DOI: https://doi.org/10.54103/1972-9901/29336

Diego Pescarini

Quantitative modelling and syntactic microvariation

ABSTRACT: This paper argues that parametric theories should account for statistically significant associations between syntactic variables that are logically and historically independent. Associations, which do not necessarily imply causation, suggest that syntactic variation is limited by abstract constraints (Parameters), which are not categorical rules. Counterexamples, which are always found in microvariation contexts, do not necessarily lead to the refutation of solid associations and, consequently, to the refutation of parametric models. It is argued that parametric models should be sufficiently plastic to adapt to unexpected patterns that are learnt from experience.

KEYWORDS: Syntax, Italo-Romance, Parameters, Microvariation, Negation.

> There is no algorithm to determine when apparently disconfirming evidence is real or is the effect of unknown factors, hence to be held in abeyance. (Noam Chomsky, *Derivation by Phase*)

1. Introduction

The syntactic analysis of closely-related languages (or dialects) within the generative paradigm has been thriving since the early 80s. Microvariation data have been approached from two complementary approaches, dubbed the *microcomparative* and the *microparametric* approach, respectively. The two overlap, but their objectives and their methodologies diverge slightly.

The microcomparative approach aims to fine-tune syntactic theory. This stream of research was inaugurated by Richard Kayne's seminal works (see Kayne 1989 for an example; Kayne 1996 for an overview) and, in relation to Italo-Romance data, it has been pursued by Italian scholars such as Cecilia Poletto, Maria Rita Manzini, and Leonardo Savoia (see references below). In these works, data from less-known

linguistic varieties allow us to endorse or discard analyses of "major" languages such as Italian or French. Methodologically, microcomparison focuses on a narrow sample of dialects to clarify issues that are relevant (viz. problematic) for current syntactic theorizing. The data that undergo analysis are normally selected from large datasets, which are collected through extensive fieldwork. Since the comparison is across languages that differ minimally, scholars can pin down a single phenomenon of interest and compare its manifestations in various languages all other conditions being equal.

The microparametric approach has a different intent. It does not focus on the representation of specific syntactic phenomena, but aims to model syntactic variation as a whole. The main intent is to understand whether and how variation is constrained by general principles, which guide/favor language acquisition by reducing the number of possible grammars that human beings can acquire. In the microparametric approach, syntactic theory is the tool that allows scholars to decompose variation into its minimal bits, which are eventually organized into hierarchies or matrixes of variables. Adam Ledgeway and Cristina Guardiano (and collaborators) have carried out microparametric analyses of Italo-Romance data (more on this in section 2). In Guardiano's works, in particular, linguistic phenomena are represented by numerical variables, which in turn undergo statistical analysis. Given the objective of microcomparative syntax, sampling is exhaustive or at least statistically significant and the empirical domain under investigation includes phenomena that, from a theoretical standpoint, are not particularly intriguing.

The present contribution elaborates on the role of microvariation data vis-à-vis current microparametric models (see section 2 for an overview). I will argue that statistically significant associations (section 3) and clusters of associations between syntactic variables that are neither logically nor historically related (section 4) provide an ideal test bed for parametric models. Section 5 concludes.

2. State of the art

In the early 80s, Chomskyan linguists proposed that the Faculty of Language comprises a series of invariable Principles and a built-in switchboard of Parameters, allowing languages to vary according to a finite set of choices, e.g. null *vs* non-null subject languages. Given its rigidity, the original model was soon surpassed by alternative views, culminating in the hypothesis (successively dubbed *Borer-Chomsky Conjecture* by Baker 2008) that variation results from the properties (technically: features) of functional elements such as Determiners, Tense markers, and Complementizers. As Borer (1984: 29) puts it, "[a]ssociating parameter values with lexical entries reduces them to the one part of a language which clearly must be learned anyway: the lexicon", whereas the syntactic/computational algorithm remains inert to change and variation (Longobardi 2001).

According to the Borer-Chomsky Conjecture, the innate nature of (micro)parameters is ultimately questioned (Boeckx, Leivada 2013). If parameters are learnt along with function words, no higher-grade universal constraint is expected to affect the

idiosyncratic featural endowment of lexical items. Variation therefore can neither result from nor be constrained by our genetic endowment, which is "apparently nearly uniform for the species" (Chomsky 2005: 6). Instead, it is Experience that "*leads to variation*, within a fairly narrow range, as in the case of other subsystems of the human capacity and the organism generally" (*ibid.*, emphasis mine).

Longobardi (2018) argues for a less radical solution, suggesting that the *format* of parameters can be reduced to a finite number of *schemata*. In a nutshell, parameters are properties of functional heads (in compliance with the Borer-Chomsky Conjecture), parameters are learnt from experience (hence, they are set according to experience), but their format, i.e. the range of properties they affect, is uniform across languages. Learning a language amounts to acquiring, for each functional head/feature, a given set of instructions, e.g. whether a certain feature is grammaticalized, whether it is involved in agreement, whether it attracts a constituent (e.g. a wh element), etc. Because of their format, parameters are logically interlocked as the relevance of certain parametric choices (e.g. X agrees with Y) depends on other parametric choices (e.g. X and Y are both grammaticalized).

