

Reports / Хроника

The 21st Annual Sergei Starostin Memorial Conference on Comparative-Historical Linguistics Moscow, HSE University, March 23–24, 2026

The 21st Annual Sergei Starostin Memorial Conference on Comparative-Historical Linguistics was held in Moscow at HSE University on March 23–24, 2026. Organized by the Centre for Comparative Studies and Phylogenetics of the Institute for Oriental and Classical Studies, HSE University, the conference typically brings together researchers working on languages from a wide range of families and geographical areas, and this time was no exception as the schedule covered traditional problems of comparative-historical linguistics — phonological and morphological reconstruction, language classification and etymology — alongside lexicostatistics, computational phylogenetics, semantic change, and language contact.

As usual, *Indo-European studies* occupy a prominent spot in the program. This time, Artyom Badeev (Institute of Linguistics of the Russian Academy of Sciences) opened the talks with **“On the origin of indefinite markers in Mitagi-Dzhalgan /Tat/ language.”** Mitagi-Dzhalgan (Tat) developed for a long period in relative isolation from the Southwestern Iranian language group, resulting in a system of indefinite pronouns and determinatives distinct from that of Middle and New Persian. The absence of a postpositive indefinite article is compensated by the grammaticalization of the numeral *ye* ‘one’ as a prepositional indefinite marker: *ye ki* ‘someone’, *ye šəy* ‘something’, *ye ædəmi* ‘some person’. A range of postpositive markers, some derived from the verbs ‘to be’ and ‘to want; to love’, points to an incomplete process of grammaticalization. At the same time, the Mitagi-Dzhalgan system shows parallels with the Northwestern Iranian Caspian languages and with Azerbaijani.

Sergei Bolotov (Institute of World Culture, Moscow State University) addressed the long-standing problem of Szemerényi’s law in his paper **“Queen against fox and the Latin prosody in Sanskrit: ‘anti-Szemerényi laws’ with their sequence.”** Rather than treating the forms traditionally associated with the law as the outcome of a single process, the paper considers a set of elementary rules operating in particular languages

and morphological environments. The resulting system accounts both for forms displaying the apparent “Szemerényi effect” and for those in which it is absent. Greek γυνή ‘woman’, for instance, behaves differently from ἀλώπηξ ‘fox’, while different types of -ντ-participles likewise show distinct but regular developments. Sanskrit and Greek participial formations provide further examples of outcomes that are regular precisely because they are not reducible to a single general rule.

Alexandra Evdokimova (Institute of Linguistics of the Russian Academy of Sciences) turned to **accent curves of the second declension in Byzantine Greek**, drawing on the BGAT corpus. Byzantine grammatical works, including those associated with Theodore Gaza and Manuel Moschopoulos, transmit numerous examples whose accentuation varies from manuscript to manuscript. Comparison of the same examples across different grammatical texts shows that some scribes consistently preferred particular accentual patterns, while others reproduced the forms found in their sources. Such variation provides evidence for the emergence and transmission of accentuation patterns in Byzantine grammatical tradition.

Elena Efimova (Tyumen State University) explored **broad-meaning verbal vocabulary in Germanic languages and its trajectories of semantic development.** The analysis focuses on the verb ‘to do’ in Gothic, Old High German, Old Norse, and Old English, combining comparative-historical analysis with distributional semantics. Thus, Germanic **dōn-* continues Proto-Indo-European **d^heh₁-* ‘to put, place’: from meanings associated with placing and producing objects, as well as creating or establishing conditions and rules, the verb has developed spatial-temporal, institutional, and abstract-causative uses. The individual Germanic languages followed these trajectories in different ways.

