025, 10th Workshop on Slavic Natural Language Processing (Slavic NLP 2025), Vienna, Austria)
- Conference talk** *Grammatical system of czas przyszły złożony in Polish: preliminary considerations* (17/07/2025, ICLC-17, Buenos Aires, Argentina)
- Conference talk** *Introduction: On Columbia School linguistics and the ethnopragmatic approach* (with L. Novotny et al., ICLC-17, Buenos Aires, Argentina)
- Conference talk** *Transcribing Transcarpathian: the challenges of historical small territorial lects digitisation* (07/06/2025, 2. SpraWIEN, Vienna, Austria)
- Conference talk** *Computer-assisted study of the Old East Slavic charters: The Smolensk Treaty of 1229* (26/03/2025, The 16th Paleoslavists Conference, Gent, Belgium)
- Seminar talk** *Analysing the patterns of language change through the lens of a corpus (on the material of East Slavic languages)* (14/04/2025, Data in Historical Linguistics, online)

- Conference talk** *Wordlist as a Corpus: Using Post-Hoc Approaches in Historical Linguistics* (24/03/2025, The 20th Sergei Starostin Memorial Conference on Comparative Historical Linguistics, Moscow, Russia)
- Conference talk** *Linguistic inquiry into the 12-14th century charters of the East Baltic region lects: Qualitative exploration of quantitative analysis* (with O. Lyashevskaya, 15/03/2025, Uncovering past with linguistic data, Riga, Latvia)
- Guest talk** *Lectometry studies of the Belarusian lects in contact: Current state and prospects* (07/03/2025, 2. Belarus Digital, Graz, Austria)
- Conference talk** *The Application of Corpus-Based Language Distance Measurement to the Diatopic Variation Study (on the Material of the Old Novgorodian Birchbark Letters)* (with O. Lyashevskaya, 02/03/2025, the Third Workshop on Resources and Representations for Under-Resourced Languages and Domains (RESOURCEFUL-2025), Tallinn, Estonia)
- Conference talk** *Computational representation of the language variation and change patterns in the historically attested dialects of Ukrainian* (20/02/2025, ACDH-CH Research Day 6, Vienna, Austria)
- Conference talk** *Multi-lect automatic detection of Swadesh list items from raw corpus data* (19/12/2024, Digital Slavic Studies panel @ 48. Austrian Linguistics Conference, Innsbruck, Austria)
- Conference talk** Digitalisierung der Dialekten und die historische-vergleichende Sprachwissenschaft [Digitilisation of dialects and the historical-comparative linguistics] (15/11/2024, Die Jahrestagung und Generalversammlung der Österreichischen Gesellschaft für Slawistik, Graz, Austria)
- Seminar talk** *It is so złożone: Why there is no way to study Polish Compound Future without the Columbia School of Linguistics* (08/11/2024, Columbia School of Linguistics seminar, Online/New York, USA)
- Conference talk** *Automatische Suche des Grundwortschatzes in altostslawischen Texten* [Multi-lect automatic identification of basic vocabulary items in the Old East Slavic texts] (24/09/2024 — 25/09/2024, 32. Treffen der Jungslawist\*innen, Vienna, Austria)
- Conference talk** *Combining string similarity measures and frequency-based metrics: a new approach to measuring language distance between Slavic lects* (with O. Lyashevskaya, 16/08/2024, 11th Inter-Varietal Applied Corpus Studies (IVACS) conference (IVACS 2024), University of Cambridge, England)
- Conference talk** *Bol'shie tipologicheskie dannye i sravnitel'no-istoricheskoe jazykoznanie: (pochti) pjatnadcat' let spustja* [Big typological data and comparative historical linguistics: (almost) fifteen years later] (23/05/2024 – 24/05/2024, Terra Hominis II, Moscow, Russia)
- Conference talk** *Podbor vneshnej gruppy v uslovijah dialekt'nogo kontinuuma: predvaritel'nye zamechanija k filogeneticheskomu analizu slavjanskih jazykov* [Outgrouping in a dialect continuum: preliminary comments on phylogenetic analysis of Slavic languages] (14/05/2024 – 15/05/2024, V. Dybo Memorial Conference, Moscow, Russia)
- Conference talk** *Perekrjostnaja ocenka jeffektivnosti predvaritel'noj geneticheskoy klassifikacii dialektov s ispol'zovaniem malyh neobrabotannyh korpusov (na materiale vostochnoslavjanskih dialektov)* [Cross-evaluation techniques for evaluation of preliminary dialect genetic classification on the small raw corpora material for East Slavic languages] (13/05/2024 – 14/05/2024, Slavic Languages: Responding to New Challenges – X, Tomsk, Russia)
- Conference talk** *Generativnye modeli i lemmatizacija russkogo jazyka: novye rezul'taty* [Generative models and Russian language lemmatisation: new results] (with A. Glazkova, D. Morozov and O. Lyashevskaya, 25/04/2024, 8th Kolmogorov Seminar on Computational Linguistics and Language Sciences, Moscow, Russia)
- Conference talk** *Measuring the language distance between the lects with high degree of inner*