Biberauer, Holmberg, Roberts, Sheehan (2010 and following publications) argue for a stronger theory, reviving the spirit of the original Principles and Parameter framework that was elaborated in the early 80s. More specifically, they entertain the hypothesis that languages are shaped by various kinds of parameters, ranging from macroparameters, which uniformly apply to multiple functional heads, to micro-parameters affecting subsets or single functional elements. The more a parametric choice is embedded in the hierarchy (see Fig. 1), the less it will impact on the syntax of a language. Macroparameters are then construed as hierarchies of lower-grade meso/micro/nano-parameters; the latter correspond to parameters in Borer-Chomsky's sense.

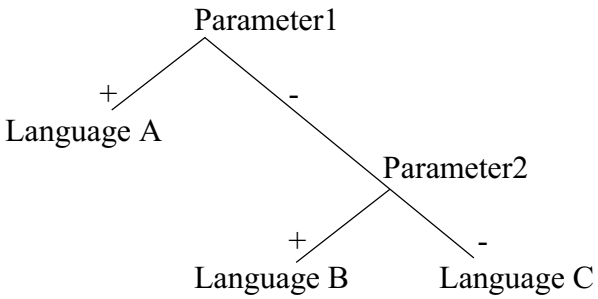

Fig. 1 Parametric hierarchy

The hallmark of parametric hierarchies is the existence of subset/superset relations in the typology of languages. Variation in the distribution of null subjects is a case in point. The dichotomy null *vs* non-null subject languages, as it was proposed in the early 80s, is too idealized. In fact, several languages allow null subjects, but only in certain contexts. Biberauer et al. (2010) argue that the null-subject parameter is better

conceived as a hierarchy like the one in Fig 1, where lower-grade factors (e.g. person) play a role in the licencing of null subjects, but only if null subjects are allowed by a higher-grade parameter. The various kinds of partial pro-drop languages belong to concentric subsets, resulting from the setting of intermediate parametric choices.

Similarly, hierarchies have been proposed to account for the make-up of inventories of functional elements, which are often implicationally structured (e.g. "if a language exhibit the element X, then it will exhibit the element Y"). Such entailments do not follow straightforwardly from general "cognitive" conditions, but from hierarchies of grammatical features organized *iuxta propria principia* (see for instance Harley, Ritter's 2002 geometry of person and number features).

2.1. The view from Italo-Romance

The parametric models introduced so far differ in several respects. They diverge with respect to the locus of parameters (functional items/heads *vs* more abstract parametric choices), but, most importantly, they differ in whether and how parameters constrain syntactic variation. Borer-Chomsky's model is the most liberal. In principle, anything goes, unless the innate and very general principles of the language faculty are violated. Other parametric models such as Roberts's or Longobardi's are more restrictive. They aim to constrain variation via parametric hierarchies, which rule out certain parametric values in the context of others, or schemata, which constrain *a priori* the properties encoded by functional heads.

The analysis of genealogically-related languages such as Italo-Romance dialects has brought fresh evidence to fuel the debate on the nature and existence of parameters, on their format, on their hierarchical organization, on the relationship between parametric hierarchies and linguistic phylogeny. Evidence from microvariation, however, does not seem conclusive. In fact, alternative and competing models of parametric variation have been endorsed on the basis of evidence from Italo-Romance varieties.

Maria Rita Manzini and Leonardo Savoia's works (e.g. Manzini, Savoia 2005) endorse the Borer-Chomsky Conjecture. For instance, they claim that the Null Subject Parameter in northern Italo-Romance 'cannot be defined for the entire language, but must be applied to the individual forms of the paradigm' (Manzini, Savoia 2005, I, 120; translation in Roberts 2014:178). By contrast, Ian Roberts's works on subject clitics (Roberts 2010, 2014) argue against this radical microparametric (or anti-parametric) approach, which would 'make the number of possible grammatical systems hyperastronomical'. Feature hierarchies might provide an intermediate explanation, a view that is endorsed in Adam Ledgeways's seminal articles on microvariation and microcontact in the Italo-Romance domain (e.g. Ledgeway 2019 on auxiliaries, Ledgeway 2020 on demonstratives, Ledgeway, Schifano, and Silvestri 2019 on differential object marking).

Poletto (2013) acknowledges the fact that morphologically inventories may be hierarchically shaped (see Benincà, Poletto 2005), but rejects the hypothesis that parametric hierarchies may explain bona fide syntactic variation across languages.

More specifically, she argues against the idea that there are multiple orders of parameters and that variation across dialects is imputed to lower-grade parameters. By the same token, similarities do not necessarily result from higher-grade parameters, which are more "stable" across time, thus preventing geographically or genealogically related languages from drifting in random directions. Instead, microvariation and typological variation across linguistic families are made from the same substance: features of functional elements (as in the Borer-Chomsky Conjecture). Departing from Borer (1984), however, Poletto assumes that functional features are not carried by "lexical entries" – in this respect she does not follow Manzini, Savoia's lexicalist account. Instead, Poletto claims that variation is encoded in the functional spine of the clause, which in turn is formed by sequencing multiple functional projections (as Poletto 2013 puts it, "a class of functional projections, all containing the primitive component x, behaves alike with respect to x").