Anastasia Garnet Romantsevski (Institute of Linguistics of the Russian Academy of Sciences; Institute of Oriental and Classical Studies, RSUH) discussed **“The Epitaph of Ursula: A Linguistic Commentary**

on Inscription CIL XIII 3909. The Latin-Gallic funerary inscription, found near Augusta Treverorum (modern Trier) and tentatively dated to the fourth or fifth century AD, is inscribed on a clay tile and provides valuable evidence for Gallo-Latin bilingualism and language contact in Late Roman Gaul. Particular attention was paid to the relationship between the daughter's Latin name *Ursula* ('Little She-Bear') and her mother's Gallic name *Artula*, interpreted as its semantic and phonetic correlate and derived from Gallic *artos* 'bear'. The paper also addressed alternative readings of the presumably Gallic final part of the inscription, *Kara matir titiulum po(suit)*. The linguistic commentary bears on both the reconstruction of the inscription and the linguistic competence of the local Treverian population, as well as on the representation of *Latinitas vulgata* in provincial epigraphy.

Artem Trofimov (School for Advanced Studies in the Humanities, RANEP; Institute of Linguistics of the Russian Academy of Sciences) considered **"Anatolian and Tocharian verbs with secondary semantic developments compared to Proto-Indo-European."** Anatolian and Tocharian are generally regarded as the first and second branches to have split from the Proto-Indo-European language community. Their lexical evidence sometimes preserves earlier meanings, as in Hittite *mer-/mar-* 'to disappear' versus Narrow Indo-European **mer-* 'to die', and Hittite *pāš-* 'to swallow' versus Narrow Indo-European *peh₃-* 'to drink'. At the same time, both branches contain verbal stems with secondary semantic developments, often associated with morphological derivation. The paper examined such cases on the basis of etymological dictionaries and databases.

On the source data of Indo-European languages (mainly Celtic and Slavic), Tatyana Mikhailova (Institute of Linguistics of the Russian Academy of Sciences; Institute of Oriental and Classical Studies, RSUH) examined **"two-step semantic shifts"**, drawing on data from DatSemShift and related databases. The discussion centered on semantic developments associated with the concepts 'life' and 'food', including the shift *to live, to be alive* → *food* and subsequent developments toward 'meat', 'rice', and 'cereals'. Particular attention was given to the interpretation of Welsh *bwyd* 'food', the history of Slavonic *žirь*, and the possibility of reconstructing an unrecorded intermediate meaning in the development *to live* → *food* → *cereal*. Insular Celtic evidence for a possible *to eat* → *cereal* shift was also considered. The broader issue is whether two- or more-step semantic shifts can be reconstructed on the basis of motivated intermediate stages, and to what extent such stages may help re-

cover meanings not directly attested in the historical record.

Linguistic contacts between Indo-European and other language families (in this case, Semitic) were also touched upon by Alexey Lyavdansky (Institute for Classical and Oriental Studies, HSE University), who examined **"Iranian Lexical Borrowings in Eastern Middle Aramaic: A Dialectological Perspective."** Iranian-Aramaic contact represents a long and complex history of interaction, but Iranian borrowings in Jewish Babylonian Aramaic and Mandaic remain less thoroughly studied than those in Classical Syriac. Several Iranian loans shared by Jewish Babylonian Aramaic and Mandaic but absent from Syriac were considered, alongside loans occurring in all three dialects whose phonetic development displays features common to Jewish Babylonian Aramaic and Mandaic. Examples include MP *zyān* 'loss, damage' and Mlr *frasaxa* 'parasang'. Such Mandaic-Jewish Babylonian Aramaic lexical and morpho-lexical isoglosses are relevant to the subgrouping of Eastern Aramaic and add to the evidence for a closer relationship between the two dialects.

Turning to other language families of Eurasia, comparative *Uralic studies* were represented by two presentations:

Valentin Gusev (Institute of Linguistics of the Russian Academy of Sciences) discussed **vowel harmony in Nganasan**, focusing on its history and present-day structure. Although the modern system resembles front-back harmony, the phonetic motivation for this opposition has been weakened by subsequent changes in the vowel system. Historical correspondences nevertheless point to two harmonic classes and make it possible to reconstruct phonetically natural front-back vowel pairs. The evidence suggests that Proto-Samoyedic non-first-syllable vocalism was not as impoverished as has sometimes been assumed. Low vowels and schwa appear particularly frequently, while mid vowels are unexpectedly more frequent than high vowels.

