



HAL
open science

Microvariation in agreement inflection. Subject clitics vs inflection

Diego Pescarini

► **To cite this version:**

Diego Pescarini. Microvariation in agreement inflection. Subject clitics vs inflection. Word Structure, 2022. <hal-03485387>

HAL Id: hal-03485387

<https://hal.science/hal-03485387v1>

Submitted on 17 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Microvariation in agreement inflection. Subject clitics vs inflection

Diego Pescarini

(CNRS, Université Côte d'Azur, BCL)

DRAFT

1. Introduction

Most Gallo-Romance languages (with the noticeable exception of Southern Occitan) and northern Italo-Romance languages exhibit so-called *subject clitics* (from now on: SCL).^{1,2} SCL are reflexes of nominative pronouns that occur adjacent to inflected verbs in tensed clauses. Whereas in some languages, such as standard French, the pronominal status of SCL is quite uncontroversial, in other languages their nature is more problematic. French, at least in its most formal registers, is a fully-fledged non-null subject language in which SCL are in complementary distribution with subject phrases. Conversely, Italo-Romance dialects show properties of null subject languages³ and SCL can often double phrasal subjects (Poletto 2000). Moreover, paradigms of SCL in Italo-Romance are highly irregular and, where interrogative inversion is still productive, the series of subject proclitics does not always coincide with that of enclitics that are found in main interrogatives. It is a highly held view that in the latter group of languages SCL are better analysed as agreement markers (see Rizzi 1986; Brandi & Cordin 1989), although similar conclusions may hold for the colloquial registers and certain diatopic varieties of spoken French (Culbertson 2010; Palasis 2015 among others).

The above issues are partly related to another research question that regards the relationship – in diachrony and in synchrony – between the make-up of verbal inflection (henceforth, INFL) and the emergence of SCL. Regardless of whether SCL are analysed as pronouns or agreement markers, the presence/absence of SCL and the property of *licensing* null subjects⁴ have been both linked to the *richness* of INFL.⁵ The intuition that SCL and INFL are (inversely) related has been debated since the works of preeminent 19th century linguists such as Meyer-Lübke 1895. To the best of my knowledge, however, the hypothesis of a systematic correlation between SCL and INFL has not been tested, but only tacitly assumed or cautiously mentioned as a tendency awaiting confirmation. In fact, the hypothesis that SCL and INFL are related is

¹ Acknowledgements.

² Throughout the article, the terms *language* and *dialect* will be used as synonyms.

³ For a recent collection of studies on null subjects and related issues, see Casalicchio & Cognola 2018.

⁴ See Taraldsen 1980 on the role of “rich” inflection in the definition of the so-called Null Subject Parameter, originally proposed by Perlmutter 1971.

⁵ Ideally, richness depends on the amount of distinctive forms across paradigms of finite verbs and across tenses, but in practice we still lack data and a feasible methodology to distinguish ‘rich’ from ‘poor’ INFL.

vague and needs to be refined. In this respect, it seems to me that at least four main research questions call for an answer:

- 1) Correlation: can we establish a statistical correlation between SCL and INFL? In my opinion, the possible relationship between SCL and INFL has remained an open question because it cannot be tackled with the canonical tools of *qualitative* linguistics and dialectology. Too many variables are in fact at play and no solid conclusion can be reached by focusing on single dialects. For this reason, we need to find a methodology to provide sound statistical measurements on a significant sample of languages.
- 2) Complementarity: are SCL in complementary distribution with distinctive forms of the verb? In fact, two possible relations may be established: SCL and INFL either exhibit isomorphic patterns of irregularities, sharing the same *morphomic* structure⁶ (in Maiden's 2005 terms; see Gaglia 2012 on Friulian dialects), or, alternatively, the irregularities in the two series (SCL and INFL) combine to maximize the overall distinctiveness of the system. According to the former hypothesis, one expects a positive correlation between the presence of syncretism (i.e. identity of exponence across paradigm cells) in INFL and the presence of syncretism and/or gaps in SCL. According to the latter hypothesis, SCL are expected to occur to *disambiguate* INFL forms that are otherwise indistinguishable. We therefore expect a negative correlation between the presence of syncretism in INFL and the presence of syncretism and/or gaps in the system of SCL.
- 3) Causation: did SCL emerge because of the loss of contrasts in the verbal system? It is a widely held view that changes in INFL – due to independent factors such as the loss of unstressed vowels – caused the emergence of SCL. This explanation reflects the supposed chronology of linguistic changes: whereas INFL was largely inherited from Latin, the syntactic change that turned free pronouns into SCL (Vedovato 2009) began around the 16th century, as documented by Vanelli 1998. Alternatively, however, one can assume, as Adams 1988 did, that it was the presence of SCL that allowed the loss of verbal endings, not vice versa.⁷
- 4) Constraints: are inventories of SCL constrained by independent grammatical principles? Following Renzi & Vanelli's 1983 generalisations, several works on subject clitics

⁶ A *morphome* (Aronoff 1994) is a morphosyntactic feature array (e.g. certain cells of the verb paradigm) that maps onto a single phonological exponent. Morphomes cannot be fully justified on morphosyntactic grounds.

⁷ In establishing a causal link between INFL and SCL, it is worth recalling that SCL were not all created equal, as Poletto 2000 argued, Consequently, not all SCL forms necessarily count with respect to the relation with INFL.

explored the hypothesis that gaps and syncretism in SCL must be explained in the light of grammatical principles, organised *iuxta propria principia* (Heap 2002, Cabredo Hofherr 2004; Benincà & Poletto 2005; Oliviéri 2011, Calabrese 2011). To the best of my knowledge, however, nobody has ever discarded the zero hypothesis, according to which shared irregularities have a purely historical explanation, in particular in closely-related dialects.

Recent scholarly works have tackled the above research questions in a principled way (Adams 1988; Sibille 2012; Roberts 2014 among others), but none from a quantitative standpoint. Sibille 2012, working on a small set of Occitan and Italo-Romance dialects, shows that distinctive SCL and INFL markers tend to occur in complementary distribution, thus optimising the number of distinctive combinations of SCL and INFL. Moreover, he argues that the kind of variation we observe with respect to SCL systems confirms the *death* (Heap 2002: 152) of the Null Subject Parameter. Roberts 2014, on the contrary, claims that the ratio between SCL and INFL distinctive forms is variable – all kinds of possibilities are attested – and this is not evidence against the Null Subject Parameter, at least in its more recent reformulation as a parameter hierarchy, rather than as a single yes/no choice. Whatever the claim – and the theoretical implications – it seems to me that correlations between SCL and INFL systems need to be tested against a representative sample of cases to avoid availability biases. This article will therefore compare paradigms of SCL and (regular) INFL in a sample of 187 Italo-Romance varieties taken from Manzini & Savoia 2005/I: 69-117.

The structure of the article is as follows: §2 elaborates on Roberts's 2014 typology of SCL systems; §3 illustrates and justifies the methodology adopted in the present study; §4 examines Manzini & Savoia's 2005 data in the light of a modified version of Roberts's 2014 taxonomy; §5 focuses on non-redundant systems (i.e. dialects in which the combined number of SCL and INFL distinctive forms equals person distinctions); §6 elaborates on the areal diffusion of SCL and INFL inventories. §7 concludes.

