Quaderni di Linguistica e Studi Orientali / Working Papers in Linguistics and Oriental Studies
n. 5 (2019), pp. 255-277
DOI: http://dx.doi.org/10.13128/QULSO-2421-7220-25968

Microvariation and Microparameters. Some Quantitative Remarks

Diego Pescarini
CNRS, Université Côte d'Azur, BCL (<diego.pescarini@univ-cotedazur.fr>)

Abstract:

This paper deals with the distribution of subject clitics in northern Italian dialects. Building on quantitative data, I argue that the observed microvariation cannot derive (only) from external linguistic factors such as contact, areal diffusion, sociolinguistic dynamics, etc. Rather, a principled feature-based analysis is needed in order to account for certain patterns of defectivity and syncretism that, although typologically rare, occur systematically in northern dialects.

Keywords: *clitics, gaps, syncretism, dialects, microvariation, Italo-Romance*

1. Introduction

As illustrated in Table 1, paradigms of subject clitics in central Romance dialects are often defective and exhibit systematic patterns of syncretism:

	Olivone	Corte	Grumello	Fornero	Piverone	Calasetta	Tayac
1p	*a*		*(a)*	*i*	*i*		
2p	*tu*	*te*	*(a) ta*	*ti*	*at*	*ti*	*tœ*
3p (m/f)	*u/ra*	*l/la*	*al/(a)*	*al/la*	*al/la*	*u/a*	*ew*
4p	*a*		*a n*	*i*	*i*		
5p	*a*		*(a)*	*i*	*i*		*vusaw*
6p (m/f)	*i*	*i/le*	*(a) i*	*i*	*a*		*zi*

Table 1. Paradigms of subject clitics in Italo and Gallo-Romance dialects

ISSN 2421-7220 (online)
www.fupress.com/bsfm-qulso

Gaps and syncretisms are sensitive to person distinctions. Previous studies revealed some robust trends in the form of implicational statements, which lend themselves to an analysis in terms of feature geometries and microparameters (Heap 2002; Manzini and Savoia 2005, 72ff; Benincà and Poletto 2005; Oliviéri 2011; Calabrese 2011).

However, before adopting microparametric or feature-based explanations, it is worth examining and – if possible – discarding alternative hypotheses. Among alternative hypotheses, one might contend that cross-linguistic variation across a set of closely related languages is always amenable to external explanations. Let us assume that some innovative speakers began to use an idiosyncratic variant V_1, which spread across social strata and nearby speaking communities until it gave rise to further subvariants $V_{2a/2b/etc.}$ and so on. In this way, paradigms of SCls gradually shifted from one type to another, yielding *prima facie* hierarchical arrays. In other words, by studying genealogically-related languages, one always 'runs the risk [to] discover shared innovations that have purely historical explanations, rather than properties that are shared because of the same parameter setting.' (Haspelmath 2008, fn 8).

The paper aims to address the above hypothesis on the basis of statistical evidence based on a dataset of 187 dialects reported in Manzini and Savoia 2005. To address the null hypothesis (i.e. 'purely historical explanations'), I will show that the geolinguistic distribution of variants does not support an account entirely based on external/historical factors.

The conclusion of the present study is in line with the premises of Longobardi and Guardiano's 2009 *Parametric Comparative Method* (PCM), which has been applied to the analysis of syntactic microvariation in Greek and (southern) Italian dialects (Guardiano *et al.* 2016). Although the methodology and the spirit of the present study are germane to the PCM, the goal of this work is much less ambitious. Longobardi and Guardiano argue that, by adopting a parametric approach, syntactic comparison is as reliable as the comparative method of historical linguistics: the clusters of languages generated by the PCM correspond to the linguistic families and groups reconstructed by means of non-syntactic comparative evidence. Then, by validating the PCM, Longobardi and Guardiano show that linguistic classifications must rely upon abstract syntactic parameters rather than superficial similarities.

The PCM approach, however, cannot be easily extended to the analysis of subject clitic systems. Since subject clitics are attested in a homogeneous linguistic area, the results of our quantitative analysis cannot be tested against a genealogical clustering. Hence, whereas the comparative method provides a benchmark to evaluate Longobardi and Guardiano's parameters, no independent evidence allows us to validate feature-geometric analyses such as Heap 2002; Benincà and Poletto 2005; Oliviéri 2011; Calabrese 2011. For this reason, the article departs from the PCM and follows a bottom-up approach to

microvariation in which higher grade accounts (e.g. feature geometries, microparameters) are supported indirectly by dismissing lower grade hypotheses (e.g. contact, analogy, etc.).

The structure of the paper is as follows: section 2 deals with the nature of subject clitics in northern Italian dialects and wonders about the relationship between subject clitics and the Null Subject Parameter (NSP); section 3 overviews the distribution of gaps and patterns of syncretism in Manzini and Savoia's 2005 sample; section 4 elaborates on the correlation between linguistic and geographical distance.

2. Subject clitics and the Null Subject Parameter

Subject clitics occur in Gallo-, Italo- and Rhaeto-Romance varieties. The null subject parameter cuts across the area of subject clitics: northern Italian dialects exhibit subject clitics, cf. (1c), but, unlike French, they are characterised by the canonical properties of null subject languages: they are not subject to the so-called *that*-trace effect, cf. (2c), and allow free inversion as in (3c). For these reasons, clitics in northern Italian dialects have been often analysed as agreement markers, rather than fully-fledged pronouns (Rizzi 1986; Brandi and Cordin 1989).

(1) a. parla italiano. (It.)
speak.3SG Italian
b. *(Il) parle italien. (Fr)
3SG.NOM= speak.3SG Italian
c. *(El) parla italian (Ver.)
3SG= speak.3SG Italian
'He speaks Italian'

(2) a. Chi hai detto che ha scritto questo libro? (It.)
who have.you said that has written this book
b. *Qui as-tu dit qu' a écrit ce livre? (Fr.)
who have=you said that has written this book
c. Ci ghe-to dito che l'a scrito sto libro? (Ver.)
who have=you said that he=has written this book
'Who did you say wrote this book?'

(3) a. È arrivato Gianni. (It.)
is arrived John
b. *Il est arrivé Jean. (Fr.)
he= is arrived John
c. L'è rivà Giani. (Ver.)
he= be.3SG arrive.PST.PTCP John.
'John has arrived'

Furthermore, in many northern Italian dialects subject clitics form a defective paradigm, as shown in (4).[1] Besides northern Italy, defective patterns have been found in some northern Occitan dialects (Kaiser, Oliviéri, Palasis 2013) and in Franco-Provençal dialects.