Besides Poletto, the heuristic value of parametric hierarchies is questioned in Cristina Guardiano's works on microvariation, primarily based on Southern Italian and Italiot Greek data (Guardiano et al. 2016, 2018; Crisma, Guardiano, Longobardi 2020). Adopting the Parametric Comparative Method (PCM, Longobardi and Guardiano 2009), Guardiano conducts a cluster analysis of Italo-Romance varieties relying exclusively on syntactic evidence from the noun phrase. Syntactic phenomena are classified according to a standardized set of diagnostics (summarized in Longobardi and Guardiano's 2009 Table A). Crucially, the clusters of dialects issued from the PCM analysis correspond to the traditional classification of Italo-Romance dialects that was independently established by dialectologists such as Pellegrini (1977).

In conclusion, microvariation data lend themselves to support alternative and competing models of parametric variation. Scholars do not agree on the kind of evidence against which previous models can be tested and evaluated. In the following sections, I suggest a new way to test parametric models by focusing on correlations between descriptive variables across genealogically-related languages.

3. Significant associations

The Romance languages exhibit several shared innovations due to the emergence of new (classes of) functional elements such as articles, clitic pronouns, subordinators, perfective auxiliaries, negative words. The syntax of these elements across the Romance languages is subject to *microvariation*, i.e. these shared innovations give rise to a multitude of language-specific and context-specific patterns. For instance, in certain Italo-Romance dialects the choice between *have* and *be* auxiliaries is person-driven (e.g. 1/2p selects *be*, 3p selects *have*), in others the auxiliary split reflects the subdivision of verbs into classes (unaccusatives, various types of reflexives, etc.), other languages mix both criteria. Furthermore, auxiliary selection is sensitive to several other syntactic factors such as tense/mood/aspect, and lexical semantic factors such as agentivity or *aktionsart*, thereby resulting in a huge spectrum of alternative possibilities (for an overview, see Loporcaro, Pescarini 2022). Despite this kaleidoscopic range of

variation, grammatical systems follow some regularities that are independent from historical or socio-cultural circumstances.

Some of these regularities have the form of correlations between *variables* that, at first sight, are not logically related. With the term variable, I refer to a linguistic feature that is "high in frequency", has "a certain immunity from conscious suppression", and is "easily quantified" (Labov 1966/1982: 49). Before introducing some examples, it is worth noting that variables are not parameters. In Longobardi's (2018) terms, the former are *manifestations* of parameters: in principle, one manifestation may result from the interaction of multiple parameters and, *vice versa*, multiple manifestations can be associated to a single parameter. Associated variables[1] are clues to unveil the complex structure of parametric systems, i.e. the relationships between parameters and variables/manifestations. With the term *associated variables*, I refer to pairs or clusters of variables (see section 4) that tend to co-occur in a significant sample of languages, although they are not trivially/logically related. Some examples from Romance are provided in (a)-(d):

a) the association between the morphosyntax of imperatives and the morphosyntax of negation (Zanuttini 1997): verbal forms that are unique to the paradigm of the imperative (for short: true imperatives) cannot co-occur with preverbal negative markers. In negative imperative clauses featuring a preverbal negation marker, the verb is either a subjunctive or an infinitive form (for short: it is a *fake* imperative).

(1) a. Fr. *mange !* → *ne mange pas !*
b. It. *mangia!* → **non mangia!* / *non mangiare!*
'eat!' → 'do not eat!'

b) the association between null subjects and rich inflectional system (Taraldsen 1980). Languages with rich inflection allow null subjects more readily than languages where inflection is often syncretic, e.g.

(2) a. It. (ˈio) ˈmaŋʤ**o**, (tu) ˈmaŋʤ**i**, (lui) ˈmaŋʤ**a**
b. Fr. *(ʒə) mãʒ, *(ty) mãʒ, *(il) mãʒ̂
'I eat, you eat, he eats'

c) the association between clitic doubling and Differential Object Marking. Doubling of objects is normally allowed in languages where arguments are introduced

1. The terms *association* and *correlation* are often used interchangeably, although they have slightly different meanings. Correlation is a measure of the linear association between variables, while association refers to any relationship between two variables. Since syntactic variables are usually binary (or dichotomous), the term association is more appropriate.

by a preposition-like particle (this association was dubbed "Kayne's generalization" by Jeaeggli 1982), e.g.

(3) a. Sp. *(lo) vi **a** tu* papa
b. It. *(*lo) vidi tuo papà*
'I saw your dad'

d) the association between number morphology and bare plurals (Delfitto and Schroten 1991; Guardiano, Stalfieri, Cambria 2022): bare nouns (i.e. determiner-less noun phrases) are not allowed in languages with impoverished number morphology on nouns, e.g.

(4) a. Fr. *j'ai mangé *(des) pommes* [pɔ̃(m̩)] *de terre*
b. It. *ho mangiato (delle) patat**e***

Possible correlations were originally found by comparing major Romance languages such as Italian, French, and Spanish in light of general theoretical principles (e.g. the Null Subject Parameter for (2) or Case theory for (3)), but very few assumptions have subsequently been tested against big datasets or following a standardized methodology to obtain replicable results. However, in the last decades generations of linguists have built on-line linguistic resources that now enable us to test empirical generalizations such as (a)-(d) on significant samples of languages.

Quantitative testing on both typological and microvariation data are opening new perspectives on parametric modelling. The remainder of this section aims to illustrate a possible avenue of research by focusing on the association in (a), i.e. the association between the morphosyntax of imperatives and the morphosyntax of negation.

