***Yan KAPRANOV, Doctor of Philological Science, Professor***
*Assistant Professor of the Faculty of Humanities and Fine Arts at the*
*University of Economics and Humanities in Warsaw, Poland*
*Professor of the Department of Theory and Practice of Translation from English*
*Kyiv National Linguistic University*
*(Kyiv, Ukraine) y.kapranov@vizja.pl*

## UTILIZING BAYESIAN NETWORKS TO MODEL
## THE NOSTRATIC PROTO-LANGUAGE

The study of historical linguistics, particularly in the realm of macro-comparative studies, has significantly evolved with the integration of computational methods. This paper aims to explore the application of Bayesian networks in modeling the Nostratic proto-language, a hypothetical language family that includes several major Eurasian language families. Our research investigates the degrees of relationship among Nostratic languages and the effectiveness of artificial intelligence (AI) tools in presenting these relationships.

**The primary objective of this study** is to demonstrate the efficacy of Bayesian networks and other computational tools in modeling linguistic relationships within the Nostratic language family. This includes identifying both close and distant degrees of relationship between languages and tracing their probable origins from the Nostratic proto-language.

Our methodology builds on the work of scholars like D. Bakker, Ch. Brown, and E. W. Holmal, who have successfully employed mathematical and computer methods to validate linguistic relationships. According to Hryhoriy Starostin, the key to such research lies in applying mathematical skills to linguistic data that cannot be verified by traditional methods.

We utilized a comprehensive research methodology, implemented in three stages:

*Quantitative Analysis of Linguistic Correspondences*: Using George Rauschenbach's mathematical formula ($S = b / (a + b)$), we calculated the phonological, morphonological, and semantic correspondences for the Nostratic etymon \*phaħ- (~ \*phəħ-) "fire; to heat". Here, 'a' represents unique indicators, while 'b' represents common indicators among languages.

*Phonological and Morphonological Analysis*: We employed Levenshtein distance as an electronic calculator to measure correspondences across different language families. To assess semantic shifts, we used Torvald Sorensen's formula ($D = 1 - 2q / (A + B)$), where 'q' denotes common meanings, and 'A + B' is the sum of all meanings for the etymon in the language families.

*Relationship Degree Analysis*: We established two degrees of relationship based on the coefficient of coincidences. A close relationship was defined for coefficients not exceeding 0.3%, while a distant relationship was defined for coefficients above 0.3%.

Our findings indicate that transformations in phonology, morphology, and semantics between the Afro-Asiatic and Indo-European language families suggest a distant relationship with a coefficient of 0.3%. Conversely, close relationships were observed between Afro-Asiatic and Kartvelian (0.15%) and between Indo-European and Kartvelian (0.26%).

Verification using Levenshtein's calculator revealed that a high number of recorded transformations (exceeding five morphological shifts) implies close linguistic relationships. Semantic development was also verified, confirming that higher coefficients indicate gradual development from the Nostratic archetype.

The application of Bayesian networks allowed us to construct a graph modeling conditional dependencies and causation among Nostratic language families. This innovative approach, combining linguistics with computational methodologies, provides new insights into the evolutionary pathways of languages.

Traditional linguistic methods are insufficient for determining genetic or areal relationships between language families. The integration of exact scientific methods and AI tools offers a more objective analysis of language development processes. Our research highlights the potential of Bayesian networks to enhance our understanding of historical linguistics and sets a foundation for future studies to delve deeper into the origins and relationships of world languages.

## REFERENCES

1. Bomhard, A. (2008). Reconstructing Proto-Nostratic (2 vols). Leiden Indo-European Etymological Dictionary Series. Internet Archive

2. Max Planck Institute for Evolutionary Anthropology. (2023). New insights into the origin of the Indo-European languages. Max-Planck-Gesellschaft