Syntax as evidence for historical relatedness

Cristina Guardiano*, Giuseppe Longobardi°
*Università di Modena e Reggio Emilia, °Università di Trieste,
cristina.guardiano@unimore.it, longbard@units

Beyond its *theoretical* success, the development of molecular biology has brought about the possibility of extraordinary progress in the *historical* study of classification/distribution of different species and human populations (e.g. Cavalli Sforza, Menozzi and Piazza 1994).

Such developments have allowed human genetics to better contribute to the common effort of archeology, philology, historical-comparative linguistics toward the reconstruction of history and prehistory (the so-called 'New Synthesis'). A crucial component of this effort aims at reconstructing relationships among peoples, languages, and cultures.

We want to suggest that, even in the cognitive sciences, purely theoretical progress in a certain discipline may in principle have analogous historical impact and equally contribute to the New Synthesis in a renovated perspective. The present research aims at establishing that the Principles and Parameters models (Chomsky 1981) can prompt progress in the study of the history and of the geographical distribution of language families. To put the problem in context, notice that, from the second half of the 19th century on, three different types of enterprises attempted to classify languages or populations separated for centuries or millennia into historically significant families:

A) For relatively shallow time depths and in particularly favorable cases, the classical linguistic comparative method, based on lexical data, can often provide sharp taxonomic conclusions, immune from the need of serious probabilistic evaluation, since the relatedness hypotheses are warranted by agreement in patently highly improbable phenomena, most notably recurrent (optimally 'regular') sound correspondences.

B) The other linguistic method so far proposed is Joseph Greenberg's (e.g. 1987, 2000) mass comparison; still based on the vocabulary (indeed, on universal lists of meanings), it suggests chronologically more far-reaching (may apply to very distant languages) but much less acceptable conclusions, because the very choice of the compared words, based simply on pretheoretical resemblance in form and meaning, is not safe from chance.

Obvious, though hardly definable, probabilistic questions arise and receive controversial and generally skeptical answers (cf. Ringe 1992, 1996).

C) A third comparative practice stems from a different discipline, population genetics. Here no question arises 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. Theoretically, any pair of populations can be safely compared and their genetic distance can be assessed. 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.

Since parameters are assumed to form a *finite* and *universal* list of discrete options, they will formally 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 previous linguistic methods based on the vocabulary. On the other side, the *a priori* probative value of similarities in parameter settings across languages is mathematically very high: e.g., the probability for two languages to coincide in the values of 30 *independent* parameters with binary *equiprobable* values is $1/2^{30}$, of three languages is $(1/2^{30})^2$; but even under less idealized conditions (especially independence turns out to be unrealistic), such probability remains quite low, so that similarities in parameter settings between two languages may suggest a significant degree of historical relationship.

Moving from some preliminary rough explorations (Guardiano and Longobardi 2005), we can establish exact correspondence sets of parameter values among languages whose degree of cognacy is independently known, in order to prove the effectiveness of the method to provide correct taxonomic insights before trying to apply it to controversial cases of historical classification. In order to collect a suitable empirical database, we propose to adopt a research strategy termed 'Modularized Global Parametrization' in Longobardi (2003). We have worked out a preliminary list of 51 binary parameters affecting the single module of DP-internal syntax, and tentatively tested their values in 26 ancient and contemporary varieties drawn from several Indoeuropean and non-Indoeuropean subfamilies. Each relevant parameter has been tentatively set for

such languages, obtaining 51 precise correspondence sets; for every pair of languages we could arithmetically count identities and differences. This procedure of lexically blind comparison, coupled with some statistical methods and computational procedures adopted from genetics and evolutionary biology, will be argued to prove adequate to generate the (largely) correct phylogenetic tree, paving the way for its future application to cases where the actual taxonomy is controversial (the 'New Synthesis' issue).

The present research program may suggest that the vocabulary and the distribution of sounds within words and morphemes are not the only domain in which languages show significant traces of their history and ancestral relations: the study of abstract and deeper mental objects, like syntactic parameters, discovered and investigated with the tools of formal cognitive science, may also provide crucial contributions to the goals and methods of historical inquiry and warrant the very possibility of formal syntax as an historical scientific paradigm.

References

Cavalli Sforza, L. Luca, Paolo Menozzi and Alberto Piazza. 1994. *The History and Geography of Human Genes*. Princeton NJ: Princeton University Press.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Greenberg, Joseph. 1987. *Language in the Americas*. Stanford CA: Stanford University Press.

Greenberg, Joseph. 2000. *Indoeuropean and its Closest Relatives: the Eurasiatic Language Family*. Stanford CA: Stanford University Press.

Guardiano, Cristina and Longobardi, Giuseppe. 2005. "Parametric Comparison and Language Taxonomy". In *Grammaticalization and Parametric Variation*, M. Batllori, M. L. Hernanz, C. Picallo, and F. Roca (eds), 149-174. Oxford: Oxford University Press.

Longobardi, Giuseppe. 2003. "Methods in Parametric Linguistics and Cognitive History". *Linguistic Variation Yearbook* 3, 2003, 101-138.

Ringe, D. 1992. "On Calculating the Factor of Chance in Language Comparison". *Transactions of the American Philosophical Society* 82 (1): 1-110.

Ringe, D. 1996. "The mathematics of Amerind". *Diachronica* 13: 135-154.