If we adopt a broader typological view, the interaction between negation and imperatives reduces to two main variables (van der Auwera, Lejeune, Goussev 2013):

- negative imperative clauses feature a true imperative, i.e. a form that corresponds to the positive imperative, or a fake imperative;
- negative imperatives feature a default negation, i.e. a form of negation that is found elsewhere (e.g. in declaratives), or a dedicated negative marker.

The latter variable is seldom at play in Romance. Most Romance varieties do not feature a dedicated negation marker in imperative clauses. Poletto and Zanuttini (2003) notice that certain Dolomitic Ladin varieties exhibit true imperative forms and a dedicated negative marker, as shown in (5a), which does not correspond to the default clausal negation marker found in declaratives such as (5b), from Poletto (2016: 841).

(5) a. No (ma) l lì (San Leonardo di Badia)
Not (prt) it= read.IMP
'Don't read it!'
b. Maria ne végn nia a ciasa.
Maria not comes not to home
'Maria isn't coming home.'

The data from the WALS (World Atlas of Language Structures; Dryer, Matthew, Haspelmath 2013) summarized in Table 1 show that the pattern in (5), which is marginal in Romance, is in fact the most widespread across the languages of the world.[2]

	True Imp	**Fake Imp**
Default Neg	113	55
Dedicated Neg	182	146

Table 1. Number of languages exhibiting fake/true imperatives and default/dedicated negative markers. Dataset: WALS (van der Auwera, Lejeune, Goussev 2013).

Moreover, the data in Table 2 show that fake imperatives are less likely to occur in languages that do not display a dedicated negation marker for imperatives, a conclusion that seems to contradict the generalization in (a). Notice, however, that an important variable is missing in Table 1: we do not know the position of negative markers with respect to the inflected verb for the 168 languages in the WALS dataset that exhibit true or fake imperatives with the default negation marker. We can find this information in another chapter of WALS (Dryer 2013), which deals with the order of negative morphemes and verbs (for the sake of simplicity, I have reduced Dryer's typology to two basic orders, V > Neg and Neg > V regardless of the affixal/clitic nature of the negative morpheme). By crossing the data of Fake/True imperatives in the languages that exhibit no dedicated negative morpheme in imperative clauses, with Dryer's (2013) data on the position of negation in negative clauses, the data show no significant association between the morphology of imperatives and the position of negation. Since the chi-square statistic is smaller than the critical value (.05), there is no significant association between these two variables.

	True Imp	**Fake Imp**
No PreV Neg	30	7
PreV Neg	42	22

Table 2. Number of languages exhibiting fake/true imperatives with/without preverbal default negation. Dataset: WALS (Dryer 2013; van der Auwera, Lejeune, Goussev 2013)

The data show that, crosslinguistically, fake imperatives are not the prevalent pattern neither in languages featuring preverbal default negators. Analogously, there is no significant association between the nature of the negator marker (clitic/affixal vs word-like) and the incidence of fake imperatives. Therefore, the generalization in (a) does not hold when examined against a broader sample (assuming that the descriptive concepts we chose are sufficiently accurate—they are not).

2. Nothing hinges on the fact that (5) is – or is not – the most widespread pattern. The statistical reliability of the WALS dataset should not be taken for granted, as an anonymous reviewer pointed out.

If we focus on northern Italo-Romance, however, the correlation between fake imperatives and the nature/position of clausal negators is robust. Fig. 2 shows the incidence of suppletive imperatives in a sample of 180 northern Italian dialects from Manzini, Savoia (2005) (data from the AIS dataset support the same conclusion, see Pescarini 2023: 239). Five type of negation systems are represented in the horizontal axis of Fig. 2: dialects with preverbal negation (N1), dialects with discontinuous negation in which the postverbal negative marker is not always mandatory (N1 (N2)), dialects with discontinuous negation in which both markers are mandatory (N1 N2), dialects with discontinuous negation in which the preverbal negative marker is not always mandatory ((N1) N2), languages with postverbal negation (N2). The line in Fig. 2 shows that the incidence of suppletive imperatives is higher in systems with N1 and decreases progressively in languages in which preverbal negative markers co-occur with or are replaced by a postverbal negative marker.

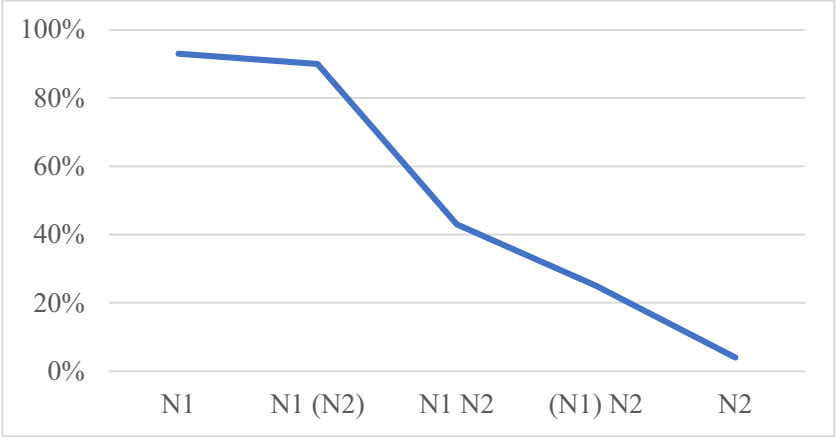

Fig. 2. Incidence of dialects displaying suppletive imperatives per negation system: dialects with preverbal (N1), discontinuous (N1 N2), and postverbal negation (N2). Data from Manzini, Savoia (2005).