(4)	Mi	_ magno	'I eat' (Ver.)
	Ti	*te* magni	'You eat'
	Lu	*el* magna	'He eats'
	Nialtri	_ magnémo	etc.
	Vialtri	_ magnì	
	Lori	*i* magna	

The above dichotomy between *clitic subject pronouns* of the French type and *subject-agreement clitic markers* of the Italo-Romance type is supported by further evidence: in northern Italian dialects, but not in French, subject clitics can double a non-dislocated subject, follow negation, and cannot be dropped under coordination:

(5)	a.	Nessuno	gli	ha	detto	nulla. (Flo.)
		none	3SG=	have.3SG	say.PST.PTCP	nothing
	b.	*Personne	il	n' a	rien	dit. (Fr.)
		none	3SG.NOM=	NEG= have.3SG	nothing	say.PST.PTCP
		'Nobody has said anything'				

(6)	a.	Un	tu	compri	mai	mele. (Flo.)
		NEG	2SG=	buy.2SG	never	apples
	b.	Tu		n' achètes	jamais	de pommes. (Fr.)
		2SG.NOM=		NEG buy.2SG	never	of apples
		'You never buy apples'				

(7)	a.	La	canta	e	la	balla (Flo.)
		3SG.F=	sing.3SG	and	3SG=	dance.3SG
	b.	Elle		chante	et	danse. (Fr.)
		3SG.F.NOM=		sing.3SG	and	dance.3SG
		'She sings and dances'				

In fact, Poletto (2000) shows that northern Italian dialects, although behaving like null-subject languages, do not always allow doubling (in particular with operator-like subjects), do not always display the order ne-

[1] The presence/absence of subject clitics may vary across clause types as the inventories of proclitics and enclitics are often dissimilar. This point will not be discussed further; what follows is based on the analysis of proclitics.

gation > clitic, and, under certain circumstances, allow the omission of certain clitic forms in coordinated structures. At the same time, corpus studies have shown that in French varieties such as colloquial metropolitan French as well as Quebec, Ontario, and Swiss varieties of French (see Culbertson 2010; Palasis 2015 and references therein), subject clitics and NP/DP subjects (including strong pronouns) co-occur even if the latter are not dislocated.

Further problems for the claim that Italo-Romance subject clitics are agreement markers come from the analysis of varieties in which subject clitics seem to occur optionally. In Paduan, for instance, third person subject clitics do not always occur (Benincà 1994). First, subject clitics are ungrammatical whenever the subject is postverbal:

(8) *El riva to fradèo. (Pad.)
He= arrives your= brother
'Your brother is coming'

With preverbal subjects, the clitic occurs if and only if the subject is left-dislocated (Benincà and Poletto 2004):

(9) a. Mario (l) compra na casa. (Pad.)
Mario (he=) buys a house
'Mario is going to buy a house'
b. Mario, na casa, no *(l) la compra.
Mario, a house, not (he=) it= buys
'Mario is not going to buy a house'

The analysis of subject clitics as agreement elements is at odds with the complementary distribution between subject clitics and non-dislocated subjects. If clitics were agreement elements, they should always occur regardless of the position of the doubled subject.

Another problem for the hypothesis that subject clitics are agreement heads comes from the presence of expletive subject clitics in impersonal clauses. For instance, the dialect of Monno in (10a) displays the non-agreeing/expletive clitic *el* with weather verbs and other impersonal predicates, whereas in the dialect of Trieste in (10b) no clitic formative occurs in impersonal clauses.

(10) a. El plof. (Monno)
3.MSG= rain.3
'It is raining'
b. Piovi. (Trieste)
rain.3
'It is raining'

Since both Monnese and Triestino are null subject languages, then one wonders about the nature of the element *el* in (10a), which occurs in the same contexts in which non-null subject languages normally require expletives. Expletives are normally regarded as placeholders, i.e., dummy elements having the same status of phrasal subjects. However, if Italo-Romance subject clitics were agreement markers, how could they satisfy any syntactic requirement related to the subject position?

Second, if subject clitics were agreement markers, they would occur in all impersonal constructions as well as in prototypical subject-less contexts such as imperatives, *contra* evidence. Renzi and Vanelli (1983) observed that expletive clitics do not always occur in all impersonal environments: they are more readily found with weather verbs and, to a lesser extent, with existentials and in impersonal *si* constructions. Some dialects require an expletive clitic to occur with the modal verb expressing impersonal necessity ('it is necessary to'), but – to the best of my knowledge – this happens if and only if the expletive clitic occurs in the remaining impersonal contexts. Hence, the distribution of expletive clitics in impersonal environments follows an implicational scale, illustrated in Table 2:

Variety	Weather verb	Existential construction	Raising construction	Arbitrary construction	Impersonal necessity
Carcare	U ciov	U j-è	U smija...	U s diz	U bsogna
Cesena	E piov	U j-è	E per...	U s dis	Ø bsogna
Monno	El plof	El g'e	El par	Ø s dis	Ø gna
Rocca P.	El piof	L'è	Ø omea	Ø se dis	Ø moza
Aldeno	El piove	Ø gh'e	Ø par	Ø se dis	Ø bisogna
	'it rains'	'there is…'	'it seems that…'	'one says'	'it is needed…'

Table 2. Expletive clitics in impersonal environments
(from Pescarini 2014 with minor modifications)

The above data challenge the idea that subject clitics are agreement heads, but alternative accounts of expletive clitics are even more problematic for the analysis of northern Italo-Romance as null subject systems. In fact, if northern Italian pronouns were considered fully-fledged expletives, they would challenge Gilligan's 1987 correlations in (11). The correlations in (11) resulted from testing Rizzi's early formulation of the Null Subject Parameter against a sample of one hundred languages. According to Gilligan's survey, only the following one-way implications hold true cross-linguistically and, crucially, three out of four generalisations predicts that null subject languages should exhibit null expletives:

(11) a. Free Inversion → expletive null subjects
b. Free Inversion → allow complementiser-trace violations
c. Referential null subjects → expletive null subjects
d. Allow complementiser-trace violations → expletive null subjects

In the light of (11), one does not expect to find overt expletives in languages allowing free inversion or *that*-trace violations. Again, the characterisation of northern Italian dialects as null subject languages leads us to the conclusion that subject clitics are not expletive pronouns, but, at the same time, the distribution illustrated in Table 2 is at odds with the hypothesis that Italo-Romance clitics are agreement markers.

Under a sub-parametrisation of the Null Subject Parameter, one might perhaps argue that northern Italian dialects are a particular kind of *partial Null Subject Languages* (Holmberg 2005) in which the presence of overt subjects is dependent on Person. Besides person-driven gaps, one might argue that also the microvariation with respect to expletive clitics – see Table 2 – follows from the partial NSL status of northern Italo-Romance dialects. In the light of this hypothesis, however, northern Italian dialects would be expected to exhibit other properties of partial pro-drop languages, which Holmberg 2005 summarises as follows:

> (a) Subject prodrop may be restricted to some persons/verb forms and is sensitive to differences of clause type, main/embedded configuration, and register;
> (b) Subject pro-drop is dependent on agreement, but the subject-verb agreement system is deficient in one way or other;
> (c) When subject pro-drop is dependent on an antecedent (a 'controller'), the controller needs to be strictly local;
> (d) There is a null third person singular inclusive generic pronoun;

In particular, northern Italian dialects are expected to resemble a partial NSL such as Brazilian Portuguese. For the sake of completeness, the distribution of null subjects in European and Brazilian Portuguese is illustrated in Table 3 (from Martins and Nunes 2018). BP differs from EP in the acceptability of null subjects, which are allowed with 1st person plural subjects, whereas they are forbidden with 2nd and 3rd singular subjects and with the inclusive impersonal *a gente*. Like northern Italian dialects, BP is not subject to the *that*-trace effect, although it is gradually losing 'free' inversion with transitive and unergative verbs (Barbosa, Duarte, Kato 2005 a.o.). Unlike northern Italian dialects, BP never exhibit expletives.