2. Roberts 2014

In most northern Italo-Romance varieties, agreement is marked redundantly on both the subject pronoun (which occurs obligatorily even if a phrasal subject is present) and on the verb:

- (1) **a** 'dørmi 'I sleep' (Carnago, Lombardy)
te 'dørmət 'you sleep'
ul/la 'dørmɜ 'he/she sleeps'
a ' dørmum 'we sleep'
(a) **dur**'mi 'you sleep'
(a) **i** 'dørman 'they sleep'

In most dialects, however, both pronouns and verbal endings exhibit patterns of defectivity and syncretism. For instance, in (1), as well as many other northern Italian dialects, 1st person (singular and plural) and 2nd person plural are all marked by the same SCL *a* (for the sake of simplicity, I will refer to 1st/2nd/3rd person plural forms as 4P, 5P, 6P). In other areas, such as in the dialects in (2), the clitics of the 1/4/5P are missing. Similar considerations hold for INFL: verbal endings are often missing – especially in the dialects in which final unstressed vowels have been lost – or syncretic, as the 3P and 6P in (2).

- (2) — 'dørmo 'I sleep' (Verona, Veneto)
te 'dørmi 'you sleep'
e/la 'dørme 'he sleeps'
— dor'mimo 'we sleep'
— dor'mi 'you sleep'
i/le 'dørme 'they sleep'

The combinations of gaps and syncretisms in SCL and INFL yield a high degree of cross-linguistic variation. Roberts 2014: 195-199 tries to put some order with the intent of showing that variation in SCL/INFL systems is not compelling evidence against higher-grade generalisations such as the Null Subject Parameter.

To build his argument, Roberts classifies dialects on the basis of a four-way taxonomy, depending on whether SCL and/or INFL exhibit a 'full set of morphological person-number distinctions, where 'full' means at least five distinct forms, and zero counts as distinct'^{8, 9}

⁸ To the best of my knowledge, Roberts's taxonomy of INFL is based exclusively on data from present tenses. As an additional criterion, I assumed that 3P and 6P are syncretic iff both masculine and feminine forms are identical.

⁹ I think that it is worth adopting and comparing two different kinds of coding: a Roberts-style coding, where zeroes are distinctive, and an alternative coding in which zeroes are non-distinctive even if a language lacks a single SCL form. Under the hypothesis that SCL emerged *in order to* disambiguate identical INFL, a missing SCL means that no additional marking was needed at the time in which SCL emerged. Gaps in SCL must be therefore regarded – according to the hypothesis under testing – as a dependent variable that is conditioned by the make-up

(Roberts 2014: 196). Four prototypical systems are therefore predicted (see (3), where [+agr] denotes full sets) and a possible correlation with the Null Subject Parameter emerges: non-null subject languages (e.g. French) are in fact expected to exhibit an asymmetry between a rich SCL system and a poor INFL system:

- (3) a. SCL[+agr] INFL[+agr] ‘fully redundant’, null subject system e.g. Carnago in (1)
 b. SCL[+agr] INFL[-agr] non-null-subject system e.g. French
 c. SCL[-agr] INFL[+agr] a non-redundant null-subject system e.g. Verona in (2)
 d. SCL[-agr] INFL[-agr] (usually) a complementary system (cf. §5)

All types are attested in Italo-Romance, but the vast majority of northern Italian dialects belong to Type (c) and (d).

- (4) Distribution of Italo-Romance dialects (source: Manzini & Savoia 2005) according to Roberts’s 2014 typology

7	SCL[+agr]	INFL[+agr]	a. ‘fully redundant’, null subject system
9	SCL[+agr]	INFL [-agr]	b. non-null-subject system
95	SCL[-agr]	INFL [+agr]	c. non-redundant null-subject system
76	SCL[-agr]	INFL [-agr]	d. (usually) a complementary system

Dialects of type (b) – which, under Roberts’s classification, are expected to exhibit properties of non-null subject languages – are found in Lombard dialects such as Grumello, Borgo di Terzo (both in the province of Bergamo) Caveragno, Soazza, Casaccia, Soglio (spoken in southern Switzerland, cf. (5)), in the Piedmontese dialects spoken in Fontane and Felizzano, and in the Franco-Provencal island of Celle San Vito.

- (5) **i** 'dro:m 'I sleep' (Soglio, Southern Switzerland)
ty 'dro:m 'you sleep'
al/la 'dro:m 'he/she sleeps'
am 'dro:m 'we sleep'
u dru'mi 'you sleep'

of INFL and by how INFL and SCL inventories interact. For this reason, I examined the statistical measures obtained from a system of coding in which zeroes in SCL and INFL are not coded alike, but the results in the two conditions have never been significantly different.

i/la 'dromɐŋ 'they sleep'

The typology in (4) is – in Roberts’s own words – “a purely descriptive device” that, in fact, does not make the right predictions regarding the null/non-null status of Italo-Romance dialects, which all behave like null subject languages, including the dialects such as (5).

As a first, preliminary conclusion, I would therefore suggest that the richness of INFL and SCL is not a solid indicator of the null/non-null subject status of a language. However, the distinction between redundant and non-redundant systems may be refined in order to prove whether a sound relationship can be established between INFL and SCL. To do so, however, we first need to exclude orthogonal phenomena that can blur statistical analyses.

3. Restricting the sample

Agreement endings are not the only way in which person distinctions are encoded in the verbal system. In fact, both regular and irregular verbs show systematic patterns of stem allomorphy and suppletivism that set apart 4/5P forms. The following data from Veronese show that 4/5P are characterised by stress shift and, with certain irregular verbs, by stem suppletion:

(6)	a Regular verb	b Irregular verb	(Veronese)				
	_ m̀agno ('I eat')	_ v̀ago ('I go')					
	te m̀agni	te v̀è					
	el m̀agna	el va					
	<table border="1"><tr><td>_ magnémo</td></tr><tr><td>_ magnì</td></tr></table>	_ magnémo	_ magnì	<table border="1"><tr><td>_ némo</td></tr><tr><td>_ v̀è</td></tr></table>	_ némo	_ v̀è	
_ magnémo							
_ magnì							
_ némo							
_ v̀è							
	i m̀agna	i va					

Allomorphy in the verbal paradigm is a trait that the Romance languages have inherited from Latin, see (7). Hence, 4/5P INFL are (and have always been) clearly distinguishable from the rest of the paradigm also in dialects with ‘poor’ inflection.

(7) EGO CANTO 'I sing'
TU CANTAS
ILLE CANTAT

NOS CANTĀMUS
VOS CANTĀTIS

ILLI CANTANT

Another complication regarding the 4P is that in several dialects of Italy the 4P results from the reanalysis of impersonal periphrases. Current 4P forms therefore feature either the impersonal/reflexive clitic *si* (as in Florentine) or a reflex of the impersonal subject HOMO (Lat. for ‘man’) followed by a verb at the 3P:

- (8) a. **e** **si** 'dɔrme (Firenze)
 SCL= *si*=sleep.3SG
 ‘We sleep’
- b. **am** 'dro:m (Soglio)
 ‘man’= sleep.3SG
 ‘We sleep.’

