Proceedings of the 17th International Scientific and Practical Conference «International Forum: Problems and Scientific Solutions»
(September 26-28, 2026).
Melbourne, Australia

No
311

PHILOLOGY AND LINGUISTICS

# Innovative Evolutionary Computing for Glottochronology and Linguistic Evolution: Trends, Paradigms, and Open Challenges

**Svitlana Krasniuk**[1]

[1]Senior Lecturer, Department of philology and translation;
*Kyiv National University of Technologies and Design*; Ukraine

**Abstract.** The article examines innovative approaches, trends, and prospects regarding the application of evolutionary computation to solve complex problems in evolutionary linguistics and glottochronology. It explores the potential for the algorithmic reconstruction of linguistic relatedness, cognate analysis, the construction of phylogenetic models, and the simulation of linguistic changes over time. The study substantiates the promise of hybrid, multi-objective, adaptive, and robust evolutionary methods in the face of the turbulence, uncertainty, and systemic risks characterizing the modern linguistic environment.
**Keywords**: *computational linguistics; evolutionary computation; evolutionary linguistics; glottochronology; language evolution; cognates; phylogenetic analysis; linguistic relatedness; evolutionary algorithms; genetic algorithms; machine learning; artificial intelligence; Big Data; multi-criteria optimization; hybrid intelligent technologies; uncertainty; turbulence; systemic risks.*

---

**Introduction.** Modern linguistics is evolving amidst the increasing complexity of its object of study, as language is a dynamic, self-organizing, and historically mutable system [1–3] whose functioning is determined by an interplay of phonetic, phonological, morphological, lexical, semantic, syntactic, sociocultural, and cognitive factors [4–6]. Particularly challenging tasks include reconstructing the historical dynamics of languages, determining the directions of their divergence, establishing genetic relationships, dating stages of linguistic divergence, and identifying the mechanisms behind the formation of language families. In this context, glottochronology emerges as a significant approach focused on the quantitative study of the rate and nature of linguistic change over time.

# PHILOLOGY AND LINGUISTICS

While traditional methods of historical-comparative linguistics remain fundamental, the sheer scale of modern linguistic data places increasingly high demands on the computational tools used for their processing [7]. Vast lexical datasets, cognate databases, multilingual corpora, phonetic and morphological characteristics, historical dictionaries, and digital language databases create a multidimensional information space; identifying stable patterns within this space requires intelligent analysis methods [8]-especially given that linguistic data are characterized by incompleteness, heterogeneity, noise, source variability, ambiguous cognate correspondences, and a dependence of results on the chosen models. The application of evolutionary computation-a class of intelligent computational methods based on the principles of natural evolution, population-based search, inheritance, mutation, recombination, and selection-is becoming increasingly relevant [9]. A key feature of these methods is their ability to search for solutions within complex spaces even when there is no simple analytical relationship between the input data and the desired result [10]. This is particularly significant for evolutionary linguistics [11-13], as reconstructing linguistic history often requires the simultaneous consideration of numerous interdependent variables and alternative historical scenarios.

In general, the use of evolutionary algorithms allows linguistic evolution to be viewed as a computationally modellable process in which changes in lexical units, phonetic features, and cognate relationships can be interpreted as sequences of transformations, while potential linguistic ancestors and branching patterns emerge as the results of a search for the most consistent evolutionary structures. While the analogy between the evolution of linguistic units and biological evolution does not imply complete identity, it provides a productive computational metaphor and a mathematical-algorithmic framework for investigating the historical dynamics of languages.

The relevance of this approach is further underscored by the dynamic nature of the contemporary linguistic research landscape [11-13]. Modern science operates amidst turbulent information flows, uncertainties regarding data origin and quality, the rapid expansion of digital corpora, shifting linguistic practices, the digitalization of communication, and the emergence of new forms of linguistic interaction. For this reason, methods in evolutionary linguistics must be capable not only of processing large volumes of data but also of adapting to incomplete, contradictory, and dynamic information environments.