	EP	BP
nós 'we'	OK	OK
vocês 'you'	OK	??
eles 'they'	OK	??
elas 'they'	OK	??
eu 'I'	OK	??
você 'you'	OK	*
ele 'he'	OK	*
ela 'she'	OK	*
a gente 'we'	*	*

Table 3. Distribution of null subjects in European and Brazilian Portuguese (from Martins and Nunes 2018)

The similarities between BP, northern Italian dialects, and other partial NSLs are quite elusive, but the key factor at play in partial systems is the representation of person features and the relationship between pro-drop and agreement (cf. Holmberg's 2005 statement in (b): 'Subject pro-drop is dependent on agreement'). In this respect, two conceptions of agreement have been advocated: a more 'morphological' view, in which the presence of subject pronouns is linked to the overt marking of verb inflection, and a more 'abstract' view, in which the presence/absence of subject clitic forms results from abstract constraints such as hierarchies of features. In the latter analysis, the externalisation of subject clitics depends on feature geometries or analogous solutions (e.g. filters) allowing/disallowing the spell-out of certain bundles of agreement features (Heap 2002; Benincà and Poletto 2005; Calabrese 2011; Oliviéri 2011).

To summarise, this section has reviewed previous proposals concerning the nature of subject clitics in northern Italian dialects. Subject clitics have been analysed as agreement markers as they can co-occur with non-dislocated subjects and, although northern Italo-Romance dialects are null subject languages, they necessarily occur in finite clauses. Under other respects, however, subject clitics do not behave like agreement markers: no dialect allows subject clitics in prototypical subject-less contexts such as imperative clauses, whereas some dialects have expletive subject clitics in certain, but not all, impersonal environments. The comparison with partial null subject languages may provide a better account of the syntax of subject clitics in northern Italian dialects, but the syntactic diagnostics observed in Germanic languages and Brazilian Portuguese are not convergent. Eventually, the comparison with partial pro-drop systems leads us to wonder about the nature of person-given gaps that characterise partial NSLs.

2. Gaps and syncretism

Paradigms of subject clitics are often defective. Certain persons of the paradigm are not expressed by a subject clitic or, if the clitic form is present, it is either optional or syncretic. For instance, in the Franco-Provencal dialect of Fenis (Laure Ermacora, p.c.) the 1st, 4th and 5th person clitics, which in the dialect of Verona in (4) were missing, are optional (in positive declarative clauses). In the Swiss dialect of Gruyère (De Crousaz and Shlonsky 2003), the optional persons are the 1st, the 3rd and the 6th. Furthermore, in Gruyère the optional clitics are syncretic as they are expressed by the vocalic exponent *i*.

(12) a. *(dze)* ˈpékɔ ɐ̃ ˈpɔma (Fenis, Franco-provençal; Laure Ermacora p.c.)
'I eat an apple'
b. *tu* ˈpékè ʌ̃ ˈpɔma
'you.sg eat an apple'
c. *iu̯* ˈpékè ɐ̃ ˈpɔma
'he/she eats an apple'
d. *(nɔ)* piˈkèn ʌ̃ ˈpɔma
'we eat an apple'
e. *(vɔ)* piˈkodè ɐ̃ ˈpɔma
'you.pl eat an apple'
f. *iu̯* ˈpékʊŋ ɐ̃ ˈpɔma
'they eat an apple'

(13) a. Me *(i)* medzo dou fre. (Gruyère, Switerland; De Crousaz and Shlonsky 2003)
'I am eating cheese'
b. Tè *te* medzè dou pan.
'You are eating bread'
c. li *(i)* medzè chin ti lé dzoa.
'He eats that every day'
d. No*no* medzin rintyé la demindze.
'We eat only on Sundays'
e. Vo *vo* medzidè avu no.
'You (pl.) are eating with us'
f. Là *(i)* medzon to cholè.
'They are eating all alone.

Building on similar data, previous studies revealed some robust trends in the form of implicational statements. Renzi e Vanelli 1983 analysed a sample of 30 dialects and put forth a set of Greenberg-style generalisations, e.g.

(14) a. If a variety has at least one subject clitic, it is 2sg.
b. If a variety has two subject clitics, they are 2sg and 3sg.
c. If a variety has three subject clitics, they are 2sg, 3sg, 3pl

The above statements can be represented by means of a chain of implications, although no single chain can account for all the patterns found so far:

(15) a. 2 > 3 > 6 > 5 > 4 > 1 (Renzi and Vanelli 1983)
b. 2 > 3 > 6 > 1 > 4/5 (Cabredo Hofherr 2004; Calabrese 2011)
c. 2 > 6 > 3 > 4 > 1 > 5 (Heap 2000)

Heap 2002, Oliviéri 2011, Calabrese 2011 among others tried to formulate higher grade generalizations by deriving Person distinction from bundles of abstract features. In fact, the organisation of clitic inventories follows more or less robust trends, rather than categorical principles. As for the presence/absence of singular clitics, for instance, data from Manzini and Savoia (2005, §2.3) revealed that a group of dialects of Trentino have the 3rd person clitic, but no 1st and 2nd person clitic (*contra* Renzi and Vanelli's first generalisation in (14). However, the 2nd person clitics is never missing if the 1st person clitic is present: as shown in the following histogram, a system with the 1st person and without the 2nd person clitic is not attested in the almost 370 dialects of Manzini and Savoia's and ASIt dataset.[2]

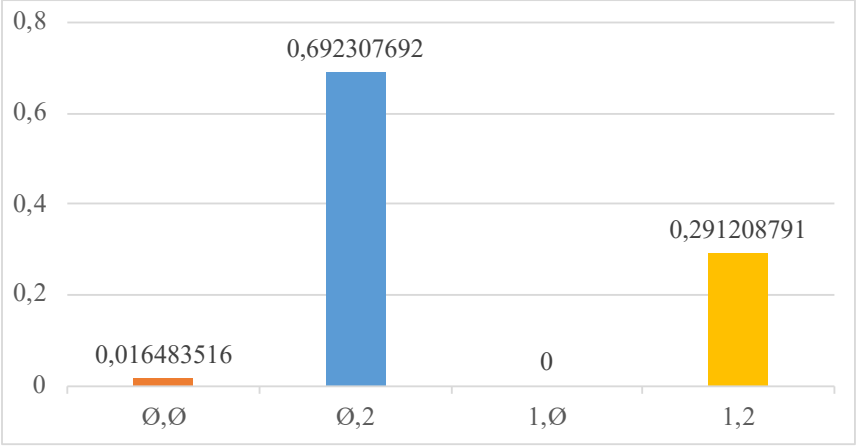


Figure 1. Presence (1, 2) vs Absence (Ø) of 1st and 2nd person subject clitics in northern Italian dialects. Key: Ø, Ø: both first person singular and second person singular are missing; Ø, 2: the first person singular is missing; 1, Ø: the second person singular is missing; 1, 2: both clitics are attested.
Sample: 182 northern Italian dialects; Dataset: ASIt database (retrieved in July 2018). Source: Loporcaro and Pescarini 2019