SCL seem to reflect the idiosyncratic behaviour of 4/5P INFL, but the interaction between SCL and INFL at the 4/5P lends itself to contradictory explanations. It is a well known fact that in most northern Italian dialects 4/5P SCL are either missing or syncretic, cf. (2) and (6) (more on this in §6). Some would therefore suggest that 4/5P clitics are either syncretic or missing at the 4/5P *because* the corresponding verbal forms are intrinsically distinctive. Others, however, might claim that 4/5P SCL and verbs have idiosyncratic forms because they share the same morphomic pattern (Gaglia 2012) or because they are targeted by the same constraints in a rule-based model of the syntax/phonology interface such as Calabrese 2011. Others would object to both explanations by highlighting the fact that no correlation between 4/5P SCL and INFL can be established because in most dialects in which 4/5P SCL are syncretic or missing the 1P SCL is syncretic or missing as well, cf. (2) and (6), more on this in §6. In conclusion, not only cannot 4/5P INFL be compared with other INFL markers all other conditions being equal, but also the relationship between SCL and INFL at the 4/5P does not provide conclusive evidence to discard any of the many theoretical options still available.

In the sections that follow, I will therefore focus on the remaining persons, which are expected to be more easily comparable across the dialects of our sample. In this respect, however, several Lombard varieties exhibit a further complication that needs elaboration. In certain dialects, etymological INFL endings had been lost and successively replaced by suffixed

SCL:¹⁰ the latter have been later reanalysed as fully fledged INFL markers such as *-i* for the 1P (from EGO), *-t* for the 2P (from TU), *-v/f* for the 5P person (from VOS),¹¹ e.g.

(9) a. **a** 'dørm-**i** (Carnago; Manzini & Savoia 2005 I: 74)

SCL sleep-1SG (<“I”)

‘I sleep’

b. **te** 'dørm-**et**

SCL sleep-2SG (<“you”)

‘you sleep’

Given the goals of the present study, one should separate the dialects in which endings *were* SCL from the other dialects in which INFL are regular reflexes of Latin endings. For this purpose, a binary numerical index has been added to the dataset (see the column “**endings<pro**” in Appendix A) to single out the varieties in which present-day INFL originated from enclitic pronouns.

Before concluding the section, a last caveat is in order concerning the verbal declension that has been used to calculate the number of distinctive INFL forms. In comparing SCL and INFL I will build on Manzini & Savoia’s 2005 data, which were elicited through a fixed questionnaire. Fortunately, the questionnaire contains the full paradigm of a regular verb – i.e. a verb that usually shows no person-driven allomorphy or suppletion – in the simple present and in the present perfect (the *passato prossimo*, according to the Italian terminology, which in northern dialects is the only available past tense). The quantitative analysis that follows is entirely based on the comparison between SCL and INFL in the present conjugation of a regular verb. Quantitative indexes, however, are supposed to vary considerably if irregular verbs such as auxiliaries would have been examined instead. One might object that high-frequency irregular verbs are more salient models of INFL systems than low-frequency regular verbs. I acknowledge the problem, but, for reasons of feasibility, I decided to focus entirely on the INFL

¹⁰ In some dialects, such “suffixed enclitics” are only found in interrogative contexts, where many northern Italian dialects exhibit remnants of verb/subject inversion. However, in certain Lombard dialects (those where the loss of final vowel caused a massive neutralisation of INFL), such “suffixed enclitics” became fully fledged INFL, which are nowadays always suffixed to verbal forms in any clausal environment.

¹¹ In this respect, the reconstruction of the morphology of the 4th person is particularly complicated as the etymological ending has been often hybridized with the enclitic pronoun HOMO occurring in the impersonal construction: cf. CANTAMU(S) ‘we sing’ vs CANTA(T) HOMO ‘one/we sing(s)’. The evolution of the two forms is strictly intertwined and, for most varieties, it is impossible to tease them apart.

endings of regular verbs and leave for further research the analysis of the SCL/INFL correlation in irregular verbs, including auxiliaries.

In conclusion, to compare a homogeneous sample of dialects and phenomena, in §§4-5 I will focus on the correlation between SCL/INFL at the 1-3P and 6P of regular verbs in the present tense in declarative clauses. Dialects in which present-day INFL endings are reflexes of SCL will not be excluded from the sample, but they will be examined separately.

4. Redundant vs non-redundant systems

This section aims to compare the structure of paradigms of SCL and finite verbs excluding 4/5P forms, which, as previously mentioned, are often singled out by patterns of allomorphy and suppletion. I first adopt a Roberts-style classification in order to test the hypothesis according to which SCL *compensate* the loss of contrasts in the verbal paradigm. To do so, we need to compare the number and array of distinctive forms in the paradigms of SCL and verbs to see whether the two correlate.

First of all, we need to calculate the number of non-distinctive forms in the verb paradigm; then, I will verify whether the number of SCL is equal/higher/lower than (or directly/inversely proportional to) the number of distinctive INFL forms. For instance, Veronese exhibit two non-distinctive verb forms (at the 3P and 6P), therefore Veronese is expected to exhibit at least one SCL in order to distinguish the third from the sixth person.

- (10) _ 'dɔrmo 'I sleep' (Verona)
te 'dɔrmi 'you sleep'
el 'dɔrme 'he sleeps'
i 'dɔrme 'they sleep'

Following Roberts's terminology, dialects such as Veronese are *redundant* because all SCL forms (including zero in the 1P) are distinctive and INFL exhibits only two syncretic forms (3P=6P).

In the following table, the 187 dialects of Manzini & Savoia's sample are organized in a 4x4 matrix on the basis of the number of contrastive forms in the INFL and SCL paradigms, respectively. Having two subparadigms of four cells (1P, 2P, 3P, 6P), the range of possible contrasts goes from 0 (all forms are identical) to 3 (all forms are different). A dialect like

Veronese in (10) is therefore classified as a dialect lacking one contrast in INFL (because 3P and 6P are syncretic) and 3 contrasts in SCL (because a single zero in the paradigm counts as distinctive, as in Roberts’s classification). Each cell of the table in (11) reports the count of dialects per type and the incidence of each type in Manzini & Savoia’s 2005 sample.

(11) Counts and incidence of dialects based on number of distinctive SCL and INFL (Sample: 187 dialects; a single zero counts as distinctive)¹²

INFL\SCL	0	1	2	3
0	-	-	-	20 (.10)
1	-	-	15 (.08)	20 (.10)
2	-	2 (.01)	19 (.10)	70 (.37)
3	-	1 (.00)	18 (.09)	21 (.11)

Grey cells correspond to redundant systems (in Roberts’s terms), i.e. languages in which the total amount of contrasts (INFL + SCL) is more than three. The dialects in the three white cells at the centre of the table (11) are *non-redundant* systems in which the total number of contrasts is equal to or lower than three. In these dialects, SCL and INFL must occur in complementary distribution to avoid syncretism across Persons (whether it is the case or not will be verified in §5). All the other possible systems are unattested in Manzini & Savoia’s sample. Crucially, they correspond to systems in which the number of distinctive INFL and SCL would be insufficient to distinguish the four persons of our sub-paradigm.