- variation (on the material of South Slavic lects) (with V. Kozak, 28/03/2024 – 29/03/2024, The 19th Sergei Starostin Memorial Conference on Comparative-Historical Linguistics, Moscow, Russia)*
- Poster** *Multi-lect automatic detection of Swadesh list items from raw corpus data in East Slavic languages (06/12/2023, 4th Workshop on Computational Approaches to Historical Language Change, Singapore. Association for Computational Linguistics, Singapore)*
- Conference talk** *The effect of (historical) language variation on the East Slavic lects lemmatisers performance (with O. Lyashevskaya et al., 18/10/2023 – 20/10/2023, Slovko 2023, Bratislava, Slovakia)*
- Poster** *Ukrainian Universal Dependencies: A New Approach to a Corpus-Based Analysis of Ukrainian Language Structures (12/10/2023 – 13/10/2023, Ukrainistik entwickeln: Ukraine in Forschung und Lehre in den Fächern Slavistische Literatur- und Kulturwissenschaft, Slavistische Linguistik im Dialog mit der Osteuropäischen Geschichte, Jena, Germany)*
- Conference talk** *The current unique state of Belarusian language and Belarusian corpus linguistics (22/09/2023 – 24/09/2023, XI International Congress of Belarusian Studies, Gdansk, Poland)*
- Conference talk** *Generativnyj podhod k lemmatizacii dialektnyh tekstov [Generative approach to the dialect texts lemmatisation] (with O. Lyashevskaya, 20/09/2023, The 47th Information Technologies and Systems Workshop, Dubna, Russia)*
- Conference talk** *From web to dialects: how to enhance non-standard Russian lects lemmatisation? (with O. Lyashevskaya, 11/09/2023 – 12/09/2023, Learning with Small Data, Gothenburg, Sweden)*
- Conference talk** *Disambiguation in context in the Russian National Corpus: 20 years later (with O. Lyashevskaya et al., 14/06/2023 – 16/06/2023, Dialog'23, Moscow, Russia)*
- Poster** *The Use of Khislavichi Lect Morphological Tagging to Determine its Position in the East Slavic Group (04/05/2023 – 06/05/2023, Tenth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2023), Dubrovnik, Croatia)*
- Conference talk** *Lemmatizacija malyh territorial'nyh lektov s ispol'zovaniem ljogkih transformerov [Small territorial lects lemmatisation with the use of light transformers] (with O. Lyashevskaya, 25/04/2023, 7th Kolmogorov Seminar on Computational Linguistics and Language Sciences, Moscow, Russia)*
- Conference talk** *Considering the classification of some East Slavic lects with the automatic search of linguistic SNP-markers (23/03/2023 – 24/03/2023, The 18th Sergei Starostin Memorial Conference on Comparative-Historical Linguistics, Moscow, Russia)*
- Conference talk** *The use of transformer models for modern and historical Russian texts lemmatisation (with O. Lyashevskaya et al., 02/10/2022 – 06/10/2022, The 46th Information Technologies and Systems Workshop, Dubna, Russia)*
- Conference talk** *Automatic collection of parallel thesaurus in dictionary/corpus joint system (with O. Lyashevskaya and V. Kozak, 22/06/2022 – 24/06/2023, The 12th International Conference on Historical Lexicography and Lexicology, Lorient, France)*
- Conference talk** *K zadache razrabotki versii korpusov NKRJa s razreshennoj neodnoznachnost'ju morfolozicheskoj i sintaksicheskoj razmetki [On the task of the development of morphologically and syntactically disambiguated RNC corpora] (with N. Vlasova et al., 25/04/2022, 6th Kolmogorov Seminar on Computational Linguistics and Language Sciences, Moscow, Russia)*
- Conference talk** *Avtomaticheskoe izmerenie vlijanija sporadicheskikh izmenenij na jevoljuciju jazyka [Automatic measurement of sporadic changes influence on the language evolution] (25/04/2022, 6th Kolmogorov Seminar on Computational Linguistics and Language Sciences, Moscow, Russia)*
- Conference talk** *The influence of dialect distance on the performance of sequence transformation systems (27/04/2022 – 28/04/2023, CorCoDial workshop,*

- Helsinki, Finland)
- Conference talk** *An HMM-based PoS Tagger for Old Church Slavonic* (with O. Lyashevskaya, 13/10/2021 – 15/10/2023, Slovko'21, Bratislava, Slovakia)
- Conference talk** *The linguistic features of XIII century Smolensk and Polotsk idioms on the material of governmentese manuscripts* (01/10/2021 – 03/10/2023, IX International Congress of Belarusian Studies, Kaunas, Lithuania)
- Conference talk** *Lemmatization in corpus-to-dictionary systems: The case study for Old Church Slavonic* (16/06/2021 – 18/06/2023, ICHLL'11, Logroño, La Rioja, Spain)
- Conference talk** *The Old Church Slavonic corpus development: preprocessing stage* (12/04/2021 – 15/04/2023, El'Manuscript-2021, Freiburg, Germany)
- Conference talk** *Corpus-dictionary system: introducing a concept* (12/11/2020 – 14/11/2020, "Linguistics Forum-2020: Language and AI", RAS, Moscow, Russia)