Consequently, the aim of this publication is to systematically examine innovative approaches, current trends, and prospects for applying evolutionary computation to solve complex, multifaceted problems in evolutionary linguistics and glottochronology-specifically, reconstructing linguistic relationships, constructing phylogenetic models, analyzing cognate relations, and modeling language change over time amidst the turbulence of the linguistic environment.

**Main Part.** Evolutionary linguistics views language as a dynamic system that is self-organizing and historically evolving. Lexical units emerge and vanish, word meanings transform, phonetic and grammatical systems undergo restructuring, and language contacts drive borrowing and convergence. The current state of a language can be viewed as the outcome of a long sequence of interdependent changes. Reconstructing this dynamic involves identifying potential developmental trajectories leading from reconstructed ancestral states to observed linguistic forms.

The analysis of basic vocabulary and cognates is of particular importance to glottochronology, as it enables the establishment of historical correspondences between languages and the assessment of their degree of relatedness. However, actual linguistic evolution is far more complex than a model of uniform change accumulation. Rates of change vary across different lexical items, and results can be skewed by borrowing, language contact, areal processes, cultural factors, incomplete historical data, and ambiguities in cognate correspondences. Consequently, modern methods must account not only for the similarity of linguistic units but also for the complex structure of factors driving linguistic divergence.

Evolutionary computation offers new possibilities in this regard. It is based on population-based search, wherein a multitude of potential solutions undergo gradual modification through mechanisms such as selection, mutation, recombination, and quality assessment. In a linguistic context, an individual solution might represent a phylogenetic structure, a system of cognate correspondences, a reconstructed linguistic form, a set of model parameters, or a combination of several characteristics. Evolutionary search enables the exploration of a vast space of possible solutions, even in cases where obtaining a precise analytical solution is practically impossible.

The application of genetic algorithms to reconstruct phylogenetic relationships between languages is particularly promising. A candidate language tree can be evaluated based on how well it aligns with data regarding cognates, phonetic correspondences, lexical distances, and other features. Iteratively modifying the solution structure allows for the identification of language tree variants that are most consistent with the source data. However, directly applying biological models to language has limitations, as linguistic evolution involves not only vertical inheritance but also horizontal contact, borrowing, and convergence. Consequently, the use of hybrid tree-network models is emerging as a promising approach.

A distinct area of research focuses on the automatic identification of cognates. When analyzing large multilingual databases, manually establishing correspondences is labor-intensive and relies on expert interpretation. Evolutionary methods can be employed to find optimal combinations of features that distinguish potential cognates from non-cognates. This process can take into account phonetic similarity, the regularity of sound correspondences, morphological structure, semantic proximity, and other linguistic characteristics.

The potential of evolutionary computation is significantly enhanced when combined with machine learning, natural language processing, and big data technologies. Machine learning can facilitate feature extraction and classification; natural language processing enables the analysis of multilingual corpora and lexical structures; and evolutionary algorithms allow for the optimization of parameters, feature spaces, or reconstruction structures. As a result, a hybrid intelligent system is formed, combining the advantages of various computational paradigms.

On this basis, a multi-level model of computational glottochronology can be formed, which includes the collection and normalization of language data, the selection of linguistic features, the determination of probabilistic cognate correspondences, the evolutionary reconstruction of relations between languages, the evaluation of temporal parameters, and the verification of the stability of the results. This approach makes it possible to move from the use of separate algorithms to complex intellectual modeling of language evolution.

An important advantage of evolutionary methods is the possibility of solving multi-criteria problems. A real reconstruction can simultaneously take into account lexical similarity, regularity of phonetic correspondences, semantic stability, probable rate of change of lexemes, presence of borrowings, geographical connections and historical restrictions. Therefore, multi-criteria evolutionary optimization makes it possible to obtain not a single decision regarding one criterion, but many compromise options characterizing various possible scenarios of language evolution.