The following Roberts-style table show the aggregate number of dialects with redundant systems and non-redundant systems:

¹² If we exclude the dialects in which INFL endings derive from pronominal forms (see §3), we observe a predictable – but not particularly significant – lower incidence of redundant systems:

INFL\SCL	0	1	2	3
0	-	-	-	11,19%
1	-	-	10,49%	11,89%
2	-	1,40%	12,59%	32,17%
3	-	0,70%	10,49%	9,09%

A: 11,19% C

B: 55,24%

(12) Types of Italo-Romance SCL/INFL systems, excluding 4/5P (source: Manzini & Savoia 2005)

22	SCL[+agr]	V[+agr]	highly redundant (6 contrasts)
110	SCL[+agr]	V[-agr]	mildly redundant (4-5 contrasts) with more contrasts in SCL than in INFL
19	SCL[-agr]	V[+agr]	mildly redundant (4-5 contrasts) with more contrasts in INFL than in INFL
36	SCL[-agr]	V[-agr]	non-redundant (3 contrasts)
0	SCL[-agr]	V[-agr]	Non-redundant and underspecified (less than 3 contrasts)

In conclusion, none of the 187 dialects in Manzini & Savoia's 2005 sample is underspecified, i.e. in no dialect the total number of distinctive SCL and INFL markers is lower than Person distinctions. This confirms previous results based on the analysis of narrower samples of dialects such as Sibille 2012.

Whereas underspecification is absent, overspecification is widespread: 151 dialects out of 187 exhibit more distinctive exponents than necessary. Overspecification usually results from a greater number of SCL than INFL markers (this happens in 110 out of 187 dialects).

These preliminary results confirm the hypothesis that SCL and INFL are somehow linked, but the relationship is not biconditional: non-distinctive INFL is a sufficient but not necessary condition for the occurrence of SCL.

In causal terms, this indicates that the emergence of SCL may have been triggered by the impoverishment of INFL, but the number and array of SCL is not predictable from the number and array of distinctive INFL (more on this in §6).

5. Non-redundant systems are complementary

In the 36 dialects that exhibit complementary systems (see the Table in (12)), SCL and INFL are expected to be combined in order to distinguish the four persons of the subparadigm we are investigating. Therefore, we want to know whether the interplay of distinctive SCL and INFL tends to maximize the number of contrasts or not. To illustrate the point, let us compare the two systems in (13): both have the same number of syncretic exponents in INFL and SCL (1 means that the marker is distinctive; 0 means non-distinctive; a single gap in the paradigm is

distinctive). (13a) is a *complementary* system in which each person is marked by a distinctive combination of SCL/INFL elements: certain persons are marked by a distinctive SCL (SCL[1] INFL[0]), others by a distinctive V (SCL[0] INFL[1]), others by a combination of two distinctive forms (SCL[1] INFL[1]), others by a combination of two forms that *intraparadigmatically* (i.e. within the paradigm of either SCL or INFL) are not distinctive, but yield an *interparadigmatic* distinctive combination (SCL[0] INFL[0]). Conversely, (13b) is a *non-complementary* system because two persons of the paradigm are syncretic, inter- and intra-paradigmatically.

(13) a. Complementary system (ED=3)

	1P	2P	3P	6P
INFL	0	0	0	1
SCL	1	1	0	0

b. Non-complementary system (ED=1)

	1P	2P	3P	6P
INFL	0	0	0	1
SCL	1	0	0	1

To find whether two systems are distinctive or not, I calculated the *Edit Distance* (ED) between the strings of binary values representing SCL and INFL, respectively. ED is the minimum number of insertions, deletions or substitutions needed to transform one string in another string. In non-redundant system of the type 1:2 such as those in (13) (i.e. dialects with one distinctive SCL and two distinctive INFL or viceversa), distinctive systems have an ED of 3, whereas non-distinctive systems have an ED of 1. Analogously, in 2:2 systems – with two distinctive INFL and two SCL – the edit distance allows us to distinguish complementary systems (with ED = 2 or 4) from non-complementary systems (ED = 0), see (14).

(14) a. Complementary system (ED=4)

	1P	2P	3P	6P
INFL	0	1	1	0
SCL	1	0	0	1

b. Complementary system (ED=2)

	1P	2P	3P	6P
INFL	0	1	1	0
SCL	1	0	1	0

c. Non- complementary system (ED=0)

	1P	2P	3P	6P
INFL	0	1	1	0
SCL	0	1	1	0

By combining the value of the ED with the number of distinctive items in INFL and SCL paradigms, we can establish whether non-redundant systems are complementary or not. The 36 non-redundant systems attested in Manzini & Savoia's sample are organized in table (15) according to two parameters: Edit Distances and number of contrasts in SCL and INF, respectively. Asterisks mark impossible distances; grey cells correspond to non-distinctive systems. The data show that all 36 non-redundant systems are distinctive:

(15) Counts of dialects with non-redundant INFL/SCL systems. Asterisks mark impossible distances; grey cells correspond to non-distinctive systems

Type\ED	0	1	2	3	4
SCL[2] INFL[1]	*	0	*	15	*
SCL[1] INFL[2]	*	0	*	2	*
SCL[2] INFL[2]	0	*	15	*	4

The table in (15) shows that, in non-redundant systems, distinctive forms of SCL and INFL are in complementary distribution as SCL occurs where INFL is syncretic or vice versa. The data confirm previous conclusions of qualitative studies (Sibille 2012:§5), but it is worth recalling that no mechanism of complementarity seems at play in the vast majority of dialects, which are characterized by redundant SCL/INFL systems (§4).

The complementarity observed in non-redundant system, coupled with the remarkable absence of underspecified systems, supports the hypothesis that the distribution of distinctive forms in SCL and INFL is somehow optimised, although nothing prevents the emergence of redundancies.

In the light of this interim conclusion, we can start answering some of the research questions mentioned in §1:

- 1) Correlation: the data analysed so far confirm that SCL and INFL are somehow linked. The absence of underspecified system in a sample of 187 dialect appears to be a rather solid conclusion. No direct correlation, however, can be established between SCL and INFL as the number of items in the two paradigms is neither directly nor inversely proportional.
- 2) Complementarity: we can discard the hypothesis that SCL and INFL exhibit isomorphic patterns of irregularities and conclude instead that SCL and INFL are in complementary distribution in non-redundant systems, which however are attested in 20% of the dialects of Manzini & Savoia's sample.
- 3) Causation: non-redundant systems seem to suggest that SCL emerged to compensate the loss of contrasts in the verbal system, but the same principle cannot account for the vast majority of dialects, in which the combined INFL/SCL inflectional system is overspecified. In this respect, a simple cause-effect relation linking the impoverishment of INFL and the emergence of SCL appears too simplistic. Alternatively, one could assume, following Adams 1988, that the presence of a rich inventory of SCL triggered (or allowed) the loss of verbal endings, not vice versa. The fact remains, however, that the loss of verbal endings resulted from independent and well-documented phonological changes, which are expected to take place regardless of the presence of SCL. Furthermore, evidence against Adams's 1988 claim is brought by those Lombard dialects in which present-day endings are grammaticalised reflexes of enclitic pronouns (cf. §3). If the presence of SCL had triggered/allowed the loss of verbal endings, what triggered/allowed, in the very same languages, the emergence of INFL markers originating from SCL?