Since implications are often contradicted by counterexamples, it is worth approaching the problem from a quantitative point of view in which categorical

[2] ASIt: Atlante Sintatico d'Italia, <http://asit.maldura.unipd.it>.

statements are turned into probabilistic generalisations. In order to weigh the generalisations on the distribution of subject clitics, I have analysed the absence vs presence of subject clitics in the sample of 187 northern Italian and Rhaeto-Romance dialects reported in Manzini and Savoia 2005. For each dialect, I surveyed the presence of subject clitics in declarative clauses. The results are summarised in Appendix 1 (for the sake of clarity, a partial screenshot of the matrix is reported in Figure 2). The first column of the Working table reports the 187 datapoints surveyed by Manzini and Savoia 2005; columns 2-7 show the presence/absence[3] ('1' *vs* '0') of subject clitic forms for each Person (recall that '1' means that the clitic is either optional or mandatory). Besides personal pronouns, the table reports the presence *vs* absence of expletive subject clitics with weather verbs (column 8).

Datapoint	**1**	**2**	**3**	**4**	**5**	**6**	**expl**
Olivone	1	1	1	1	1	1	1
Semione	1	1	1	1	1	1	1
Quarna sopra	1	1	1	1	1	1	1
Moncalvo	0	1	1	0	1	1	1
Valmacca	1	1	1	0	0	1	1
Breme	1	1	1	1	1	1	1
Castellinaldo	1	1	1	1	1	1	1
Inveruno	1	1	1	1	1	1	1
Carnago	1	1	1	1	1	1	1
Martignana di Po	1	1	1	1	1	1	1
Casorezzo	1	1	1	1	1	1	1
Arconate	1	1	1	1	1	1	1
Solbiate Arno	1	1	1	1	1	1	1
Càdero	1	1	1	1	1	1	1
San Benedetto Po	1	1	1	1	1	1	1
Saguedo	1	1	1	1	1	1	1
Stienta	1	1	1	1	1	1	1

Figure 2. Working table (see Appendix 1)

I used the R Package 'rworldmap' (South 2011) to plot the results on six geographical maps, one for each Person (key: red points mean that the dialect exhibits no clitic form). 3rd Person and, to a lesser extent, 6th person clitics are almost always present, although, as mentioned in §2, in many eastern

[3] Clitics are considered present in the system even if they are optional.

dialects the occurrence of 3rd person clitics is subject to further conditions, yielding the impression that the presence of the clitic is optional:

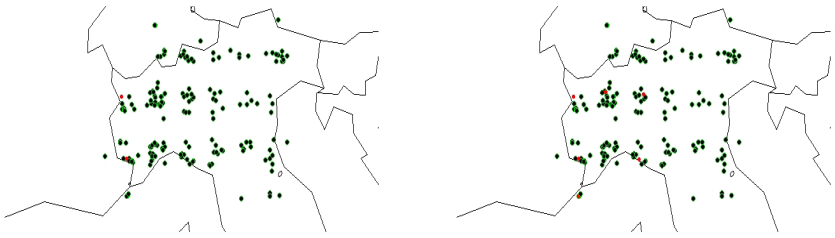


Figure 3. Presence/absence of third person (left) and sixth person clitics (right)

4th and 5th person clitics are often missing, in particular in northeastern dialects. As for the 4th person, it is worth noting that in many Lombard dialects the 4th person results from the reanalysis of an impersonal periphrasis formed by the clitic *om* < HOMO followed by the verb at the third person. Although these dialects do not have a proper 4th person clitics, they have been reported in green in the following map; the presence of the *om* < HOMO formative is reported in column 9 of Appendix 1.

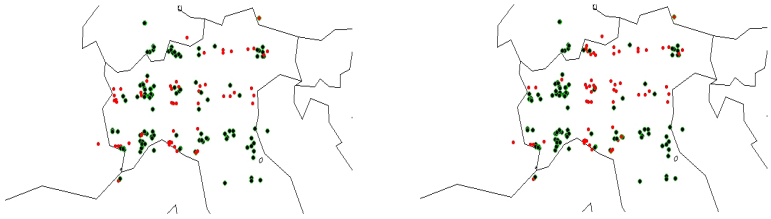


Figure 4. Presence/absence of fourth person (left) and fifth person clitics (right)

As shown in the following maps, the 1st person is frequently missing (like the 4th and the 5th), whereas the 2nd person is almost always mandatory:

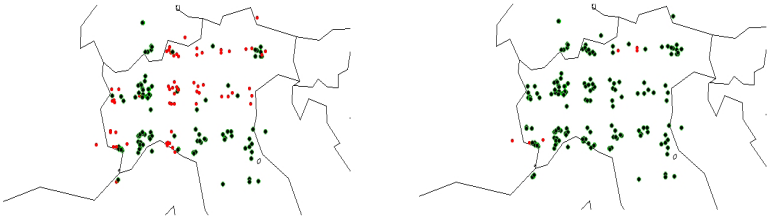


Figure 5. Presence/absence of first person (left) and second person clitics (right)

The above maps confirm – at large – previous impressionistic generalisations. However, to assess linguistic generalisations, we need to turn our empirical generalisations into probabilistic measurements. I first calculated the correlation between the presence/absence of clitics across persons. The following table reports the correlation indexes calculated on the basis of the data from Appendix 1 (Manzini and Savoia's 2005 187 dialects). Each cell of the table reports the degree of correlation for each pair of personal pronouns. Two personal pronouns correlate positively if, for each dialect and each pair of persons, clitics are either present or missing.