If we reduce this pattern of variation to two binary variables (presence of preverbal negation *vs* true/fake imperatives, we can easily show that the correlation is statistically significant. Table 1 shows the number of dialects that instantiate each of the four logical patterns in Manzini, Savoia's sample. The result of the chi square test is highly significant ($p < 0.00001$), i.e. the probability that these two variables are randomly associated is extremely low.

	True Imp	**Fake Imp**
No PreV Neg	77	20
PreV Neg	7	83

Table 3. Number of dialects exhibiting fake/true imperatives with/without preverbal negation. Dataset: northern Italian dialects from Manzini, Savoia (2005).

The data in Fig. 2/Table 3 are solid evidence supporting Zanuttini's claim in (a), at least for the (Italo)-Romance area, despite the occurrence of some outliers/counter-examples. In fact, the three huge volumes of Manzini, Savoia 2005 contain counterexamples for any descriptive generalization proposed so far on Italo-Romance dialects, including (a). For instance, Manzini, Savoia (2005) show that dialects of the Livo type in (6a) have suppletive imperatives with a postverbal negation of type *no*, whereas varieties of the Mercato Saraceno type in (6b) display true imperatives that are preceded by a preverbal negative marker.

(6) a. ma'ɲarlo nɔ (Livo, TN)
eat.inf=it neg
'Don't eat it!'
b. na be (Mercato Saraceno, FC)
neg dring.imp
'Don't drink!'

Besides weakening empirical generalizations such as (a), counterexamples also question the validity of the theoretical interpretation of generalizations. To illustrate this point, let us introduce a toy-model of fake imperatives, in (7). According to (7), the mechanism producing fake imperatives is the lack of verb movement: let us suppose that imperatives in (7b) move higher (i.e. more to the left) than the inflected verbs in declaratives, see (7a). According to this simplified analysis, it has been claimed that preverbal negation somehow disrupts V's movement in (7c), thus preventing the verb from acquiring imperative morphology, even if the clause maintains imperative force (Zanuttini 1997 among others).

(7) a. [$_{\text{Declarative clause}}$ … [V …]
b. [$_{\text{Imperative clause}}$ V [~~V~~ …]
c. [$_{\text{Imperative clause}}$ * V NEG [V …]

The analysis in (7) accounts for the "well-behaved" varieties, those that comply with the generalization in (a), but the other 27 varieties that are exemplified in (6)—and many languages of the world examined in the WALS—are incompatible with (7). Manzini, Savoia (2005) reached the conclusion that both the empirical generalization in (a) and its theoretical interpretation in (7) are to be rejected. Benincà, Poletto (2004), on the contrary, make the effort to refine the account of (apparent) counterexamples without discarding (a) and (7). For instance, they postulate that in dialects such as (8), which show the same pattern as in (5a), a null preverbal negative marker (represented as ø$_{\text{NEG}}$) hinders verb movement, thus barring imperative morphology.

(8) ø$_{\text{NEG}}$ movra=t mia! (Albinea, RE)
move.INF=yourself NEG
'Don't move!'

Besides technicalities, the gist of Benincà and Poletto's (2004) analysis is that the linguistic competence of speakers is sufficiently *plastic* to accommodate counter-examples like (8), especially in dialects of the Emilian area where the loss of the preverbal negative marker is relatively recent, witness the attestations of languages of the (N1) N2 type. The fact that the preverbal negative marker remains silent ($ø_{NEG}$) is a microparameter in Borer's sense, i.e. a property of a single lexical element that learners acquire from experience. In this view, micro/nano-parameters (i.e. parameters of single lexical items) can be seen as adaptive responses that aim to circumvent a possible violation of higher-grade parameters and, hence, as a proof of the existence of higher-grade parametric choices.

To summarize, microvariation data from Italo-Romance show a robust association between two properties (Table 3, Fig. 2) that, crosslinguistically, are not systematically associated (Table 2). The association is not confirmed by typological data and, moreover, the WALS data suggest that the association found in Italo-Romance is *typologically marked*: it is dominant in northern Italo-Romance dialects, but it is marginal across other linguistic families and groups (see Table 1). Additionally, the generalization is challenged by several counterexamples. The generalization is therefore challenged twice: at the macro level (viz. typologically) and at the micro level.

In my opinion, however, neither the validity of the generalization in (a) nor its parametric interpretation are undermined by counterevidence.

The existence of outliers within a linguistic (sub)group is not necessarily at odds with parametric models. All parametric models acknowledge the existence of micro/nano-parameters allowing speakers to accommodate what they learn from experience as proposed by Benincà and Poletto (2014) with respect to (8). But crucially, microparameters cannot account for all variation patterns, in particular for solid associations like the one illustrated so far.

Similarly, the fact that statistically significant associations within a linguistic group are not confirmed by broader typological evidence is expected under a parametric model, which aims to account for systematic patterns of variation, regardless of their incidence across other linguistic families.

4. Clusters of correlations

Starting from the assumption that evidence from statistically significant associations between linguistic variables are a solid test-bed for parametric models, we can start collecting more (meta)data from existing datasets (on-line databases, atlases, grammars, etc.) to find other correlations that are not amenable to historical/contact explanations and see whether and how correlations interact with each others. In the preceding section we dealt with an association between two simple variables (X─Y), but we cannot exclude associations with a more complicated format, e.g. X&Z─Y (meaning: the incidence of variable Y is conditioned by the co-occurrence of variables X and Z).