In conclusion, it seems to me that the data introduced so far do not provide conclusive evidence for the causal analysis of the emergence of SCL (question 3). Analogously, the comparison between SCL and INFL does not shed light on the nature of the constraints shaping (redundant) SCL systems, which are characterised by quasi-regular patterns of gaps and syncretisms. Such solid tendencies, which have been debated since Renzi & Vanelli 1983, will be discussed in the next section.

6. Typological vs geographical dispersion

So far, dialects have been analysed as isolated systems, although they co-exist in a continuous space where waves of linguistic innovations are free to spread. Hence, when we look at the distribution of a certain phenomenon across genealogically-related languages that are spoken in a contiguous area, we cannot abstract away completely from external factors that may compromise the statistical significance of quantitative analyses. In fact, one may assume that “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). In my opinion, Nichols’s view is too drastic. I would rather assume, following Evans & Levinson 2009 that internal and external factors are both at play, but never annihilate each other. Extralinguistic factors (“cultural factors” *lato sensu*) may allow certain grammatical traits to spread from a language to another, but internal factors (“biological factors” *lato sensu*) are supposed to hinder the diffusion of marked features/phenomena and enhance the diffusion of unmarked ones. Unmarked traits are then expected to spread more readily than marked ones, but, at the same time, one cannot exclude that a marked trait could become prevalent if it is perceived as prestigious. At the same time, internal factors are expected to trigger/allow the emergence of similar, unmarked phenomena in discontinuous areas (*leopard spots*, in Poletto’s 2013 terminology) via independent, internal change.

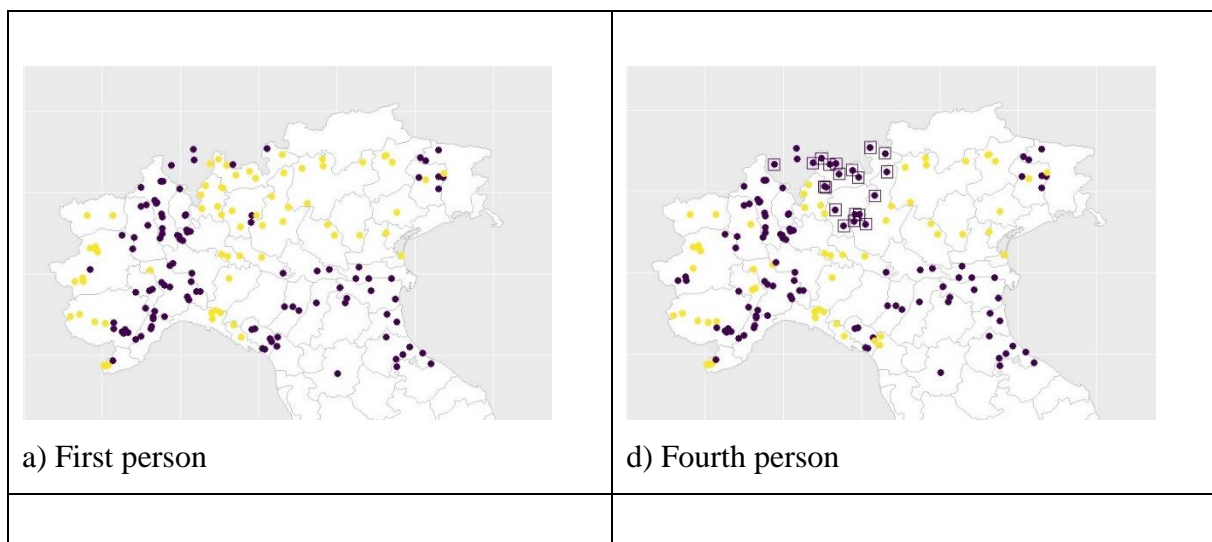
Whether inventories of SCL are constrained by inner factors or are ‘historical accidents’ is still an open question. Benincà & Poletto 2005, for instance, argue that the distribution of “vocalic” SCL in northern Italian dialects is constrained by internal principles. Vocalic clitics are uninflected SCL that, despite being called ‘vocalic’, are singled out by syntactic and pragmatic properties (Poletto 2000): they often merge with complementizers, they may precede inflected SCL, they normally precede the preverbal negator (if present), they are not inverted, in some dialects they turned into discourse particles (Benincà 1983). Vocalic SCL, which – etymologically – are supposed to be reflexes of the Latin pronoun EGO ‘I’ – are often syncretic. In many dialects, vocalic clitics are found at the 1P and 4P:

- (16) a. **A mangiu** (Oneglia, Piedmont)
 ‘I eat.’
 b. **A mangiammu**
 ‘We eat.’

In other dialects such as Bolognese in (17), the vocalic SCL occurs at the 1P, 4P, and 5P:

- (17) a. **A magn** (Bologna, Emilia-Romagna)
 ‘I eat.’
 b. **A magnén**
 ‘We eat.’
 c. **A magnè**
 ‘You.PL eat.’

As Benincà & Poletto 2015 pointed out, vocalic clitics yield a narrow number of possible patterns of syncretism. In Manzini & Savoia’s sample, for instance, 80% of the dialects show a syncretic exponent for the 1P and 4P (as in (16)), whereas a remarkable 67% of the dialects have a single syncretic exponent for the 1P/4P/5P (as in (17)). Analogous considerations hold true for gaps. The distribution of gaps (= missing SCL) in declarative clauses¹³ is plotted in the following six maps, one for each Person (yellow points mean that the dialect exhibits a gap). 2P, 3P and, to a lesser extent, 6P SCL are almost always present, whereas 1P, 4P and 5P SCL are often missing. It is worth noting that in many Lombard dialects – those marked by a square in Fig. 2d – the 4P person results from the reanalysis of an impersonal periphrasis formed by the clitic *om* < HOMO followed by the verb at the third person (cf. §3). These dialects therefore lack a specific/etymological form for the 4P. If we discard the squared points in the map of the 4P person, the areas where 1P, 4P and 5P person clitics are attested (or missing) partially overlap.



¹³ The array of SCL in interrogative clauses may not correspond to the one in declaratives, as we know since Renzi & Vanelli’s 1983 preliminary survey.