Leonid Zaitsev (Lomonosov Moscow State University) presented **"Stepwise reconstruction for morphophonology and Uralic consonant gradation."** The starting point is E. A. Helimski's reconstruction of a complex Proto-Uralic system of consonant alternations involving rhythmic and syllabic gradation. One difficulty of this reconstruction lies in comparing relatively recent material from a language long separated from its closest relatives with the proto-languages of other branches. Forest Nenets, which preserves a subsystem of cluster gradation, provides an additional intermediate stage. In comparison with Tundra Nenets

and Nganasan, these data allow consonant gradation to be reconstructed at the Proto-Nenets and Proto-(Northern) Samoyed levels. Stepwise reconstruction thus makes the intermediate systems comparable with other Uralic proto-languages and strengthens the case for Proto-Uralic consonant gradation.

The field of *Turkic studies* was represented by Alexander Savelyev (Institute of Linguistics of the Russian Academy of Sciences; Institute for Oriental and Classical Studies, HSE University; Institute for History and Archaeology, Ural Branch of the Russian Academy of Sciences), who gave a talk on **“Chuvash wC-type clusters in the light of Proto-Turkic reconstruction and Altaic comparison.”** The paper systematically examines cases in which Chuvash *wC-type clusters correspond to a plain *C in Narrow Turkic. Given the binary structure of the Turkic family, possible Proto-Turkic reconstructions must also take Altaic parallels into account. The evidence argues against treating Chuvash *w as an “intrusive” element: it primarily continues Proto-Turkic *b and *g, with a few other sources occurring more rarely. The analysis also points to the need for a more systematic incorporation of Bulgharic evidence into Proto-Turkic reconstruction.

Regarding the areal relationships of specific Turkic languages, Timofey Kuzmenko (Institute of Asian and African Studies, Lomonosov Moscow State University) discussed **“Possible Middle Chinese borrowings in the Salar language.”** Phonetic patterns in loanwords from Northern Chinese dialects suggest that some items may have been borrowed from Middle Chinese. This would point to earlier and closer contacts between Salar and Middle Chinese than those implied by Ming historical chronicles and Salar folklore, which give 1370 as the date of the Salar “migration” to the territory of present-day China. The evidence may also be relevant to glottochronological estimates placing the separation of Salar and Turkmen in 1220. The paper seeks to identify a possible group of direct Middle Chinese loans and distinguish them from later borrowings from Northern Chinese dialects and Southern Mongolian languages.

Turning to the Caucasus and, specifically, the *Kartvelian family*, Roman Pavlov presented **“Principles for building an etymological database of the Kartvelian languages.”** The discussion starts from the problems posed by the existing Kartvelian etymological dictionaries, particularly the quality and reproducibility of the input data used for external comparison. Some phonological correspondences do not have sufficiently well-established distributions; dialectal Svan, Megrelian, and Laz material is often presented without precise dialectal or source information; and the selection

of verbal forms does not always adequately represent the distribution of ablaut grades. The proposed database structure is intended to address these issues and draws on the experience of such projects as *Lexikon der indogermanischen Verben*, *Nomina im Indogermanischen Lexikon*, and *A North Caucasian Etymological Dictionary*.

The adjacent (but not genetically related) *Nakh-Daghestanian family* was covered by two presentations. In one of them, Timur Maisak and Steven Kaye (Institute of Linguistics of the Russian Academy of Sciences; HSE University; University of Surrey) discussed **“Irregular presents of motion verbs in Andic languages: comparative data and a hypothesis on their origin.”** Most Andic languages, with the exception of Akhvakh and Tindi, possess irregular present-tense forms of the motion verbs ‘go (away)’ and/or ‘come’. The present forms, or derived imperfective stems, contain a special marker with a lateral *ʃ* or *ɬ* in place of the regular present-tense or imperfective marker. This element resembles a reduced form of the general motion verb ‘go, walk’, which is not always preserved as an independent lexeme. The authors propose that the irregular forms originated in combinations of the aorist (perfective stem) of a directed motion verb with the present tense of the general verb ‘go, walk’, the latter subsequently becoming grammaticalized as a tense suffix or stem marker.

