The historical reality of biolinguistic diversity

Giuseppe Longobardi\textsuperscript{1}, Cristina Guardiano\textsuperscript{2},
Luca Bortolussi\textsuperscript{3}, Andrea Sgarro\textsuperscript{4}, Giuseppina Silvestri\textsuperscript{5}, Andrea Ceolin\textsuperscript{6}

\textsuperscript{1}Univ. York, \textsuperscript{2}Univ. Modena e Reggio Emilia, \textsuperscript{3}Univ. Trieste, \textsuperscript{4}Univ. Pisa

Goals. Arguing that the historical application of the biolinguistic model can complement molecular anthropology to model out a ‘grammatical anthropology’ as a new discipline at the crossroads of cognitive, biological and historical sciences. Background. In many respects, P&P models of UG are a conceptually plausible answer to the problem of explanatory adequacy (Chomsky 1964). However, empirically, parametric theories are not yet sufficiently corroborated, since nobody has so far indisputably assessed their effectiveness to the acquisition of grammatical diversity by implementing a parameter setting system over a realistic collection of parameters (Fodor 2001, Yang 2003; cf. Chomsky 1995:7). It is therefore debatable that a P&P model has actually attained substantial explanatory adequacy, though progressing beyond language-specific descriptive adequacy. To address the need for more solid arguments in favor of P&P, Longobardi (2003) suggested the opportunity of: i) adopting a Modularized Global Parametrization strategy, aiming at studying together relatively many (closely interacting) parameters in relatively many languages within the circumscribed domain of small modules of grammar; ii) beginning to aim at further testing grounds and levels of success, i.e. at satisfactory accounts of the actual distribution of grammatical diversity in time and space (historical adequacy). Methods. Elaborating on previous work (Longobardi/Guardiano 2009), a sample of more than 50 carefully identified binary parameters in DP-syntax, set in over 30 languages, is focused on; it is complemented with a set of hypotheses about UG constraints, defining two levels of deductive structure: one determines the traditional covariation of properties following from the same parameter, the other encodes an extraordinarily rich implicational hierarchy among parameters themselves (more pervasive than hinted in Baker 2001), largely responsible for hierarchies of size (e.g. Biberauer/Roberts 2012). Phylogenetic programs of biostatistical derivation have been applied to this database to formally measure syntactic diversity and generate hypotheses of phylogenetic trees and networks. Specific mathematical procedures (a sampling algorithm capable of dealing with the universal constraints imposed on parameter setting) have been elaborated on purpose, to compute the width of potential diversity allowed by this fragment of UG and to evaluate the significance of the one observed in the actual language set. First results. The distribution of actual syntactic distances is statistically highly significant. The results have been measured against independently known historical data (from comparative linguistics, history, genetics), with largely correct correlations: given a non-trivial set of languages, the description of their variation provided by the systematic parametric analysis of a whole compact domain quite exactly recapitulates their known history and relationships. The reality of a P&P model of the language faculty, therefore, receives strong and original support from its historical adequacy. Further testing. Recently, the use of structural traits (superficial grammatical patterns) has been advocated and tested for conclusions on language phylogenies, the status of universals, and the modeling of grammatical evolution (Dunn et al. 2011). Now, since parameters try to represent ‘abstract’ differences, often exhibiting a high degree of deductive depth with respect to surface contrasts, counting similarities in patterns rather than in parameter values could turn out to provide different outcomes when quantitatively assessing areal or genealogical relatedness. In order to test this idea, the same experiments above have been repeated using a choice of the surface descriptive patterns derived from the parameters, rather than the parameter values themselves (i.e. comparing E-languages rather than I-languages). Again, the results were plotted against the same independently known historical variables. This experiment allows one to empirically test the parameter-pattern controversy and Dunn et al.’s alleged conclusion that implicational universals and
UG are not supported by the extant distribution of structural diversity in the world’s languages. The first computations suggest that pattern-based phylogenies are by no means more significant or more revealing than those founded on abstract parameters and that the latter better represent actual historical linguistic relations. Further corroborating these results, we argue that biolinguistic models of diversity encode a higher level of reality than surface-oriented typologies.

**References**


Fig. 1: UPGMA Tree from 56 syntactic parameters

Fig. 2: UPGMA Tree from 113 corresponding surface patterns

---

1 **Legend**. Wo = Wolof; StB = Standard Basque; WB = Western Basque; Hu = Hungarian; Fin = Finnish; Ar = Arabic; Heb = Hebrew; Hi = Hindi; Ma = Marathi; Nor = Norwegian; Da = Danish; Ice = Icelandic; E = English; D = German; Wel = Welsh; Ir = Irish; Rus = Russian; Po = Polish; Slo = Slovenian; SC = Serbo-Croat; Blg = Bulgarian; Gri = Grico (Salentino Greek); BoG = Bovese Greek (Calabria, Grecaanico); Grk = Greek; It = Italian; Cal = Calabrese; Sal = Salentino; Sic = Sicilian; Ptg = Portuguese; Sp = Spanish; Fr = French; Rm = Rumanian.
Evidence from historical linguistics in support of two dominant hypotheses – the Aryan migration view and the out-of-India hypothesis – is presented and briefly assessed. In conclusion, it is observed that the current understanding in historical linguistics favours the Aryan migration view though the substratum view is questionable. Language families in the Indian subcontinent. (Source of the map used in this figure could not be determined; the map was obtained from the internet). The IA language family (Source: Masica C 1991 The Indo-Aryan Languages (CUP) p 449). The Dravidian language family Linguistic diversity has been defined in a broad sense as the range of variations exhibited by human languages (www.terralingua.org). The Ethnologue (Gordon, 2005, www.ethnologue.com) considers that there are 6,912 languages in the world today, but 2/45 SUS.DIV position paper research task 1.2 some of the languages included are just considered varieties or dialects in other accounts. The spread of English has been visualized in terms of three circles representing the historical and sociolinguistic profile of English in different parts of the world (Kachru, 1985). The inner circle includes the countries that are traditionally considered the bases of English, where English is the first language for the majority of the populations: UK, USA, Ireland, Canada, New Zealand, Australia. Linguistic diversity index (LDI) may refer to either Greenberg's (language) Diversity Index or the related Index of Linguistic Diversity (ILD) from Terralingua, which measures changes in the underlying LDI over time. Greenberg's Diversity Index (LDI) is the probability that two people selected from the population at random will have different mother tongues; it therefore ranges from 0 (everyone has the same mother tongue) to 1 (no two people have the same mother tongue). The ILD measures how the LDI