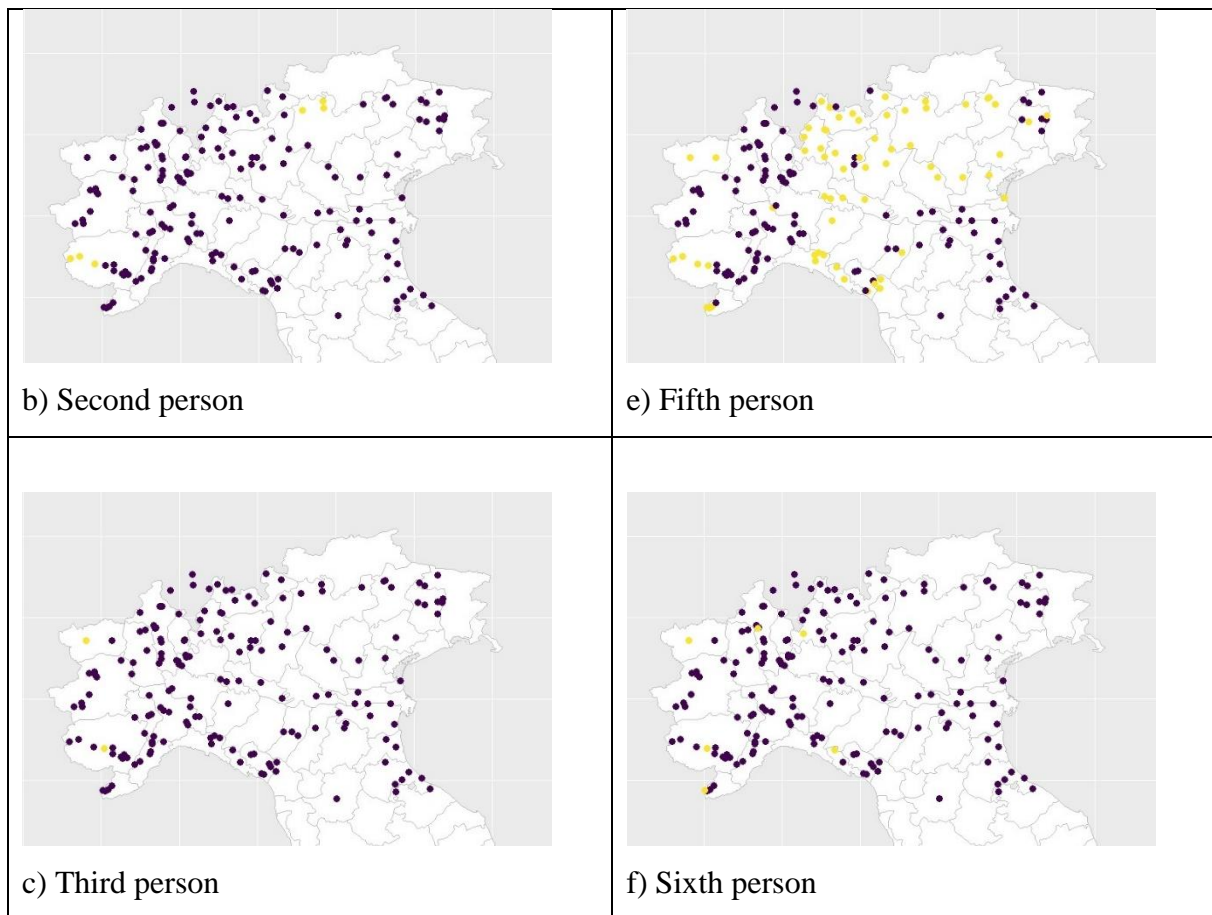


Figure 2 - absence vs presence of subject clitics.

Sample: Manzini & Savoia 2005 (187 datapoints)

Legenda: ● absence vs ● presence of subject clitics

According to the above data, the cross-linguistic distribution of syncretisms and gaps is not chaotic: 1P, 4P and 5P SCL are ‘shy’ (in Calabrese’s 2011 terms): in northeastern dialects these SCL are missing, while in northwestern dialects they often have a syncretic vocalic exponent.

Higher grade generalizations have then been proposed in the attempt of deriving the systematic distribution of gaps and syncretism from grammatical principles such as feature geometries or filters on Person distinctions, which constrain how person and number features are externalised (Heap 2002, Cabredo Hofherr 2004; Benincà & Poletto 2005, Floricic and Molinu 2008, 46-47; Oliviéri 2011, Calabrese 2011). However, within a co-evolutionary model of morphosyntactic change, in which grammatical systems are shaped by both “biological” and “cultural” factors, one may suggest that certain patterns are more widespread than others not because they instantiate a less marked feature configuration, but because certain arrays spread over contiguous areas for external reasons, e.g. sociolinguistic dynamics.

In this section, I try to discard the latter hypothesis by showing that the geolinguistic distribution of SCL is more scattered than INFL systems, although SCL systems are less

variable than INFL systems. In a nutshell, I intend to demonstrate that SCL systems are geographically dispersed, but typologically homogeneous, whereas INFL systems are typologically diverse, although they are distributed more homogeneously than SCL in contiguous geolinguistic areas.

To illustrate typological dispersion, the following unrooted dendograms show how SCL and INFL systems are clustered. On the basis of the distribution of distinctive and non-distinctive items (0/1) we can calculate the Euclidean distance between each paradigm type (e.g. 1111, 1100, 1011, etc.) and plot these distances in a bidimensional space: the longer the line connecting two leaves, the greater the distance between systems. The dendogram in Fig. 3 and 4 represents distances between INFL and SCL systems in Manzini & Savoia's sample: the graph in Fig. 3, representing INFL systems, has more leaves than the one in Fig. 4, representing SCL; distances are therefore shorter in INFL and no INFL system stands out clearly. Systems of SCL, conversely, are less variable and dialects with a full set of distinctive SCL (in the high right corner of the dendogram) are clearly set apart from the others.

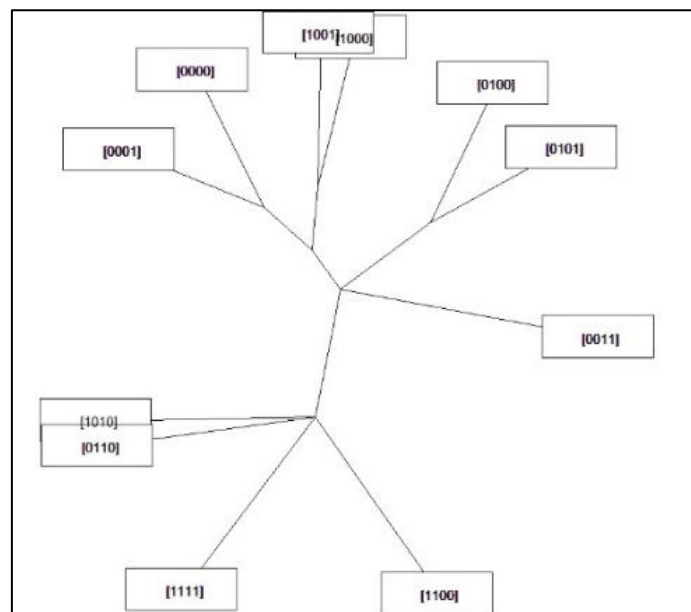


Fig. 3 - Cluster analysis of INFL. 1 and 0 correspond to, respectively, distinctive and non distinctive forms in a four-cells paradigm (1-3P, 6P). Each leaf is a type of system, e.g. 1100, 1110, etc.; lines are proportional to distances) Sample: 187 dialects from Manzini & Savoia 2005.