2P	3P	4P	5P	6P	EXP	
0,24	0,15	0,61	0,68	0,15	0,44	1P
	0,28	0,39	0,39	0,33	-0,06	2P
		0,45	0,51	0,86	0,37	3P
			0,76	0,50	0,49	4P
				0,57	0,44	5P
					0,44	6P

Figure 6. Correlation between persons: gaps
(sample: 187 dialects from Manzini and Savoia 2005, cf. Appendix 1)

The presence of the expletive clitics with meteorological verbs does not correlate with the presence of any personal pronoun and, from now on, expletive clitics will not be examined anymore. The degree of correlation between personal forms is represented in the following radar plot: each vertex of the hexagon represents a person (1P, 2P, etc.) and each coloured line represents the degree of correlation between one Person and the other five; the closest the line is to the vertex, the highest the correlation:

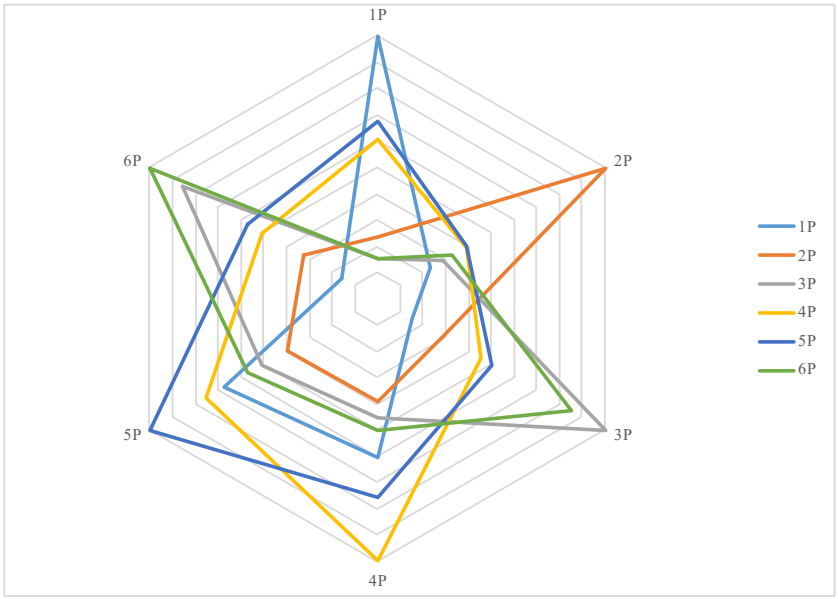


Figure 7. Radar plot of the correlation matrix in Figure 6

The 4th and 5th person exhibit a high degree of correlation; hence, the shape of the two lines is very similar. The same holds for the 3rd and 6th person, which almost overlap.

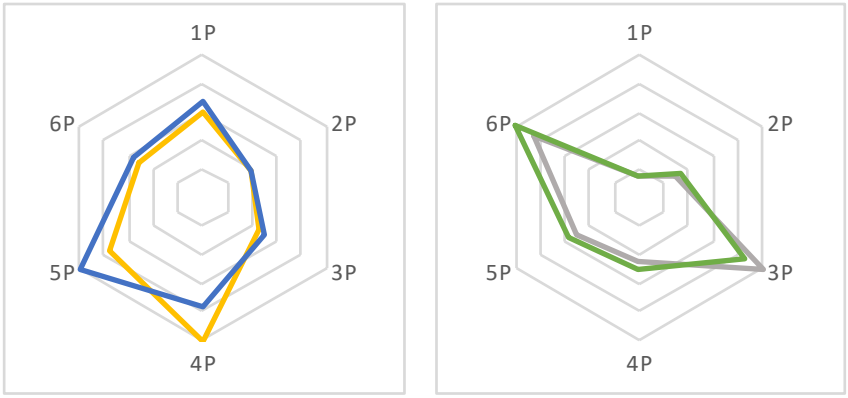


Figure 8. Radar plots of the Fourth/Fifth Person (left) and Third/Sixth Person (right)

Conversely, the 2nd Person exhibits a very low degree of correlation with any other clitic. The 1st Person has a very puzzling interaction as it correlates with the 4th and 5th Person. The correlation 1st/4th is expected given that the 4th person denotes a set containing the speaker, but the pattern formed by the 1st/4th/5th person is typologically rare and it cannot be derived from a non-disjunctive set of person features.

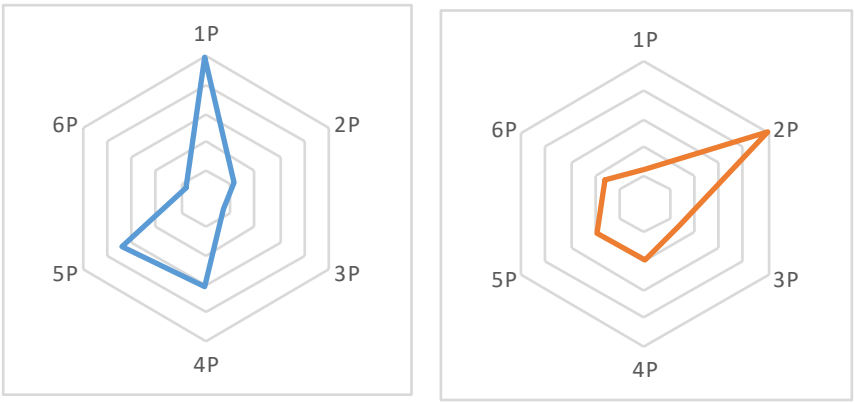


Figure 9. Radar plots of the First Person (left) and Second Person (right)

The specificity of the 1st/4th/5th cluster is further confirmed by patterns of syncretism, i.e. identity of exponence. The following histogram shows the incidence of various patterns of syncretism in Manzini and Savoia's sample of 187 dialects. The bars show the diffusion (number of dialects) of each pattern of syncretism, e.g. the first bar means that the pattern '145', in which the 1st, 4th, and 5th Person are syncretic, is attested in 50 dialects of Manzini and Savoia's sample. Some dialects exhibit two syncretic exponents, e.g. the bar labelled '145&36' represents the number of dialects having one exponent for the 1st, 4th, and 5th Person and another exponent for the 3rd and 6th person. In tabulating the data about syncretism, several factors have been examined: for instance, 3rd and 6th person forms have been considered syncretic iff the masculine and the feminine forms are syncretic; in the case of dialects allowing the co-occurrence of multiple formatives I considered the resulting complex form as a single clitic; I assumed no principled distinction between syncretism and homophony.

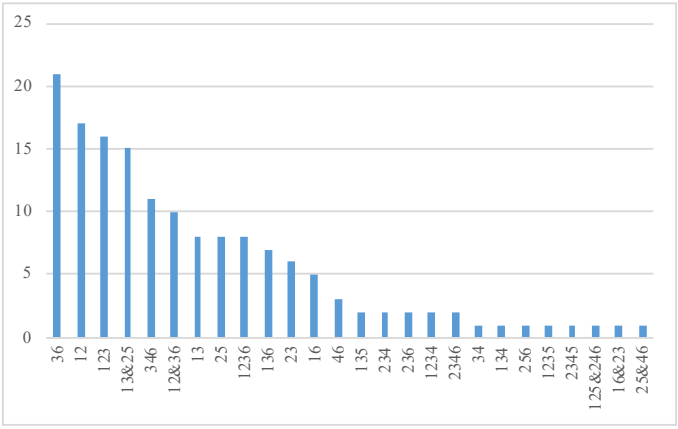


Figure 10. Incidence of various patterns of syncretism. Dataset: Manzini and Savoia 2005

The possible patterns of syncretism are rather constrained: 80% of the dialects show a syncretic exponent for the 1st and 4th Person, whereas a remarkable 67% of the dialects have a single syncretic exponent for the 1st/4th/5th cluster. The 6th person is involved in several patterns of syncretism, whereas the 3rd person is syncretic only with the 6th person. Lastly, the 2nd person is involved in only 2 (very complex) patterns of syncretism, which confirms the impression that the 2nd person clitic has no interaction with the rest of the system.