To illustrate this point, this section elaborates on another phenomenon that, like fake imperatives, is associated with the position of clausal negation markers. In northern Italian dialects, the position of preverbal negation can be pinpointed by looking at its position with respect to subject clitic pronouns. In the subset of varieties where negation is preverbal, subject clitics can either precede or follow preverbal negation, as in (9a) and (9b), respectively.[3]

(9) a. **al** **(na)** ˈdɔrma ˈmia (Agazzano, PC)
he= not= sleep.3SG neg
b. **no** **l** ˈdɔrme ˈmia (Verona)
not= he= sleep.3SG neg
'he does not sleep.'

In the light of (9), it is tempting to revise the generalization in (a) about fake imperatives. One may hypothesize that the position of preverbal negation with respect to subject clitics may affect the licencing of true/fake imperatives, thus accounting for the outliers that do not comply with (a). Let us suppose, for the sake of the argument, that preverbal negation is compatible with true imperatives if and only if it precedes subject clitics. The data in Table 4, however, show that there is no significant association between these two properties: true imperatives are found regardless of the position of preverbal negation with respect to subject clitics (p = .87).

	True Imp	**Fake Imp**
Neg > SCls	2	6
SCls > Neg	18	47

Table 4. Number of dialects exhibiting fake/true imperatives and in which SCls precede/follow negation. Dataset: Manzini, Savoia (2005).

The data in Table 4 do not shed light on the counterexamples discussed in the previous section. However, we can try to relate the variable in (9) with other properties of subject clitics.

Pescarini (2022) found that the microvariation exhibited by northern Italian dialects with respect to the syntax of subject clitics is less kaleidoscopic than part of the recent literature suggests. Chi-square tests showed that the following variables are significantly associated.

3. An anonymous reviewer pointed out that some dialects have a dual series of clitic formatives, which occur before and after the negation marker. The pattern is often attested in Friulian dialects. The question whether this pattern result from a syntactic parameter or from some morphological readjustment will remain open to further discussion. As for the purpose of the present study, only person-formatives were examined, i.e. subject clitic items carrying person agreement, e.g. *tu* 'you', *(e)l* 'he', etc.

- Doubling: the (non) occurrence of subject clitics with non-dislocated subjects (e.g., the wh element *who*), which in some dialects is barred, as shown in (10);[4]
- Expletives: the (non) occurrence of subject clitics with non-thematic predicates, e.g. weather verbs in (11);
- Gaps: the (non) occurrence of subject clitics in all persons of the paradigm, e.g. in (12) the 1sg subject clitic is missing.

(10) a. ˈmarjo el ˈriva doˈmaŋ (Verona)
Mario 3SG.M.NOM= arrive.3SG tomorrow
'Mario (he) will arrive tomorrow.'
b. ʧi ˈriv(a) (*elo)
who arrive.3SG =3SG.M.NOM
'Who will arrive?'

(11) a. el ˈriva
3SG.M.NOM= arrive.3SG
'He/she is coming.'
b. (*el) ˈpiove
3SG.M.NOM= rain.3SG
'It rains'

(12) a. __ ˈrivo
arrive.1SG
b. te ˈrivi
2SG= arrive.2SG

Veronese, in (10)-(12) never exhibits subject clitics in the above environments, whereas other northern Italian dialects exhibit subject clitic formatives in one or more of the contexts in (10)-(12). By studying the occurrence of subject clitics in these three clausal environments, two significant associations emerge. Gaps are (negatively) associated with expletives (as already noticed by Renzi and Vanelli 1983 on the basis of a narrower dataset). In turn, expletives are positively associated with doubling (doubling is rarely found in languages without expletives). No direct association is found between the paradigmatic structure of clitics (i.e., gaps) and doubling. Associations are schematized in (13). Recall that associations do not necessarily imply causation.

4. It is well-known, since Poletto's (2000) seminal work, that subject clitics are less likely to co-occur with operator-like subjects such as WH elements and bare quantifiers. By focusing on WH subjects, we can therefore individuate the varieties in which subject clitics are mandatory across the board. An anonymous reviewer pointed out that, in several varieties, main interrogatives require enclisis, which adds at least another variable to the association scheme. However, since third person subject clitics are usually found both in enclisis and proclisis, I believe that, for the purposes of the present study, the effect of orthogonal variables can be safely disregarded.

(13) gaps ─ expletives ─ doubling

We can now test if the above three variables are associated with the fourth variable introduced in this section, namely the position of preverbal negation with respect to subject clitics (in the contexts where subject clitics are attested). Table 5 reports the counts of dialects that exhibit/do not exhibit gaps in the first person singular (the clitic form that is most likely to be missing) and the position of negation with respect to the clitic forms that are attested (usually the second singular and/or third persons). The two variables are associated ($p < 0.00001$): no variety with the order negation > subject clitic (which, historically, is an innovation) displays a full array of subject clitics.

	Neg > SCls	**SCls > Neg**
Gap	11	16
No Gap	0	49

Table 5. Number of dialects where the 1sg is missing and other SCls precede/follow the preverbal negation. Dataset: Manzini, Savoia (2005).

