

Giuseppe Longobardi, Cristina Guardiano, Guido Cordoni,  
Andrea Ceolin and Dimitar Kazakov

## **Syntactic and Genetic Evolution of Human Populations: an Interdisciplinary Approach to Reconstructing Linguistic History**

The combination of linguistics and genetics provides powerful tools to probe human history. Whether their results converge or diverge, strong conclusions can be reached in the reconstruction of our past. We explore the possibility of broad-scale congruence between language and gene transmission by performing a new type of calculation of linguistic distances in Eurasia for a selected dataset of 37 languages from 11 traditionally irreducible linguistic phyla and the corresponding populations. To calculate linguistic distances, we adopt the Parametric Comparison Method (PCM, Longobardi and Guardiano 2009). To assess language relatedness, The PCM uses as taxonomic characters a class of items which had never been exploited before, that is, syntactic parameters of universal grammar. The syntactic database contains 91 binary parameters connected by an intricate network of implicational relationships, suggested by original theoretical and typological research. In spite of claims to the contrary, in both traditional and formal grammar (Newmeyer 2005, Lightfoot 2006, Anderson 2017 to cite just a few), we show that this dataset:

- (1) a. retrieves a deep historical signal in reconstructing all the major known splits into families and subfamilies in Eurasia;
- b. suggests statistically assessable hypotheses about deeper phyla which cannot be achieved and evaluated with lexical methods.

Then, we present evidence of a substantial overall gene-language correlation among the studied Eurasian populations and demonstrate that:

- (2) a. the evolution of syntax is predominantly vertical (i.e. tree-like, as a result of diachronic change, rather than through contact), and the remaining variations can be explained through demic movements as evidenced in the genetic record;
- b. when syntax-based taxonomies do not fully agree with independent (e.g. lexical-based) predictions, it is always the case that there is evidence of considerable genetic contact that explains the observed discrepancy;
- c. language-gene mismatches point to movements of languages without significant demic displacement, never the opposite.

On these grounds, we derive three conclusions about this interdisciplinary domain, which can be named *glotto-genetics*:

- (3) a. the correlation of genetic diversity and geographic distance is always higher than the one between linguistic diversity and geographic distance ( $Gen \sim Geo > Ling \sim Geo$ );
- b. the correlation of genetic diversity and syntactic diversity ( $Gen \sim Ling$ ), at the relevant long range and density of sampling, is moderate to strong, unless specific sources of mismatch apply;
- c. there has been long-distance gene replacement across Eurasia only if accompanied by language replacement, but not the contrary (languages can travel alone, genes cannot).