To sum up, the crosslinguistic distribution of gaps and syncretism follows very robust trends. Some of the above trends are quite predictable: for instance, the 1st and 4th person tend to pattern alike in many linguistic systems of the world, arguably because of the pivotal role of the feature [speaker]. Other patterns, however, are typologically uncommon: for instance, most (Italo)Romance dialects exhibit a robust subsystem of subject clitic elements formed by the 1st/4th/5th person. This cluster cannot be defined by a non-disjunctive set of person features and is typologically rare.

4. The correlation between linguistic and geographical distance

The fact that the 145 pattern is so widespread does not necessarily support internal explanations. In fact, 'which language spreads in a spread zone is a matter of historical accident, and this historical accident can distort the statistical distribution of linguistic types in an area' (Nichols 1992: 23). For instance, Figures 3-5 show that gaps are more frequent in north-eastern than in north-western dialects, which may indicate that the diffusion of gaps is – in part – an areal phenomenon.

It is then fair to assume that the distribution of gaps and syncretism is due to both internal and external factors. In this respect, subject clitics are an interesting case study as they exhibit an extreme degree of variability although they are attested in a densely populated area and, diachronically, emerged in a relatively short diachronic span (from the 16th c. onwards). Hence, demic diffusion can-

not account for the complex geographic distribution of patterns of syncretism and gaps. At the same time, however, the array of pronominal forms discussed so-far does not provide conclusive evidence for feature-based models. As previously mentioned, linguistic systems may be shaped by external forces – 'cultural traditions' in Evans and Levinson's 2009 terms – that must be disentangled from biological constraints. For instance, how can we understand whether the 145 pattern results from biolinguistic constraints or is a cultural 'artefact'?

In order to answer the above question, we need a methodology to demonstrate that the systematic tendencies observed so far are not 'a matter of historical accident'. If so, we would predict a certain degree of correlation between linguistic and geographic distance: one might suppose that a given pattern emerged in a single dialect, for unknown reasons, and then spread to the surrounding area through language contact and sociolinguistic dynamics. Historically, the basin of the river Po and the surrounding mountains have always been a well-interconnected area, where people and goods circulated rather freely despite the geopolitical fragmentation. Given this socio-historical scenario, one would expect linguistic innovations to spread homogeneously in contiguous areas regardless of biolinguistic constraints on the make-up of pronominal inventories.

Alternatively, one may hypothesize that patterns of gaps and syncretism (e.g. the 145 pattern) are due to a biolinguistic constraint preventing or hindering the externalization of certain clitic forms. Then one would expect to find the same pattern scattered in non-contiguous dialects (Poletto's 2013 *leopard spots*), regardless of socio-historical factors.

In the remainder of the present section, I focus on the data contained in Appendix 1 (see Figure 11) to verify whether the microvariation displayed by clitic systems correlates or not with geographic distance.

Datapoint	1	2	3	4	5	6	expl	4=HOMO	SCL GAPS: # exponents	SCL GAPS: # gaps	SCL SYN: presence	SCl SYN: patterns
Olivone	1	1	1	1	1	1	1	0	6	0	1	145
Semione	1	1	1	1	1	1	1	0	6	0	1	145
Quarna sopra	1	1	1	1	1	1	1	0	6	0	1	145
Moncalvo	0	1	1	0	1	1	1	0	4	2	0	
Valmacca	1	1	1	0	0	1	1	0	4	2	0	
Breme	1	1	1	1	1	1	1	0	6	0	1	145
Castellinaldo	1	1	1	1	1	1	1	0	6	0	1	145
Inveruno	1	1	1	1	1	1	1	0	6	0	1	145
Carnago	1	1	1	1	1	1	1	0	6	0	1	145
Martignana di Po	1	1	1	1	1	1	1	0	6	0	1	145
Casorezzo	1	1	1	1	1	1	1	0	6	0	1	145
Arconate	1	1	1	1	1	1	1	0	6	0	1	145
Solbiate Arno	1	1	1	1	1	1	1	0	6	0	1	145
Càdero	1	1	1	1	1	1	1	0	6	0	1	145

Figure 11. Working table (upper part; see Appendix 1)

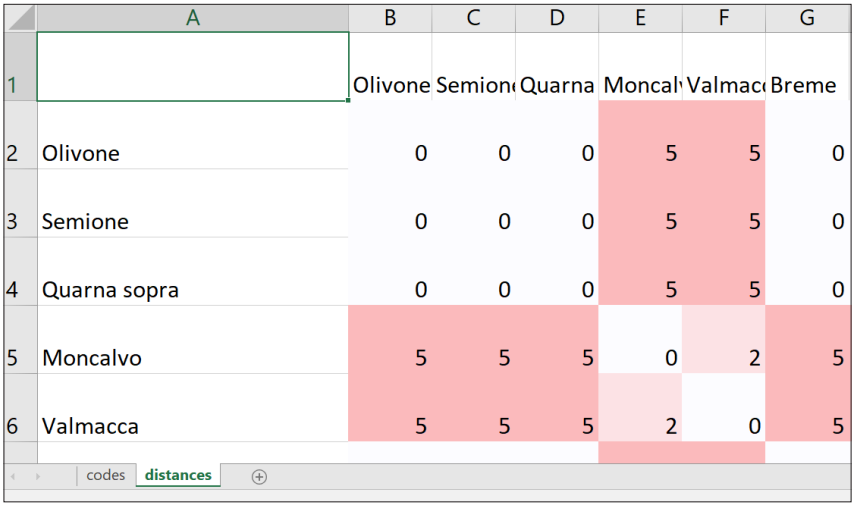
From the above table, I obtained a 'code' for each datapoint. The code is divided into two parts: the first six figures after the letter G(aps) represent gaps, whereas the figures after the letter S(yncretism) show whether the dialect has syncretic exponents and, if so, which Persons of the paradigms are involved in the syncretic pattern. For instance, a dialect with the code 'G023456S45' is a dialect in which the 1st person clitic is missing, while the 4th and 5th person are syncretic. The table containing the 'codes' of each variety is reported in Appendix 2 (Figure 12 shows the upper part of the table).

