

The history and geography of syntactic and genetic diversity

Giuseppe Longobardi

Linguistics and genetics provide powerful tools to probe the history of human populations. If their results are proven to eventually converge, very strong conclusions can be reached in reconstructing our past. Therefore, addressing linguistic variation in this perspective is one of the main challenges of modern theories of language. In the past few decades, indeed, molecular anthropology and its quantitative models have changed our understanding of the peopling of entire continents. However, it is still unclear how demographic and cultural processes (only the former leaving consequences at the genetic level) interact to shape human diversity in such wide and diverse areas. To understand this problem, I will address the possible parallelism between language and gene transmission on a broad scale, a question anticipated by Darwin's (1859) prediction of a global congruence of biological and linguistic variation. So far comparisons of linguistic and genetic diversity have been undermined in areal scope and/or quantitative resolution by their use of traditional linguistic classifications, based on lexical cognacy: the latter fail to safely establish comparison beyond relatively time-shallow language families, and tend to undergo non-discrete variation and environmental selection.. Therefore, a change in data-quality, as well as in statistical accuracy, is also necessary in linguistics, to produce wide-range demographic hypotheses within a comprehensive 'glosso-genetic' framework.

For this purpose, I will present a comparison of gene/language diversity in a sample of populations spanning from the Atlantic to the Bering Strait, we quantify linguistic relationships through the recent Parametric Comparison Method (PCM) which uses universally definable and discrete (binary) grammatical differences (syntactic parameters), rather than lexical items. With linguistic and genetic collaborators a database was built of such polymorphic grammatical loci (2100 parametric states, 75 each for 28 languages from 9 traditionally irreducible linguistic phyla), which deductively define several thousands of syntactic phenomena. Resulting language distances were compared with those defined through whole-genome autosomal SNP markers in 1303 individuals from corresponding populations (Figure 1). The effect of spatial relationships between populations was subsequently modeled through 5 different types of geographical measurement.

The populations/languages sampled are scattered on purpose across large territories, because we specifically aim to capture long-range/long-term trends of migration and cultural diffusion in Eurasia, rather than micro-areal and micro-temporal peculiarities. In principle, combinations of such events may distribute languages (grammars, in our study) and genes either in close correlation

(together, or at least along the same geographical routes) or in an unrelated way: in the latter case, either grammars could be culturally transmitted with unremarkable gene movements, or conversely ‘left behind’ by populations expanding into different territories.

Our experiments revealed an unprecedentedly high language/gene correlation (0.53); furthermore, when controlled for geography, it turned out 5 to 7 times higher (and statistically more significant) than correlations previously discovered using other linguistic variables (phonemic inventories). Thus, the results largely fulfill Darwin’s expectation, suggesting that, as a rule, grammars and genes have diffused together, with few constrained exceptions: in such cases language features have traveled without massive gene displacement, but never the contrary. Through this apparatus, the generative theory of grammatical variation can be used to draw actual historical and demographic conclusions: wide-scale diffusion of syntactic properties across much of the Old World appears to have occurred through robust demic migrations, with limited cases of elite dominance, but no major displacement of entire linguistically subdued/assimilated populations.