In the case of uncertainty, the assessment of the stability of the obtained solutions becomes important. If small changes in the original data cause a significant reconstruction of the language tree, the corresponding reconstruction should be interpreted with caution. On the contrary, the preservation of basic structures during the change of data, parameters or algorithms indicates greater stability of the result. Therefore, a promising computational glottochronology system should include sensitivity analysis, uncertainty assessment, re-simulation, and comparison of alternative reconstructions.

The modern linguistic environment is also characterized by the dynamism of the processes of language change. Migrations, social transformations, intensive language contacts, technological changes and the development of digital communication can significantly affect the speed and direction of language evolution. This creates prerequisites for the transition from static language trees to dynamic models capable of taking into account changes in language characteristics over time, non-stationarity of parameters and changes in the intensity of language contacts.

An important prospective direction is the modeling of turbulent language changes. Traditional models often assume a relatively stable environment for evolution, while the historical dynamics of languages can change dramatically under the influence of wars, migrations, cultural contacts, and other systemic events. From a computational point of view, this requires the use of models with changing parameters and the possibility of forming alternative development scenarios.

Systemic risks also arise directly in the process of computational modeling: due to incompleteness of the original base, transcription errors, incorrect identification of cognates, borrowings, or excessive dependence of the result on the selected mathematical model. Therefore, a reliable system should provide for data quality control, anomaly detection, uncertainty assessment, and robustness of the computational result.

The combination of evolutionary calculations with graph analysis, probabilistic modeling and expert knowledge of a linguist has a special perspective. In such an architecture, the evolutionary algorithm performs the function of search and optimization, machine learning methods - detection of regularities and classification, graph models - presentation of complex language relationships, and the expert ensures the formation of linguistic restrictions and interpretation of results. At the same time, automation does not replace a specialist, but expands his capabilities when working with large and complex data sets.

Additional possibilities are provided by genetic programming, which allows to evolve as parameters, but also the rules or structures of transformations themselves. For evolutionary linguistics, this creates prerequisites for the automatic search for formalized rules of linguistic changes, including regular phonetic correspondences. Evolutionary search can also be applied to the reconstruction of protolanguage forms, when it is necessary to determine a hypothetical ancestral state that is most consistent with a system of observed correspondences.

Glottochronological dating can also be reinterpreted through the lens of evolutionary optimization. Instead of assuming a constant rate of linguistic change, models with variable parameters-accounting for the specificities of individual lexemes, historical periods, and language contacts-can be employed. In this context, an evolutionary algorithm seeks parameter combinations that yield the best fit between the reconstructed dynamics and available historical data.

Thus, shifting from the "one algorithm, one result" principle to a "multiple models, scenario space" approach offers significant promise. Given the incompleteness and ambiguity of historical data, a set of possible structures-along with assessments of the stability of specific connections-can be more informative than a single reconstructed tree. This allows for a more accurate handling of uncertainty and helps avoid overly categorical historical-linguistic conclusions.

# PHILOLOGY AND LINGUISTICS

Adaptive algorithms, capable of adjusting search parameters based on the characteristics of specific linguistic data, will play a crucial role in the future development of evolutionary computation. Simultaneously, there is a growing need for interpretable evolutionary models, where computational results are accompanied by a description of the factors that shaped them. For the humanities, it is important not only to obtain an optimal solution but also to understand the linguistic logic behind its formation.

Looking ahead, it may be possible to create digital evolutionary-linguistic platforms that integrate historical corpora, dictionaries, cognate databases, phonetic and morphological characteristics, geographic data, and evolutionary optimization algorithms. Such systems will be capable of investigating not only the historical relationships between languages but also the potential trajectories of their evolution under various conditions. In this context, forecasting should be understood primarily as scenario modeling rather than as a precise prediction of the future state of a language.