	1	2	3
1	**Datapoint**	**gaps+syncretism**	gaps
2	Olivone	G123456S145	123456
3	Semione	G123456S145	123456
4	Quarna sopra	G123456S145	123456
5	Moncalvo	G023056S	023056
6	Valmacca	G123006S	123006
7	Breme	G123456S145	123456
8	Castellinaldo	G123456S145	123456
9	Inveruno	G123456S145	123456
10	Carnago	G123456S145	123456
11	Martignana di Po	G123456S145	123456
12	Casorezzo	G123456S145	123456
13	Arconate	G123456S145	123456
14	Solbiate Arno	G123456S145	123456
15	Càdero	G123456S145	123456
16	San Benedetto Po	G123456S145	123456
17	Saguedo	G123456S145	123456
18	Stienta	G123456S145	123456
19	Revere	G123456S145	123456

Figure 12. Table with codes (Appendix 2)

Having a 'code' for each dialect, I calculated the linguistic distance between each pair of datapoints. The linguistic distance is calculated as the edit distance (or Levenshtein distance) between the two codes, i.e. the minimum number of operations (e.g. removal, insertion, or substitution of a character) to transform one string into the other. For instance, the edit distance between the dialect of Quarna sopra (G123456S) and Moncalvo (G023056S) amounts to 5 because two characters are substituted in the first part of the code and three are deleted from the second part.

Since edit distances are calculated pairwise, the result is a 187x187 symmetric matrix (187 is the number of datapoints), which is reported in Appendix 3; Figure 13 focuses on of the upper-left corner of the matrix:



	A	B	C	D	E	F	G
1		Olivone	Semion	Quarna	Moncalv	Valmac	Breme
2	Olivone	0	0	0	5	5	0
3	Semione	0	0	0	5	5	0
4	Quarna sopra	0	0	0	5	5	0
5	Moncalvo	5	5	5	0	2	5
6	Valmacca	5	5	5	2	0	5

codes distances

Figure 13. Section of the matrix of linguistic distances (see Appendix 3)

Having a matrix of the linguistic distances, I used the R package Geosphere (Hijmans 2017) to calculate the geographical distances between the same 187 datapoints of Manzini and Savoia's sample. The geographical distance is calculated on the basis of geographic coordinates and is therefore a geodesic distance, i.e. distance as the crow flies.

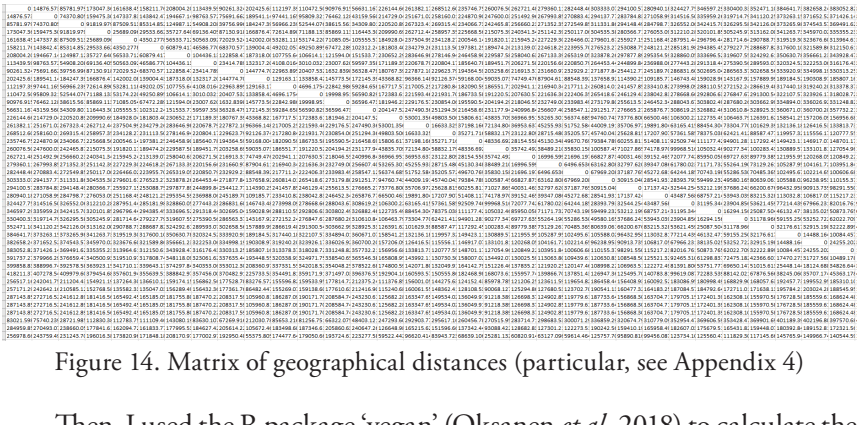


Figure 14. Matrix of geographical distances (particular, see Appendix 4)

Then, I used the R package 'vegan' (Oksanen *et al.* 2018) to calculate the correlation between the two matrixes: the matrix of linguistic distances and

the matrix of geographic distances (Mantel test). The result of the Mantel statistic is an index of 0.05931 (significance: 0.014), which means that there is no correlation between linguistic and geographical distances with respect to the inventories of subject clitics. Given such a low degree of correlation, it is fair to conclude that the robust tendencies found since Renzi and Vanelli's 1983 work cannot be accounted for under a pure geolinguistic explanation.

5. Discussion and conclusion

This article focused on the make-up of paradigms of subject clitics in northern Italian dialects. Subject clitics are a solid test bed to develop a methodology in order to evaluate internal vs external hypotheses on the emergence of linguistic variation.

Disentangling biological constraints from 'cultural' effects (*lato sensu*) is an aspect of linguistic research that, in my opinion, is still underdeveloped. Following Evans and Levinson 2009, it is fair to assume that '[s]triking similarities across languages […] have their origin in two sources: historical common origin or mutual influence, on the one hand, and on the other, from convergent selective pressures on what systems can evolve.' Hence, '[t]he dual role of biological and cultural-historical attractors underlines the need for a coevolutionary model of human language, where there is interaction between entities of completely different orders – biological constraints and cultural-historical traditions.' At present, however, we have no sound methodology to disentangle biological constraints from cultural-historical factors, in particular in the realm of *microvariation*, i.e. the study of genealogically-related languages. By studying genealogically-related languages, one always 'runs the risk [to] discover shared innovations that have purely historical explanations, rather than properties that are shared because of the same parameter setting.' (Haspelmath 2008: fn 8).

This objection has never been recast on the basis of empirical evidence because qualitative analyses do not provide any solid argument to reject the null hypothesis that microvariation is essentially chaotic. The null hypothesis is programmatically neglected by syntacticians, who prefer to support stronger hypotheses until falsification. This strategy is rewarding until the stronger hypotheses are reasonably falsifiable, but the increasing complexity of parametric models is hindering our capacity to analyse the enormous amount of empirical evidence we gathered.

For instance, in the last decades many data on subject clitics have been collected and, on the basis of these data, some solid tendencies have been found. These tendencies however yielded an unresolved tension between explanatory and descriptive adequacy. According to Manzini and Savoia (2005, I, 120) the data on gaps show that the Null Subject Parameter 'cannot be defined for the entire language, but must be applied to the individual forms

of the paradigm' (translation in Roberts 2014:178). Whereas Roberts 2014 argues against this radical microparametric approach, which would 'mak[e] the number of possible grammatical systems hyperastronomical'. Feature hierarchies might provide an intermediate explanation, by constraining the way in which (subject) agreement features are externalized across languages, but we need a methodology in order to assess the proposed models and tackle the following questions:

1) To what extent are the above empirical generalisations solid? Since no generalisation is exceptionless, it is worth knowing whether a given statement is true in 99% or 5% of the cases. In other words, we need to turn from categorical statements to probabilistic generalisations.
2) To what extent do the observed patterns result from random or extra-linguistic factors?

In this article, I provided (preliminary) statistical evidence to address the above questions. I confirmed that certain persons of the paradigm – in particular, the 145 cluster – exhibit the same behavior with respect to gaps and syncretism. Since this cluster is not a natural class found in other linguistic groups/families, one wonders about whether the above pattern results from a biolinguistic constraint or, alternatively, from random historical evolutions. To find out, I focused on the correlation between linguistic and geographical distance. I used a dialectometric approach to calculate pairwise linguistic distances regarding the structure of paradigms of subject clitics. Then I calculated the correlation between linguistic and geographical distances, which is surprisingly low. This means that the external explanation by itself cannot account for the observed cross-linguistic trends in the evolution of pronominal systems and some internal (biolinguistic?) constraints must be hypothesized along the lines of Heap 2002; Benincà and Poletto 2005; Oliviéri 2011; Calabrese 2011.