Likewise, the association between the position of negation and the presence of expletive subject clitics in impersonal clauses (specifically, in clauses containing a weather predicate, see Tab. 6) is significant at $p < .05$ ($p = 0.04$).

	Neg > SCls	**SCls > Neg**
No Expletive	4	7
Expletive	7	58

Table 6. Number of dialects in which SCls occur with weather verbs and SCls precede/follow prevernal negation. Dataset: Manzini, Savoia (2005).

Lastly, no significant association (at $p < .05$) is found between the position of negation and the co-occurrence of subject clitics and subject wh elements (WH doubling), see Table 7.

	Neg > SCls	**SCls > Neg**
No Doubling	1	11
Doubling	7	21

Table 7. Number of dialects in which SCls double *who* and precede/follow preverbal negation. Dataset: Manzini, Savoia (2005).

In the light of evidence from Table 5-7, a fourth variable can be added to the cluster in (13), as shown in (14). The data introduced so far showed that the position of preverbal negation with respect to subject clitics (PrevNegPos) is associated with two other syntactic variables such as the occurrence of expletive clitics in impersonal clauses and the occurrence of subject clitic formatives in all persons.

```
(14) Gaps ─ Expletives ─ Doubling
       |       |
      PrevNegPos
```

The variable PrevNegPos is in turn logically linked to a super-ordinated variable, which is the position of clausal negation markers (preverbal vs postverbal). The former (PrevNegPos) is relevant if and only if a language displays a preverbal negative marker (PrevNeg), which in turn is associated with fake imperatives, see Section 3. In (15), the arrow represents a logical entailment, whereas lines represent statistically significant associations between variables.

```
(15) Gaps ─ Expletives ─ Doubling
       |       |
     PrevNegPos
         ↓
      PrevNeg
         |
  Fake imperatives
```

The cluster of associations in (15) results from the statistical analysis of the distribution of syntactic variables in Manzini, Savoia's (2005) sample of dialects. Not all dialects conform to (15) because, as previously mentioned, our linguistic competence is shaped by various kinds of factors (Chomsky 2005), including unpredictable historical factors that can be accommodated by acquirers. I have assumed that linguistic acquisition is adaptive, i.e. like a neural network, it can change through growth and reorganization in response to new information. However, the fact that certain clusters of variables, like those in (15), are relatively stable—statistically speaking—indicates that syntactic competence possesses some degree of rigidity, which, if I am correct, can be imputed to an abstract system of parameters.

5. Conclusions

This contribution aimed to explore a new test bed for parametric theory. Syntactic parameters are often postulated on the basis of comparative evidence from genealogically unrelated languages. Within a sample of genealogically-related languages such as Italo-Romance dialects, the results of parametric analyses are more controversial.

Most scholars argue or assume that the analysis of microvariation, i.e. variation across genealogically-related languages/dialects, may fine-tune parametric models by specifying lower-ranked parametric choices (microparameters), which target single functional items.

Poletto (2013) departs from this view, arguing that microvariation does not necessarily result from microparameters. In a similar vein, I argued that, by sifting data through statistics, microvariation data can contribute to the discovery of higher-grade parametric choices. Significant associations between syntactic variables (those that have no plausible historical explanation in languages that are otherwise very similar) may shed new light on the existence, nature, format, and organization of syntactic parameters *beyond* idiosyncratic (viz lexical) microparameters.

Bibliographical references

Auwera, J. van der, Lejeune, L. 2013, *The Prohibitive*, in M.S. Dryer, M. Haspelmath, (eds.), *WALS Online* (v2020.3). Zenodo. https://doi.org/10.5281/zenodo.7385533 (Available online at http://wals.info/chapter/71, Accessed on 2024-04-12.)

Baker, M. 2008, *The macroparameter in a microparametric world*, in Biberauer, Th. (eds.), *The limits of syntactic variation,* Amsterdam, Benjamins: 351-374.

Benincà, P., Poletto, C. 2005, *On Some Descriptive Generalizations in Romance*, in G. Cinque, R. S. Kayne (eds.), *The Oxford Handbook of Comparative Syntax*, Oxford, Oxford University Press: 221-258.

Biberauer, T., Holmberg, A., Roberts, I., Sheehan, M. 2010, *Parametric Variation: Null Subjects in Minimalist Theory*, Cambridge, Cambridge University Press.

Boeckx C., Leivada E. 2013, *Entangled Parametric Hierarchies: Problems for an Overspecified Universal Grammar*, «PLoS ONE» 8(9): e72357.

Borer, H. 1984, *Parametric Syntax*, Dordrecht, Foris.

Chomsky, N. 2005, *Three Factors in Language Design*, «Linguistic Inquiry» 31(1): 1-22.

Crisma, P., Guardiano, C., Longobardi G. 2020. *Syntactic diversity and language learnability*, «Studi e saggi linguistici» 58(2): 99-130.

Delfitto, D., Schroten J. 1991. *Bare plurals and the number affix in DP*, «Probus» 3: 155-186.

Dryer, M.S. 2013, *Order of Negative Morpheme and Verb*, in M.S.Dryer, M. Haspelmath (eds.), *WALS Online* (v2020.3) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.7385533 (Available online at http://wals.info/chapter/143, Accessed on 2024-04-12.)