Thus, the scientific potential of evolutionary computation lies not in the mechanical transfer of biological algorithms to linguistics, but in their adaptation to language as an open, socio-cognitive, and historically dynamic system. Amidst turbulence, uncertainty, and systemic risks, this entails a shift toward adaptive, multi-criteria, hybrid, scenario-based, and robust modeling capable of simultaneously accounting for the complexity of linguistic data and the uncertainty inherent in reconstructed historical processes.

Conclusions. The conceptual study conducted suggests that evolutionary computation represents a promising class of intelligent methods for addressing the complex challenges of modern evolutionary linguistics and glottochronology. Its application enables the computational reconstruction of linguistic relationships, the automated analysis of cognates, the construction and optimization of phylogenetic structures, the reconstruction of proto-language forms, the modeling of rates of linguistic change, and the investigation of alternative evolutionary trajectories.

It has been established that the specific nature of the linguistic object necessitates the adaptation of classical evolutionary algorithms. Language is not a simple system of vertical inheritance; its development involves borrowing, contact, convergence, varying rates of change across different elements, and the influence of social and cultural factors. Consequently, hybrid tree-network models, multi-objective optimization, and algorithms capable of accounting for the heterogeneity and non-stationarity of linguistic processes hold significant promise.

Hybrid intelligent systems-combining evolutionary computation with machine learning, natural language processing, Big Data, graph analysis, and probabilistic modeling-have been identified as having the greatest potential. Within such an architecture, evolutionary algorithms can handle search and optimization tasks; machine learning can perform pattern detection and classification; and expert linguists can define constraints, as well as interpret and verify the results.

The study substantiates the shift from seeking a single reconstruction to generating a multitude of alternative evolutionary scenarios. This approach is particularly relevant to glottochronology, as historical linguistic data are often incomplete and ambiguous. Evaluating a multitude of valid solutions makes it possible to examine the robustness of reconstructed relationships and to distinguish relatively stable patterns from results that are sensitive to initial assumptions.

It has been shown that under conditions of turbulence, uncertainty, and systemic risk, robustness becomes the most critical characteristic of a computational linguistic model. Therefore, a promising system should incorporate procedures for data quality assessment, anomaly detection, sensitivity analysis, the testing of alternative parameters, and re-modeling. This enables a significant increase in the reliability of interpreting the results of evolutionary reconstruction.

A promising avenue of research is the development of adaptive evolutionary algorithms capable of adjusting search parameters based on the structure of the linguistic data being analyzed. Concurrently, the importance of multi-objective evolutionary optimization is growing, as it allows for the simultaneous consideration of lexical, phonetic, morphological, semantic, geographic, and historical characteristics.

The development of interpretable evolutionary models-where reconstruction results are accompanied by a description of the factors that shaped them-holds particular scientific promise. This is crucial for evolutionary linguistics, as computational efficiency must be balanced with linguistic interpretability and the potential for expert verification. Overall, the future development of evolutionary computation within evolutionary linguistics and glottochronology should be linked to the emergence of an adaptive, hybrid, multi-criteria, scenario-based, and robust paradigm for the computational modeling of language evolution. Such a paradigm is capable of combining advances in evolutionary algorithms, artificial intelligence, and digital linguistics to enable a deeper investigation into the patterns of historical language change.

A promising outcome of future research could be the creation of an intelligent evolutionary-linguistic platform capable of integrating multilingual cognate databases, digital corpora, phonetic and morphological features, historical-geographic data, and evolutionary optimization algorithms. Such a platform would facilitate a shift from isolated computational experiments to the systemic modeling of language evolution, wherein the reconstruction of the linguistic past is conducted as a dynamic process of searching for, comparing, and verifying a multitude of possible historical-linguistic scenarios.