Another talk, by Mikhail Zhivlov (Institute of Oriental and Classical Studies, RSUH; Institute for Oriental and Classical Studies, HSE University), addressed **“Laryngeal theory or consonant stems? A problem in Proto-Andic reconstruction.”** The discussion centered on nouns with unusual correspondences involving resonants, vowel length, and nasalization. The word ‘water’, represented by forms such as Andi *ʔen* and Chamalal *ʔɪ*, provides a particularly clear example. Whereas the *North Caucasian Etymological Dictionary* reconstructs resonant-laryngeal clusters, an alternative analysis treats these forms as consonant stems ending in resonants. The two approaches lead to different interpretations of the Proto-Andic consonant system.

Several talks addressed contemporary issues in the classification and internal history of *Sino-Tibetan languages*. Anton Kogan (Institute of Oriental Studies of the Russian Academy of Sciences) proposed a classification of **Northwestern Tibetan dialects** based on glottochronological counts. The resulting dated genealogical tree provides a basis for hypotheses concerning the chronology of the spread of particular Tibetan varieties in the Upper Indus region and their contacts with the languages of the local pre-Tibetan population. The proposed chronology was compared with genetic data.

Sophia Khafizova (School of Linguistics, HSE University) gave a talk on the **“Classification of West-Gyalrongic languages based on the analysis of basic lexicon.”** Her study focuses on two branches of West Gyalrongic: Horpa, including Eastern Geshiza and *rTa’u*, and Khroskyabs, including Lavrung (Yelong and Guanyinqiao) and Khroskyabs Wobzi. The classification is based on basic-lexicon wordlists, and the results are intended to provide a basis for further etymological analysis.

Additionally, Li Andong (Peoples’ Friendship University of Russia) addressed some issues of Chinese (Sinitic) dialectology in **“On Some Phonetic Aspects of the Influence of Chinese Idioms on Mandarin.”** The phonetic norm of Mandarin in the People’s Republic of China is based on Beijing pronunciation, but a number of direct and indirect exceptions reflect other traditions. Orthoepic standardization affected both native “colloquial” and borrowed “literary” readings of Chinese characters, sometimes retaining a single reading and sometimes assigning different readings to specific domains of use. Southern Chinese varieties have also left their mark on phonetic loanwords, including proper names. Such forms may differ considerably from their original foreign pronunciation in Beijing Mandarin, while Southern traditions continue to survive within Putonghua. New borrowings may likewise show dialectal influence.

Comparative *Yeniseian studies* formed the principal topic of George Starostin’s (Institute for Oriental and Classical Studies, HSE University; Santa Fe Institute, New Mexico) presentation **“From ‘hyper-reconstruction’ to ‘optimisation’ (on the example of comparative Yeniseian material).”** The paper considered the tendency for established proto-language reconstructions to become increasingly elaborate through the addition of controversial phonetic correspondences, poorly substantiated material, internal reconstruction, “word-family” arguments, or hypothetical contacts with neighboring language families. The result may be a “hyper-reconstruction” in which a systematically established core is combined with debatable or highly improbable elements. The current state of Proto-Yeniseian reconstruction was reviewed, beginning with S. A. Starostin’s classic 1982 study and including subsequent alternative proposals. Rather than further complicating the reconstruction, the paper advocates “optimisation”: typological oddities and suspicious redundancies can be removed by identifying additional positional distributions and intra-family borrowings. As one example, the velar and uvular fricatives **x* and **χ* in the 1982 reconstruction can be eliminated without seriously affecting the system as a whole.

Turning further East — to the field of comparative *Austronesian studies* — we find Albert Davletshin’s (Institute of Oriental and Classical Studies, RSUH) talk on **“Laryngeal cooccurrence constraints in Proto-Polynesian.”** Polynesian roots generally exclude adjacent homorganic consonants, while codas are absent from most Polynesian languages. Within the Proto-Polynesian system, the laryngeals **ʔ* and **h* do not normally cooccur within a root, except that **h* may occur next to the rounded vowels **o* and **u*; **ʔ* is likewise incompatible with **k*. These restrictions are considered both from an articulatory and from a historical perspective. They also bear on the reconstruction of the Proto-Polynesian lexicon and may help account for the double reflexes **s* and **h* of Proto-Central Pacific ***c/**z*.