Guardiano, C., Michelioudakis, D., Ceolin, A., Irimia, M., Longobardi, G., Radkevic, N., Silvestri, G., Sitaridou, I. 2016, *South by South East. A Syntactic Approach to Greek and Romance Microvariation*, «L'Italia Dialettale» 77: 96-166.

Guardiano, C. et al. 2018. *Parametric comparison and dialect variation. Insights from Southern Italy*, in L. Repetti, F. Ordóñez (eds.), *Romance Languages and*

Linguistic Theory 14: Selected papers from the 46th Linguistic Symposium on Romance Languages (LSRL), Stony Brook, NY, Amsterdam, Benjamins: 103-133.

Guardiano C., Cambria, M., Stalfieri, V. 2022, *Number Morphology and Bare Nouns in Some Romance Dialects of Italy*, «Languages» 7: 255.

Harley, H., Ritter, E. 2002, *Person and Number in Pronouns: A Feature-Geometric Analysis*, «Language» 78(3): 482-526.

Jaeggli, O. 1982, *Topics in Romance Syntax*, Dordrecht, Foris.

Kayne, R.S. 1989, *Facets of Romance past participle agreement*, in P. Benincá (ed.), *Dialect Variation and the Theory of Grammar: Proceedings of the GLOW Workshop in Venice, 1987*, Berlin, De Gruyter: 85-104.

Kayne, R.S. 1996, *Microparametric Syntax. Some Introductory Remarks*, in J.R. Black, V. Motapanyane (eds.), *Microparametric Syntax and Dialect Variation*, Amsterdam, Benjamins, ix-xviii.

Labov, W. 1966, *The Linguistic Variable as a Structural Unit*, «Washington Linguistics Review» 3: 4-22.

Ledgeway, A. 2019, *Parameters in the Development of Romance Perfective Auxiliary Selection*, in M. Cennamo, C. Fabrizio (eds.), *Historical Linguistics 2015. Selected Papers from the 22nd International Conference on Historical Linguistics, Naples, 27-31 July 2015*, Amsterdam, Benjamins: 343-384.

Ledgeway, A. 2020, *Rethinking Microvariation in Romance Demonstrative Systems*, in A. Bárány, Th. Biberauer, J. Douglas, S. Vikner (eds.), *Syntactic Architecture and its Consequences. II. Between Syntax and Morphology*, Berlin, Language Science Press: 451-490.

Ledgeway, A., Schifano, N., Silvestri, G. 2019, *Differential Object Marking and the properties of D in the dialects of the extreme south of Italy*, «Glossa» 4(1): 51.

Longobardi, G. 2001, *Formal syntax, diachronic minimalism, and etymology: The history of French* Chez, «Linguistic Inquiry» 32(2): 275-302.

Longobardi G. 2018, *Principles, Parameters, and Schemata: A radically under-specified UG*, «Linguistic Analysis» 41(3-4): 517-558.

Longobardi, G., Guardiano, C. 2009. *Evidence for Syntax as a Signal of Historical Relatedness*, «Lingua» 119: 1679-1706.

Loporcaro, M., Pescarini, D. 2022, *Variation in Romance*, in A. Ledgeway, M. Maiden (eds.), *Cambridge Handbook of Romance Linguistics*, Cambridge, Cambridge University Press.

Manzini, M.R., Savoia, L. 2005, *I dialetti italiani e romanci*, Alessandria, Edizioni dell'Orso.

Pellegrini, G.B. 1977, *Carta dei dialetti d'Italia*, Pisa, Pacini.

Pescarini, D. 2022, *Expletive Subject Clitics in Northern Italo-Romance*, «Languages»: 7: 265.

Pescarini, D. 2023. *La concordanza negativa nei dialetti altoitaliani: i dati dell'AIS*, in V. Faraoni, L. Filipponio, T. Paciaroni, S. Schmid (eds.), *Prospettive di ricerca in linguistica italiana e romanza. Studi offerti a Michele Loporcaro dagli allievi e dai collaboratori zurighesi*, Pisa, Edizioni ETS: 235-252.

Poletto, C. 2013, *Leopard spot variation: What dialects have to say about variation, change and acquisition*, «Studia Linguistica» 67: 165-183.
Poletto, C. 2016, *Negation*, in A. Ledgeway, M. Maiden (eds.), *The Oxford Guide to Romance Language*, Oxford, Oxford University Press: 833-846.
Poletto, C. 2000, *The Higher Functional Field. Evidence from Northern Italian Dialects*, Oxford, Oxford University Press.
Renzi, L., Vanelli, L. 1983, *I pronomi soggetto in alcune varietà romanze*, in P. Benincà *et al.* (eds.), *Scritti in onore di G.B. Pellegrini*, Pisa, Pacini: 120-145.
Roberts, I. 2010, *Agreement and Head Movement: Clitics, Incorporation, and Defective Goals*, Cambridge (Mass.), MIT Press.
Roberts, I. 2014, *Subject clitics and macroparameters*, in P. Benincà, P., A. Ledgeway, N. Vincent (eds.), *Diachrony and Dialects: Grammatical Change in the Dialects of Italy*, Oxford: Oxford University Press: 177-201.
Taraldsen, K. 1980, *On the Nominative Island Constraint, Vacuous Application and the That-Trace Filter*, Indiana University Linguistics Club.
Zanuttini, R. 1997, *Negation and Clausal Structure: A Comparative Study of Romance Languages*, Oxford, Oxford University Press.