

Is a 'History and Geography of Human Syntax' meaningful?

In addition to its theoretical impact, the development of molecular biology has brought about the possibility of extraordinary scientific progress in the study of historical classification and geographical distribution of different species and different human populations (cf. Cavalli Sforza, Menozzi, and Piazza 1994: *The History and Geography of Human Genes*).

We want to suggest that parametric theories of grammatical variation can prompt analogous progress in the study of the history and geographical distribution of different language families.

In fact, this work aims at unifying two traditionally unrelated lines of investigation:

- the formal study of syntactic variation in generative grammar
- the reconstruction of historical relations among languages (phylogenetic taxonomy)

The pursuit of this approach will be argued to seriously question the traditional belief in the orthogonality of grammatical typology and language diachrony/genealogy, enabling us to tentatively suggest a positive answer to the following problem, that we conventionally label 'Humboldt's problem':

Are the *typological* and the *genealogical* (phylogenetic) classifications of languages significantly isomorphous?

From the second half of the 19th century on, three different types of enterprises attempted to classify languages and/or populations separated for centuries or millennia into historically significant families:

A) for relatively shallow time depths and in particularly favored cases, the classical linguistic comparative method, based on inspection of the lexicon, can often provide sharp taxonomic conclusions, immune from the need of serious probabilistic evaluation, since the relatedness hypotheses are warranted by few patently highly improbable phenomena, most notably recurrent (optimally 'regular') sound correspondences. These properties largely solve the problem of a safe choice of *comparanda*, allowing for solid conclusions.

B) Beyond the classical one, the only other linguistic method so far proposed is Joseph Greenberg's (e.g. 1987, 2000) mass comparison; still based on the lexicon, it suggests more far-reaching but much less rigorous and widely acceptable taxonomic conclusions, because the very choice of the compared entities, based on pretheoretical resemblance among arrays of words, is less safe from chance. Obvious probabilistic questions, often of unmanageable complexity, arise and receive controversial answers.

C) A third comparative practice stems from a different discipline, population genetics (cf. Cavalli Sforza *et al.*, *op. cit.*): no question arises here about the comparability of the basic entities, since they are drawn from a *finite* and *universal* list of *biological* options: a blood group must be compared to the same blood group in another population, obviously, not to other sorts of genetic polymorphisms. The only issue concerns the statistical and empirical significance of the similarities discovered. This is why population genetics is considered so useful to complement linguistics in the task of classifying populations and languages.

We propose that the contribution of linguistics proper to such issues can be completely renovated on the grounds of parametric generative theories, which in principle allow one to bring modern cognitive science to bear on issues of cultural variation and historical explanation. Since parameters form a *finite* and *universal* list of discrete *biological* (though culturally set) options, they resemble the set of polymorphisms studied by population genetics and potentially enjoy similar (and perhaps even greater) formal advantages, overcoming in principle all questions on the choice of *comparanda* affecting linguistic methods based on the vocabulary. On the other side, the *a priori* probative value of parametric comparison is mathematically very high: e.g. just 30 binary independent parameters generate 2^{30} languages = 1,073,741,824. The probability for

two languages to coincide in the values of 30 independent parameters with binary equiprobable values = $1/2^{30}$, of three languages = $(1/2^{30})^2$, i.e. less than one in one billion billions.

We will test and exploit such a potential by establishing exact comparisons of parameter values among some languages whose degree of cognation is independently known, in order to prove the effectiveness of the method to provide historically correct taxonomic insights before applying it to controversial cases.

For the past fifteen years, a number of scholars have studied the parametric variation of the structure of nominal phrases in several languages. Relying on this work and on a number of original specific proposals and data collections, we have worked out a preliminary list of 41 binary parameters affecting DP-internal syntax and tested their values in over 20 ancient and contemporary varieties drawn from several Indoeuropean and non-Indoeuropean subfamilies (including Modern Italian, French, European Portuguese, Latin, Classical Greek, Modern Greek, Gothic, Old English, Modern English, German, Bulgarian, Serbo-Croat, Arabic, Hebrew, Hungarian, Wolof, among others). Each relevant parameter has been tentatively set for such languages, obtaining up to 41 precise correspondence sets of parameter values, and for every pair of languages we could arithmetically count identities and differences. In our formalism, the relative distance between any two languages is expressed by a coefficient, which consists of an ordered pair of positive integers $\langle i, d \rangle$, where i is the number of identities and d the number of differences. This procedure of lexically blind comparison, coupled with especially designed empirical and statistical methods, will be argued to prove adequate to generate the essentially correct phylogenetic tree, as resulting from traditional methods of lexical comparison. This is why the answer to Humboldt's problem above appears (perhaps surprisingly) largely positive.

The application of a parametric analysis to such issues of historical variation, in turn, yields a number of important insights for parameter theory itself from both a synchronic and a diachronic point of view.

The elaboration of a compact variation model for a whole syntactic subdomain has led to some general conclusions:

- we will show how parameters, even in a limited area of grammar, appear to be tightly interrelated and not freely orthogonal to each other (cf. Fodor 2000): some parameters become irrelevant (in different senses, to be technically distinguished) in a language due to the setting of other parameters (or to variation in the composition of the lexicon: Kayne 2003);
- parameters can be classified according to their relative *historical* weight: 1) certain parameters may understate or overstate identities and differences because they trigger a high number of irrelevant settings in the other parameters. Their diachronic resetting will be much more consequential (catastrophic in Lightfoot's 1999 sense) than that of others; 2) some parameters appear to be genealogically more significant than others because they are diachronically stabler. The possibility that this may be due to different degrees of exposure to external triggers and other modules at the interface, along the lines of Keenan's insight about Inertia, will be explored.

In the attempt to progress 'beyond explanatory adequacy' this line of investigation may thus lead toward new questions and a better understanding of classical ones in the cultural and natural history of language.

Chiara **Gianollo** (Universita' di Pisa), Cristina **Guardiano** (Universita' di Modena e Reggio Emilia), Giuseppe **Longobardi** (Universita' di Trieste)