XLV Convegno della Società Italiana di Glottologia
Il cambiamento linguistico fra natura e cultura
Pisa, 21-23 ottobre 2021

The limits of syntactic change and the structure of the language faculty

Giuseppe Longobardi
University of York

(joint work with A. Ceolin, C. Guardiano, M-A. Irimia, D. Kazakov)

Work in formal syntax has very forcefully argued in favor of a cognitive faculty of language on the basis of the logical problem or paradox of language acquisition. Thus, this work has focused on what may be called the ontogenetics of grammars. Ceolin et al. (2020, 2021), using a parametric comparative method (PCM, Longobardi & Guardiano 2009) and building phylogenies from generative grammars, highlighted two surprising properties of historical syntactic variability: A) within established families, language phylogenies generated from syntactic parameters reproduce the classical lexical-etymological ones extremely well (contrary to a century-long tradition, unchallenged even in recent formal frameworks: see Anderson & Lightfoot 2004, a.o.); B) syntax also retrieves a statistically significant signal chronologically deeper than the lexical one, capable of evaluating controversial macro-families (e.g. Altaic, Ural-Altaic,…see Fig. 1). Such conclusions show that investigating variation in syntax has noticeable consequences for the study of human history. Here we argue that they are relevant to the theory of human mind as well, because they prompt a deeper question:

> *How come syntax encodes so strong a historical signal throughout long periods of transmission (cross-generational acquisition) in contrast to other linguistic levels?*

Call this problem the *paradox of syntactic history*. To appreciate its significance and puzzling status, consider some differences with the lexical-etymological signal, which roams through a domain of variation (Saussurean arbitrariness) poorly constrained, if at all, by a specific neurocognitive faculty (in the sense e.g. of Chomsky 1975, ch. 1):

1. first, as noted, the signal retrieved by lexical comparison appears 'shorter' in time (saturates earlier) than that of syntactic phylogenies (point B) above);
2. second, we will show that the typical rate of shared parameter values in two languages separated for two millennia is ~80%, much higher than the most optimistic retention rates of the original core vocabulary proposed in any classical lexico-statistic approach;
3. the retained common lexicon would often be unrecognizable were it not for the existence of regular sound correspondences, which are the product of phonological regularity imposed by general cognitive principles (even on variable rules in Labov's 1973 sense); in spite of no comparable regularity holding for syntax, the latter provides a better signal;
4. Points 1. and 2. hold true in spite of the fact that the lexicon is arguably subject to some pressure toward stability imposed by mutual understandability (Gilliéron 1918); diachronic parameter resetting is restricted though being less likely to be so functionally conditioned.

Hence, phylogenetic inquiry reveals the diachronic stability of syntactic parameters

XLV Convegno della Società Italiana di Glottologia
Il cambiamento linguistico fra natura e cultura
Pisa, 21-23 ottobre 2021

(essentially Keenan's 2009 *Inertia*), which proves stronger than that of lexical and phonological entities. This suggests that syntactic variability is both sufficiently **wide** to code for history to a detailed level of resolution and remarkably **constrained** to preserve its signal for a long time. In turn, it tells us something about cognition and learning, perhaps no less than the *paradox of language acquisition* and *poverty-of-stimulus*: namely, that there must be a universal cognitive faculty for syntax and also that it must have a specific structure.

Indeed, to go into detail, the observed situation is unexpected under at least two possible extreme hypotheses about the initial conditions of the mind: 1) given that syntactic variation is often assumed to be binary and finite or very limited (unlike lexical arbitrariness), if there were no language faculty (essentially as in Evans and Levinson 2009) and third factor conditions at all, or even if a rigid universal core (e.g. mapping of syntax to an invariant C-I interface: Chomsky 2014, Chomsky, Gallego, Ott 2019) were surrounded by completely unstructured variability, the syntactic signal would be no deeper and probably less informative/more chance-like than that of lexical variation, being heavily affected by homoplasy and back-mutations; 2) if, instead, there were only a rich but rigid and pervasive UG with little pre-defined room for variation (as perhaps in original '20 questions' models of parameters, Fodor 2001), there would not be enough resolution to obtain such detailed and correct phylogenies. Neither model would suffice to account for both the *persistence* and the *resolution* of the diachronic syntactic signal.

Then, a third model of the language faculty is necessary: along an optimized and probably universal mapping to C-I, even parts of the variable externalization system (mapping abstract representations to the S-M interface) must be governed by a large, flexible but constrained and interdependent, structure of UG and third factor. We show how it is precisely this tight interdependence of parametric characters that, once modeled and computationally measured, contributes to explaining the stability of the signal at least in two crucial respects: the lack of substantial homoplasy and the illusion of parallel evolutions.

In conclusion, the length of historical signal seems characteristic of systems of knowledge guided by a structured cognitive faculty, though flexible to setting information from the environment. Thus, through the paradox of syntactic history, the *phylogenetics* of grammars is of no less relevance to cognitive science than the *ontogenetics* of grammars so central in the generative bio-linguistic tradition.

Fig. 1 (from Ceolin et al. 2021) – A UPGMA tree calculated using our syntactic distances. The groups of languages identified by the phylogenetic algorithm as low nodes (which are then represented each as a unique node) have been used simply as an input to a purpose-designed statistical test and are colored in blue, while the branches which have been shown to be related as the output of the test are in green. The relationships in these green branches can be considered 'validated' by both the statistical test and the UPGMA clustering algorithm. They include several nodes that would be unprovable and highly controversial families according to standard etymological methods. In black the clusters not supported by this test.

XLV Convegno della Società Italiana di Glottologia
Il cambiamento linguistico fra natura e cultura
Pisa, 21-23 ottobre 2021

Wolof
Mandarin
Cantonese
Japanese
Korean
Basque_Central
Basque_Western
Malagasy
Yukaghir
Hungarian
Khanty
Estonian
Finnish
Mari
Udmurt
Buryat
Evenki
Even_1
Even_2
Uzbek
Yakut
Turkish
Kazakh
Kyrghiz
Arabic
Hebrew
Tamil
Telugu
Archi
Lak
Irish
Welsh
Pashto
Marathi
Hindi
Romanian
French
Spanish
Portuguese
Salentino
Siciliano
Italian
Greek_Salento
Greek_Calabria
Greek
Greek_Cypriot
Bulgarian
Polish
Russian
Serbo-Croatian
Slovenian
Afrikaans
Dutch
German
Icelandic
English
Danish
Faroese
Norwegian