Andrea Ceolin (UniMoRE), Guido Cordoni (Surrey), Cristina Guardiano (UniMoRE), Monica Irimia (UniMoRE), Dimitar Kazakov (York), Giuseppe Longobardi (York) and Giuseppina Silvestri (UCLA)

Who likes to travel alone? Movements of populations and languages across Eurasia

1. Syntax and migrations. In this work, we explore the possible parallelism between migrations of populations and movements of languages through genome-language comparisons across Eurasia, relying on quantitative tools on both sides. To compare languages across multiple families, we use generative syntactic parameters. We show that such characters produce plausible phylogenies against established taxonomic knowledge and allow one to measure language distances. This permits to evaluate the congruence between linguistic and genetic distances across populations with a range that is not achievable through lexical data, and with results that are chronologically much deeper than those obtained from the comparison of phonemic inventories.

2. Materials and methods. To perform our experiments, we use a dataset of 94 binary parameters, which were set in 37 Eurasian languages belonging to 11 different linguistic phyla, i.e. at least 50% of the major stocks (and salient isolates) traditionally recognized in Eurasia (Europe, Western, Central and North-Eastern Asia): they are scattered at relatively high distance throughout a vast territory ranging from the Atlantic to the Pacific, so as to maximize the presumable historical depth of the diversity represented in the sample. First, we generate syntactic taxonomies through two different types of phylogenetic algorithms, distance-based and Bayesian character-based methods, which produce consistent results. Then, we correlate the syntactic distances with the genetic distances of 37 corresponding populations, calculated from about 300k SNPs. To perform the correlation, we use several Mantel tests: each correlation was tested using three different methods (Pearson, Spearman, and Kendall). On the whole, the three methods exhibit more than 95% correlation with each other and produce consistent results supporting the same conclusions.

3. Results. Through this approach, we have uncovered some significant broad-scale gene/language congruence, not accounted for simply in terms of geographic distance, which suggests that the grammar of most languages was often transmitted along with ancestral speakers' migrations. Indeed, syntactic history has appeared more geography-independent than gene transmission, correlating with geography prevailingly as a consequence of physical displacement of populations. Some local exceptions to the overall congruence have emerged, falling into differently motivated types, explained by plausible conditions on language diffusion and gene-language combinations.
Our results can be summed up as follows:

(1) a. the correlation between syntactic and genetic distances across the Eurasian populations is on average high, especially so at the two extremes of the macro-continent;
b. although the correlation between syntactic diversity and geographic distances appears at first sight generally significant, it becomes virtually null and/or non-significant, when one controls for the mediating effect of genetic diversity;
c. few cases of partial mismatch between traditional linguistic phylogenies and our syntax-based results (e.g. the positions of Bulgarian or Farsi). This situation suggests points of secondary contact effects: interestingly, in all such cases there is evidence of genetic admixture (and of some phonemic convergence) among the relevant populations. This supports Thomason and Kaufman's (1988) thesis that syntactic borrowing may indeed arise, though in contexts of particularly deep contact, such as strong demographic admixture.

4. Towards a glottogenetic history of Eurasia. The patterns we discovered can be summarized into five general principles, which seem to hold across the history of glottogenetic movements of the Old World:

(2) a. *Horizontality* principle: the correlation of genetic diversity with geography is higher than between syntactic diversity and geography ($D_{GEN}/D_{GEO} > D_{SYN}/D_{GEO}$).

b. *Congruence* principle: the correlation of genetic diversity and syntactic diversity at a continental scale is significantly high ($D_{SYN}/D_{GEN} > r=0.50$), unless one of the specific sources of mismatch identified under (1)c applies; about half of it (more, in high-congruence areas) is geography-independent.

c. *Unbalanced admixture* principle: in situations of long-standing population contact without full language replacement, the degree of resulting genetic admixture is higher than that of interference stemming from language contact (in stark contrast to the common-sense observation that migrations produce language replacement in the space of few generations).

d. *Independent Travel* principle: across the macro-continent there has been salient long-distance gene replacement only if accompanied by language replacement; language replacement, instead, has occasionally taken place independently of genes: *languages can travel alone, genes never do*.

5. Conclusions. Using syntactic parameters aims to mirror the step taken by genetic anthropology in shifting the burden of long-term evolutionary accounts from phenotypes to genotypes. The revealing picture of the glottogenetic history of Eurasia so obtained supports the idea that generative syntax provides powerful tools for cross-family historical comparison of languages and for the interdisciplinary study of divergence and convergence of human populations, thus promoting cognitive sciences to the frontline of historical research.

Map of the 37 languages/populations. BLACK=Basque (Bas); RED=Indo-European; LIGHT BLUE=Finno-Ugric; DARK BLUE= (hypothetical) Altaic; OLIVE GREEN=Semitic; PURPLE=Archi; DARK GREEN=Dravidian; ORANGE=Sinitic; GREY=Yukaghir; YELLOW=Japanese, BROWN=Korean.