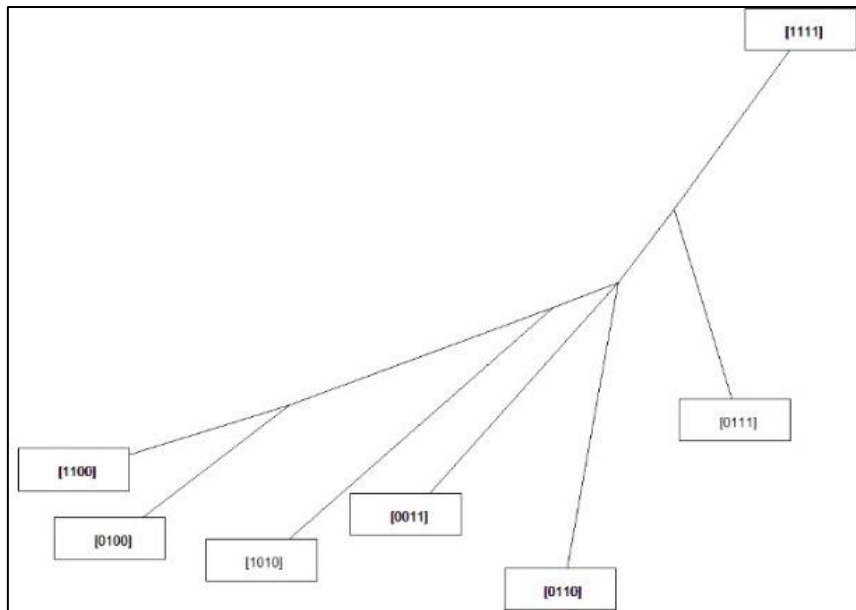


Fig. 4 - Cluster analysis of SCL. 1 and 0 correspond to, respectively, distinctive and non distinctive forms in a four-cells paradigm (1-3P, 6P). Each leaf is a type of system, e.g. 1100, 1110, etc.; lines are proportional to distances) Sample: 187 dialects from Manzini & Savoia 2005.

The difference between INFL/Fig. 3 and SCL/Fig. 4 lends itself to two alternative explanations. One may suggest that SCL are less variable because SCL, unlike INFL, are constrained by markedness principles such as feature geometries and the like. Alternatively, one may object that SCL are less variable because they emerged in the early modern period (Vanelli 1998), whereas INFL are reflexes of Latin endings and, as such, had been drifting for centuries before SCL emerged. Hence, typological dispersion does not provide conclusive evidence to disentangle internal from external factors of variation.

A possible solution may be to compare typological dispersion with geographic dispersion. Let us assume that innovations spread as waves across genealogically-related dialects that are spoken in contiguous areas. This kind of diffusion is expected to be geographically regular: varieties that are closer to the centre of diffusion of the wave will be more innovative than peripheral areas, which will retain more conservative traits (or will develop their own, isolated innovations). By the same token, one expects INFL systems, which have been subject to many successive waves of innovations (coming from various directions), to have a rather discontinuous geographical diffusion. On the contrary, innovations that are relatively more recent such as SCL are expected to be distributed in contiguous area. If it was not the case, one would suggest instead that SCL are geographically dispersed, but typologically similar (cf. Fig. 4), because they are constrained by internal principles that allowed the emergence of similar systems in discontinuous areas, regardless of socio-historical factors.

To test this hypothesis, I calculated the correlation between linguistic and geographical distances for both SCL and INFL. If patterns of SCL were as geographically dispersed as (or less dispersed than) patterns of INFL, one would conclude that paradigms of SCL are not affected by internal factors and their make-up is a “historical accident”. Conversely, if types of SCL paradigms were more dispersed than INFL, one would conclude that internal factors are likely to be at play in SCL inventories.

To find out, I calculated the geographical (geodesic) distances between 63 datapoints of Manzini & Savoia’s 2005 sample. I selected only dialects spoken in the basin of the River Po (hence, Lombard, Piedmontese, Emiliano-Romagno, Venetan, and Friulian dialects) and I selected only dialects spoken in datapoints located below 500 metres of altitude. I also excluded all those Lombard dialects in which INFL markers are reflexes of postverbal SCL (§3). I thus obtained a set of dialects spoken in a relatively homogeneous area, which are not separated by significant orographic barriers that would make geodesic distances unreliable.

Then I used the R package *Vegan* (Oksanen *et al.* 2018) to calculate the correlation between geographic and linguistic distances (Mantel test). A previous study on SCL showed no correlation between geographic and linguistic distances ($r = 0.05931$; significance: 0.014, see Author 2019). The present study departs from Author 2019 in three respects:

1. the sample has been narrowed to 63 dialects spoken in the basin of the river Po, to avoid issues in the calculation of geographic distances;
2. only 1-3P and 6P SCL/INFL have been coded, for the reasons discussed in §3;
3. the present experiment intends to compare the correlation between geographic dispersion and typological dispersion of both SCL and INFL, whereas previous studies have focused only on the correlation with respect to SCL. Taking the value of INFL as a baseline, I will be able to determine whether the linguistic/geographic distance correlation index of SCL is higher, lower, or equal to the one of INFL.

The results of the Mantel test are summarised in table (18). Each index measures the degree of correlation between geographic and linguistic distances in the SCL and INFL domains, respectively.

(18) Comparison between correlation indexes: SCL vs INFL.

SCL/geographic distances	INFL/geographic distances
Mantel statistic r: 0.04213	Mantel statistic r: 0.1889
Significance: 0.15	Significance: 0.001

The results in (18) show that there is no possible correlation between linguistic and geographical distances with respect to inventories of SCL. Hence, SCL do not exhibit many variants (see Fig. 4), and those variants are randomly distributed across geographic areas. By contrast, a weak, but significant positive correlation (18%) is found between INFL and geographic distance: the farther the datapoints from each other, the higher the number of differences. Given the difference between the correlation indexes for SCL and INFL, it seems to me that it is fair to conclude that the robust typological tendencies observed in SCL are not a “historical accident”.

In conclusion, the fact that the types of SCL inventories are relatively few, see Tab. 4, and geographically scattered brings indirect evidence for language-internal explanations of the kind suggested or discussed in previous works such as Heap 2002, Cabredo Hofherr 2004; Benincà & Poletto 2005, Floricic and Molinu 2008, 46-47; Oliviéri 2011, Calabrese 2011.

7. Conclusion

This article has focused on the relationship between the presence and shape of subject clitics and verbal endings in a sample of Italo-Romance dialects.

It has been a widely held view since the works of 19th century Romance scholars that subject clitics emerged to *compensate* the loss of verbal endings, which had been eroded by independent morpho-phonological changes. In the 20th century, the question has been revived in the light of generative theorizing concerning the distinction between null and non-null subject languages. Rich inflection was in fact considered as one of the many factors triggering or allowing null subjects.

The debate about the status of ‘rich’ inflection was in turn related to the debate concerning the nature of subject clitics: (standard) French has been often regarded as a non-null subject language in which subject clitics are fully fledged pronouns because INFL is impoverished, whereas Italo-Romance subject clitics have been often analysed as redundant agreement markers licensing null subjects (see Brandi and Cordin’s 1981 and Rizzi’s 1986).