References

Benincà, Paola. 1994. *La variazione sintattica*. Bologna: Il Mulino.

Benincà, Paola, and Cecilia Poletto. 2004. "Topic, Focus and V2: Defining the CP Sublayers." In *The Structure of CP and IP: The Cartography of Syntactic Structures, Volume 2*, ed. by Luigi Rizzi 52-75. Oxford, MS: Oxford UP.

Benincà, Paola, and Cecilia Poletto. 2005. "The Third Dimension of Person Features." In *Syntax and Variation*, ed. by Leonie Cornips, and Karen P. Corrigan, 265-299. Amsterdam: Benjamins.

Barbosa, Pilar, M.E.L. Duarte, and Mary A. Kato. 2005. "Null Subjects in European and Brazilian Portuguese." *Journal of Portuguese Linguistics* 4 (2): 11-52.

Brandi, Luciana, and Patrizia Cordin. 1989. "Two Italian Dialects and the Null Subject Parameter." In *The Null Subject Parameter*, ed. by O. Jaeggli, and K. Safir, 111-142. Dordrecht: Kluwer.

Cabredo Hofherr, Patricia. 2004. "Les clitiques sujet du français et le paramètre du sujet nul." *Langue française* 141: 99-109.

Calabrese, Andrea. 2011. "Investigations on Markedness, Syncretism and Zero Exponence in Morphology." *Morphology* 21: 283-325.
Culbertson, Jennifer. 2010. "Convergent Evidence for Categorical Change in French: From Subject Clitic to Agreement Marker." *Language* 86: 85-132.
De Crousaz, Isabelle, and Ur Shlonsky. 2003. "The Distribution of a Subject Clitic Pronoun in a Franco-Provençal Dialect and the Licensing of Pro." *Linguistic Inquiry* 34 (3): 413-442.
Evans, Nicholas, and Levinson, Stephen. 2009. "The Myth of Language Universals: Language Diversity and Its Importance for Cognitive Science." *Behavioral and Brain Sciences* 32: 429-492.
Guardiano, Cristina *et al.* 2016. "South by South East. A Syntactic Approach to Greek and Romance Microvariation." *L'Italia Dialettale* 77: 96-166.
Haspelmath, Martin. 2008. "Parametric Versus Functional Explanations of Syntactic Universals." In *The Limits of Syntactic Variation*, ed. by T. Biberauer, 75-107. Amsterdam: Benjamins.
Heap, David. 2002. "Split Subject Pronoun Paradigms: Feature Geometry and Underspecification." In *Current Issues in Romance Languages: Selected Papers from the 29th Linguistic Symposium on Romance Languages* (LSRL), Ann Arbor, 8-11 April 1999, ed. by T. Satterfield *et al*, 129-144. Amsterdam: Benjamins.
Heap, David, and Michèle Oliviéri. 2013. "On the Emergence of Nominative Clitics in Romance Dialects", paper presented at the *7th European Dialect Syntax (EDiSyn) Workshop*, University of Konstanz (13-15 June 2013).
Holmberg, Anders. 2005. "Is There a Little *pro*? Evidence from Finnish." *Linguistic Inquiry* 36: 533-564.
Oksanen, Jari, F.G. Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, P.R. Minchin, R.B. O'Hara, G.L. Simpson, Peter Solymos, M.H.H. Stevens, Eduard Szoecs, and Helene Wagner. 2018. "Vegan: Community Ecology Package." R package version 2.5-3, <https://CRAN.R-project.org/package=vegan> (06/2019).
Hijmans, Robert J. 2017. "Geosphere: Spherical Trigonometry" R package version 1.5-7, <https://CRAN.R-project.org/package=geosphere> (06/2019).
Lobo, Maria, and Ana M. Martins. 2017. "Subjects." In *Manual of Romance Morphosyntax and Syntax*, ed. by Andreas Dufter, and Elisabeth Stark, 27-88. Berlin: De Gruyter.
Longobardi, Giuseppe, and Cristina Guardiano. 2009. "Evidence for Syntax as a Signal of Historical Relatedness." *Lingua* 119 (11): 1679-1706.
Loporcaro, Michele, and Diego Pescarini. *Forthcoming*. "Variation in Romance." In *Cambridge Handbook of Romance Linguistics*, ed. by M. Maiden. A. Ledgeway. Cambridge: Cambridge UP.
Manzini, M. Rita, and Leonardo M. Savoia. 2005. *I dialetti italiani e romanci. Morfosintassi generativa*. Alessandria: Edizioni dell'Orso.
Martins, Ana M., and Jairo Nunes. 2018. "Brazilian and European Portuguese and Holmberg's 2005 Typology of Null Subject Languages." Universidade de Lisboa, Universidade de São Paulo: ms.
Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. University of Chicago Press.

Oliviéri, Michèle. 2011. "Typology or Reconstruction: the Benefits of Dialectology for Diachronic Analysis." In *Romance Languages and Linguistic Theory 2009. Selected papers from 'Going Romance' Nice 2009,* ed. by J. Berns, *et al.*, 239-253. Amsterdam: Benjamins.

Palasis, Katerina. 2015. "Subject Clitics and Preverbal Negation in European French: Variation, Acquisition, Diatopy and Diachrony." *Lingua* 161: 125-143.

Pescarini, Diego. 2014. "La distribuzione dei clitici soggetto espletivi: tipologia e prospettive parametriche." *L'Italia dialettale* 75: 229-246.

Poletto, Cecilia. 2000. *The Higher Functional Field. Evidence from Northern Italian Dialects*. Oxford, MS: Oxford UP.

Poletto, Cecilia. 2013. "Leopard Spot Variation: What Dialects Have To Say About Variation, Change and Acquisition." *Studia Linguistica* 67 (1): 165-183.

Renzi, Lorenzo, and Laura Vanelli. 1983. "I pronomi soggetto in alcune varietà romanze." In *Scritti in onore di G.B. Pellegrini*, ed. by P. Benincà, *et al.*, 120-145. Pisa: Pacini.

Rizzi, Luigi. 1986. "On the Status of Subject Clitics in Romance." In *Studies in Romance Linguistics*, ed. by O. Jaeggli, and C. Silva Corvalan, 391-419. Dordrecht: Foris.

Roberts, Ian. 2014. "Subject Clitics and Macroparameters." In *Diachrony and Dialects: Grammatical Change in the Dialects of Italy*, ed. by P. Benincà, A. Ledgeway, and N. Vincent, 177-201. Oxford, MS: Oxford UP.

South, Andy. 2011. "Rworldmap: A New R package for Mapping Global Data." *The R Journal* 3 (1): 35-43.