Measuring syntactic diversity in Southern Italy: a microparametric approach

Introduction. This paper argues for the importance of syntactic comparison for the contrastive analysis and classification of languages closely intertwined geographically, genealogically and sociolinguistically. All such cases are well represented in Southern Italy, within Romance and between Romance and Greek. In this area, we collected syntactic data from the nominal domain in 8 Romance varieties (three in Sicily: Ragusa, Mussomeli and the Gallo-Italic dialect of Aidone; two in Calabria: Reggio Calabria and Verbicaro; one in Salento and one in Campania), and two Greek ones (Salento Greek and Southern Calabria Greek). To better delineate the internal structure of the two language groups, we extended our comparisons to a set of "standard" Romance languages (Italian, French, Spanish, Portuguese and Rumanian), a set of Greek varieties (Standard Modern Greek, two varieties of Cappadocian Greek, Romeyka Pontic, Cypriot), and their ancestors (Latin and two varieties of Ancient Greek).

Methods. Our analytic tool is the Parametric Comparison Method (PCM, Longobardi and Guardiano 2009), that opens the possibility of combining insights from formal grammar, historical-comparative linguistics, quantitative sociolinguistics (dialectometry), and computer-assisted techniques to study vertical and horizontal transmission of languages (Guardiano et al 2015).

Questions. The goal of this paper is to show that the PCM can be successfully implemented in order to answer the following research questions: (1) Does dialect syntax retain a significant historical signal in contexts of massive horizontal transmission (robust amount of secondary convergence)? (2) Does dialect syntax retain a horizontal signal? (3) Does the distribution of syntactic diversity correlate with geographical factors on the micro-comparative scale?

Data and experiments. We analyzed the internal structure of the nominal domain in the varieties listed above using Longobardi et al's (2015) list of 75 binary parameters. To achieve higher resolution, especially within Southern Italy Romance, we explored the Romance varieties of our sample under a finer-grained perspective. This allowed us to single out and parametrize aspects of variation (in nominal syntax) specific to them. This task required the investigation of 8 additional newly devised micro-parameters (parameters associated with certain intensionally definable classes of lexical items, in the sense of Roberts 2012). The new parameters were formulated aiming at the most constrained theory possible, i.e. the lowest number of characters that can cover all the unveiled surface manifestations.

Taxonomic results. After computing parametric distances, we built syntactic trees and networks and performed various statistical experiments. Our results show that parametric information alone is a good predictor of the genealogical histories of Greek and Romance in Southern Italy, in spite of some detectable but controllable amount of secondary convergence. The experiments also reveal that language contact did not have strong enough impact to obscure the genealogical relations among the varieties examined: the Romance group is systematically classified separately from the Greek one (even when the ancient languages are included). Having established that parametric syntax can detect historical processes of diversification/continuity/(or even) convergence, it remains to be seen if the distances computed by the system can be translated into a phylogeny successfully distinguishing vertical from horizontal processes. Furthermore, the internal classification of the two groups resembles the dialectal distribution of the varieties: the Italo-Romance varieties of Sicily are grouped together and the Extreme Southern Unity is successfully recognized. Unsurprisingly, the two Upper Southern varieties do not fall together: Campano clusters with Italian, while Northern Calabrese is the outlier of the whole Italo-Romance group, outside the cluster of Extreme and Upper Southern varieties. The outlier status of Nothern Calabrese is presumably due to its isolation: it is the only representative

in our sample of the "Lausberg area", which shares various features with other plausible outliers of the Italo-Romance group.

Patterns of horizontal transmission. The distribution of parameter values shows that some particular subdomains are particularly sensitive to areal constraints, while others are more homogeneous diachronically (Guardiano 2014). On the whole, there are two kinds of restrictions on syntactic borrowing. First, given the sociolinguistic setting of the area and the asymmetries between Romance and Greek, parametric contact happens to be monodirectional, from Romance into Greek (Guardiano et al 2015). Second, contact-induced syntactic change is clearly constrained by pre-existing patterns in the language: resetting of parameter α from value X to Y in language A as triggered by interference of language B only takes place if a subset of the strings that contribute to constituting a trigger for value Y of parameter α in language B already exists in language A (Guardiano et al's (2015) 'Resistance Principle'; see also Sitaridou 2014: 52). For instance, the subdomain of adjectival modification in Southern Italy Greek and Southern Italy Romance has been identified as a point of lesser resistance to syntactic borrowing (Guardiano 2014, Guardiano and Stavrou 2014). Indeed, in accordance with Guardiano et al's 'Resistance Principle', it can be shown that the parameter resettings which gave rise to identical parameter values in this subdomain result precisely from the availability, in both groups, of superficially postnominal adjectives, though with distinct underlying representations.

Conservatism and diversification. The distribution of syntactic distances within Greek and Romance, especially when compared to their ancestors, shows that the Greek varieties display more salient internal diversification than Romance. On the other hand, they provide a stronger impression of closeness (conservatism) with respect to their plausible direct ancestor (*koiné* Greek) than Romance with respect to Latin. These two partly contradictory conclusions (diversity and conservatism) can be reconciled by noting that indeed both diversity and innovativeness especially (or exclusively) concern the Italiot varieties most affected by Romance interference and Romeyka (affected by Turkic interference). Both these dialect groups share not only contact with non-Greek languages but also isolation from "Mainland" Greek (thus not sharing its innovations). It is further possible to single out the parameters responsible for the remarkable evolution of the Romance nominal syntax from the Latin one (along the lines suggested in Longobardi 2012), as well as those most responsible for the diversification among the Greek varieties.

Conclusions. Our results provide positive answers to all our three research questions: (1) Historical relatedness is successfully reflected in parametric distances and the resulting clusterings in tree/network representations. Also, the need for finer-grained parameters for higher resolution suggests that not all parameters have the same stability, and therefore different parametric changes can non-accidentally trace splits of different historical depth, i.e. that parameters do carry a historical signal. (2) It is plausible to identify selective subdomains of syntax that are more sensitive to contact-induced parameter resetting or, vice-versa, more impermeable to the pressure of contact. However, our crucial conjecture is that this must be always relativized to the general parametric layout of the languages in contact. (3) Geographical factors are reflected in syntactic taxonomies. In particular, physical barriers seem to play a significant role in syntactic transmission. The cases of presumable geographical influence seem not to affect the overall syntactic phylogeny. We conclude that the PCM can not only be used for the purpose of long-range historical comparisons, but also be successfully implemented as a device to describe and explain microvariation at a local and chronologically shallower level.