**References:**

[1] Smits, T., & Wevers, M. (2023). A multimodal turn in Digital Humanities: Using contrastive machine learning models to explore, enrich, and analyze digital visual historical collections. Digital Scholarship in the Humanities, 38(3), 1267-1280. https://doi.org/10.1093/llc/fqad008
[2] Goncharenko S. Intelligent technologies for modern sociolinguistics (2025). Science: Development and Factors its Influence : Proceedings of the 6th International Scientific and Practical Conference (October 6-8, 2025; Amsterdam, Netherlands). - Amsterdam: Scientific Collection «InterConf», 2025. - pp. 61-64.
[3] De Luca, E. W., Fallucchi, F., Ghattas, B., & Spielhaus, R. (2024). The digital transformation processes for supporting digital humanities researchers in text analysis. Journal of Documentation, 80(2), 378-391. https://doi.org/10.1108/JD-07-2022-0143
[4] Zhang, Z., Song, W., & Liu, P. (2024). Making and interpreting: Digital humanities as embodied action. Humanities and Social Sciences Communications, 11, Article 13. https://doi.org/10.1057/s41599-023-02548-3
[5] Grieve, J., Bartl, S., Fuoli, M., Grafmiller, J., Huang, W., Jawerbaum, A., Murakami, A., Perlman, M., Roemling, D., & Winter, B. (2024). The sociolinguistic foundations of language modeling. Frontiers in Artificial Intelligence, 7, 1472411. https://doi.org/10.3389/frai.2024.1472411

[6] Rettberg, J. W. (2022). Algorithmic failure as a humanities methodology: Machine learning's mispredictions identify rich cases for qualitative analysis. Big Data & Society, 9(2). https://doi.org/10.1177/20539517221131290

[7] Goncharenko S. (2025). BIG Data in modern literary studies. Science: Development and Factors its Influence : Proceedings of the 6th International Scientific and Practical Conference (October 6-8, 2025; Amsterdam, Netherlands). - Amsterdam: Scientific Collection «InterConf», 2025. - P. 82-85.

[8] Goncharenko S. (2025). Big semi-structured data & deep ANN in computational linguistics. Science and Global Challenges in the Modern World : Proceedings of the 2nd International Scientific Conference (Leicester, United Kingdom, 5 October 2025). - Lulu Press, Inc., 2025. - pp. 133-136.

[9] Krasnyuk M.T., Naumenko M.A. (2024) Efektyvne zastosuvannia henetychnykh alhorytmiv u vyrishenni bahatoekstremumnykh optymizatsiinykh zadach v menedzhmenti konkurentnoho pidpryiemstva [Effective application of genetic algorithms in solving multi-extremum optimization problems in the management of a competitive enterprise]. Grail of Science, №41, June 2024, pp. 65-73. URL: https://archive.journal-grail.science/index.php/2710-3056/issue/view/05.07.2024/29 [in Ukrainian].

[10] Kulynych Y., Krasnyuk M., Krasniuk S. (2022). Efficiency of evolutionary algorithms in solving optimization problems on the example of the fintech industry. Grail of Science, №14-15, May 2022, (pp. 63-70) DOI: https://doi.org/10.36074/grail-of-science.27.05.2022

[11] Bungum, L., & Gambäck, B. (2016). Evolutionary Algorithms in Natural Language Processing. In Proceedings of the Workshop on Natural Language Processing and Computational Linguistics. https://www.semanticscholar.org/paper/Evolutionary-Algorithms-in-Natural-Language-Bungum-Gamb%C3%A4ck/d2424a7f83be06b3abd99485ee22f8fd90cb602c

[12] Araujo, L. (2009). How evolutionary algorithms are applied to statistical natural language processing. Artificial Intelligence Review, 28(4), 275-303. https://doi.org/10.1007/s10462-009-9104-y

[13] Manzoni, L., Jakobovic, D., Mariot, L., Picek, S., & Castelli, M. (2020). Towards an evolutionary-based approach for natural language processing. arXiv preprint. arXiv:2004.13832. https://arxiv.org/abs/2004.13832