Evgeniya Korovina (Institute of Linguistics of the Russian Academy of Sciences) also used the Austronesian family as her source material while examining the role of the analytical method in **“Method as a Variable: Classifying the Languages of East Polynesia.”** Different clustering procedures may produce substantially different results from the same dataset, making it important to distinguish genuinely stable nodes from those introduced by a particular model. East Polynesian provides a useful testing ground: although the subgroup itself is clearly defined, its internal classification remains disputed, especially with regard to Hawaiian and Rapa Nui. Using basic vocabulary lists of different lengths, the paper compares classical distance-based procedures with phylogenetic methods, including algorithms capable of accounting for homoplasy. The broader question is what evidence is sufficient to establish a genealogically meaningful subgroup.

The African and American continents were represented by one talk each this time. Anna Smirnitskaya (Institute of Oriental Studies of the Russian Academy of Sciences) explored **the correlation between the cognitive mechanism of semantic shift and its morphological manifestation in the EvoLex etymological dictionary of Khoe-Kwadi.** The various types of semantic-shift realization described in DatSemShift 3.0 link lexical meanings at different semantic distances. Such relationships may arise through borrowing, metaphor, external similarity, or metonymy; in Khoekhoe, for example, *mú* means both ‘to see’ and ‘eye’. The greatest semantic distance is found among cognates displaying radial semantic development, as in the reflexes of PKK *g!uri*: ‘fat’ in Khoe ‘fat’ and Kora ‘height’, or PKK *ǀae*: ‘page’ in Naro and ‘arrow tip’ in Khoe. The analysis is based on the EvoLex comparative dictionary of Khoe-Kwadi languages.

Javier Carol (CONICET–Universidad Autónoma de Entre Ríos, Argentina) presented **“Person indices in Proto-Mataguyan.”** The Mataguyan family comprises Wichí, Chorote, Nivaclé, and Maká. Four principal personal paradigms can be reconstructed for the proto-language: active indicative, inactive indicative, non-indicative, and possessive. Considering these paradigms together is essential, since reflexes of the same index frequently occur in different paradigms across the daughter languages. The comparison was extended to Proto-Guaycuruan, a closely related family: most, if not all, personal indices in the two families have cognates. This evidence modifies conclusions based on Mataguyan alone and also sheds new light on the internal diversification of Proto-Mataguyan, which appears somewhat different when viewed through personal morphology rather than lexical data.

Finally, the Conference would be incomplete without at least one significant talk on the general methodology of historical linguistics, and on this occasion Mikhail Vasilyev, Arseniy Manusov, and Mikhail Saenko (Institute of Oriental Studies of the Russian Academy of Sciences; School of Linguistics, HSE University; Institute of Slavic Studies of the Russian Academy of Sciences) presented **“New linguistic software for constructing, dating and analyzing lexicostatistic trees.”** The prototype, developed since 2025, is designed for constructing and dating phylogenetic classifications on the basis of lexical data. It accepts both standard lexicostatistical databases, in-

cluding those created in Starling or Excel, and pre-calculated cognacy percentages. Among its features are an enhanced neighbor-joining procedure capable of producing non-binary nodes, the option of manually defining tree structures, and statistical tools for assessing the reliability of the source data and the resulting classification. Confidence intervals for individual nodes and graphical visualization of the results are also provided; further functions are planned for subsequent versions.

As usual, the 21st Sergei Starostin Memorial Conference brought together a particularly broad range of approaches to the reconstruction and classification of languages. Traditional comparative-historical analysis was represented alongside lexicostatistical and phylogenetic methods, while questions of semantic change, grammaticalization, language contact, and the organization of etymological data received considerable attention. The diversity of the material — from Iranian, Uralic, Nakh-Daghestanian, and Indo-European languages to Polynesian, Kartvelian, Turkic, Tibetan, Aramaic, Mataguyan, and Gyalrongic — once again reflected the wide comparative perspective that has long characterized the Starostin Memorial Conference.

The complete program of the Conference, along with Russian- and English-language abstracts and presentations from most of its participants, can be found online at the «Tower of Babel» web site (URL: starlingdb.org/meetings.php?lan=en).