This article aimed to translate an abstract concept such as ‘rich inflection’ into quantitative indexes and to explore a methodology to test previous hypotheses on the basis of a relatively broad set of data. The methodology is entirely based on the comparison of distinctive forms (including zeroes). This kind of methodology can be applied to dialects, such as the Italo-Romance ones, in which the presence of subject clitics is mandatory. Data concerning other linguistic areas such as Occitan, where subject clitics are optional (Heap 2002, Oliviéri 2011, Sibille 2012) do not lend themselves to the same analysis.

The first result of my inquiry is that Italo-Romance subject clitic systems are never underspecified. In certain dialects, subject clitics and distinctive endings are in complementary distribution, but in the majority of languages subject clitics outnumber distinctive inflectional endings. The absence of underspecified systems militates in favour of diachronic and synchronic explanations claiming that the presence of subject clitics is related to the richness of inflection. At the same time, however, the predominance of overspecified systems indicates that further factors are at play in shaping inventories of subject clitics.

The second result is that, from a geolinguistic perspective, inventories of subject clitics are distributed randomly across dialects, whereas the distribution of inflectional systems tends to be more regular from a geolinguistic standpoint: close dialects exhibit more similarities in the inflectional system than in clitic systems. This brings further support to the claim that inventories of subject clitics do not reflect the array of inflectional endings. Some further factors are probably involved in the emergence of subject clitics because, although subject clitics exhibit relatively few patterns of variation, such patterns are geographically scattered.

This article remains agnostic regarding the nature of the factors constraining inventories of subject clitics. Such factors might be either feature geometries/filters, as proposed in the previous literature (Heap 2002, Benincà & Poletto 2005, Calabrese 2011), or *third factors* in the sense of Chomsky 2005, i.e. factors that are related to computational efficiency and/or processing costs. To find out, we probably need to abandon the ‘flat’ representation of paradigms adopted in this study (where inventories were mapped into sequence of 1 and 0) and adopt a compositional representation of person distinction based on simpler binary feature, e.g. participant, speaker, plural, etc.

References

- Adams, Marianne. 1988. Les effets V2 en ancien et en moyen français. *Revue québécoise de linguistique théorique et appliquée* 7(3): 13-39.
- Benincà, Paola. 1983. Il clítico a nel dialetto padovano. In Benincà, Paola *et al.* (eds.), *Scritti in onore di G.B. Pellegrini*. Pisa: Pacini, 25-35.
- Benincà, Paola & Poletto, Cecilia. 2005. The third dimension of person features. In Leonie Cornips and Karen P. Corrigan (eds), *Syntax and Variation*. Amsterdam: Benjamins, 265–299.
- Brandi, Luciana & Cordin, Patrizia. 1989. Two Italian Dialects and the Null Subject Parameter. In O. Jaeggli, and K. Safir (eds), *The null subject parameter*. Dordrecht: Kluwer, 111-142.
- 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.
- Casalicchio, Jan & Cognola, Federica (eds.). 2018. *Null Subjects in Generative Grammar. A Synchronic and Diachronic Perspective*. Oxford: Oxford University Press.
- Chomsky, Noam. 2005. Three Factors in Language Design. *Linguistic Inquiry* 36(1): 1-22
- Culbertson, Jennifer. 2010. Convergent evidence for categorical change in French: From subject clitic to agreement marker. *Language* 86: 85-132.
- Evans, Nicholas & Levinson, Stephen. 2009. The myth of language universals: language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32: 429–492.
- Gaglia, Sasha. 2012. The Friulian subject clitics. Realisation and paradigmatic structure. In Sascha Gaglia and Marc-Olivier Hinzelin (eds), *Inflection and Word Formation in Romance Languages*. Amsterdam: Benjamins, 83-118.
- Heap, David. 2002. Split subject pronoun paradigms : Feature geometry and underspecification. In Satterfield, T. *et al.* (eds), *Current Issues in Romance Languages: Selected Papers from the 29th Linguistic Symposium on Romance Languages (LSRL)*, Ann Arbor, 8-11 April 1999. Amsterdam: Benjamins, 129-44.
- Maiden, Martin. 2005. Morphological autonomy and diachrony. *Yearbook of Morphology 2004*: 137–175.
- Manzini, Maria Rita & Savoia, Leonardo. 2005. *I dialetti italiani e romanci. Morfosintassi generativa*. Alessandria: Edizioni dell’Orso.
- Meyer-Lübke, Wilhelm. 1895. *Grammaire des langues romanes* (II). Paris : Welter
- Nichols, Johanna. 1992. *Linguistic Diversity in Space and Time*. University of Chicago Press.

- Oksanen, Jari *et al.* 2018. Vegan: Community Ecology Package. R package version 2.5-3.
<https://CRAN.R-project.org/package=vegan>
- Oliviéri, Michèle. 2011. Typology or Reconstruction: the Benefits of Dialectology for Diachronic Analysis. In Berns, J. *et al.* (eds.), *Romance Languages and Linguistic Theory 2009. Selected papers from 'Going Romance' Nice 2009*. Amsterdam: Benjamins, 239-53.
- Palasis, Katerina. 2015. Subject clitics and preverbal negation in European French: Variation, acquisition, diatopy and diachrony. *Lingua* 161: 125-143.
- Perlmutter, David. 1971. *Deep and Surface Structure Constraints in Syntax*. New York: Holt, Rinehart and Winston.
- Poletto, Cecilia. 2013. Leopard spot variation: What dialects have to say about variation, change and acquisition. *Studia Linguistica* 67(1): 165-183.
- Poletto, Cecilia 2000. *The Higher Functional Field. Evidence from Northern Italian Dialects*. Oxford: Oxford University Press.
- Renzi, Lorenzo and Vanelli, Laura (1983). 'I pronomi soggetto in alcune varietà romanze'. In Benincà, Paola *et al.* (eds.), *Scritti in onore di G.B. Pellegrini*. Pisa: Pacini, 120-145.
- Rizzi, Luigi. 1986. On the status of subject clitics in Romance. In O. Jaeggli, and C. Silva Corvalan (eds), *Studies in Romance Linguistics*. Dordrecht: Foris, 391-419.
- Roberts, Ian. 2014. Subject clitics and macroparameters. In Paola Benincà, Adam Ledgeway, and Nigel Vincent (eds), *Diachrony and Dialects: Grammatical Change in the Dialects of Italy*. Oxford: Oxford University Press, 177-201.
- Sibille, Jean. 2012. Syncrétisme des formes verbales et des clitiques sujets dans plusieurs variétés romanes vernaculaires et en français standard. In Eva Lavric, Wolfgang Pöckl (ed.), *Internationalen Arbeitstagung zum romanisch-deutschen und innerromanischen Sprachvergleich*. Innsbruck: Peter Lang. 341-354.
- Taraldsen, Tarald. 1980. *On the NIC, vacuous application and the that-trace filter*. Indiana University Linguistics Club.
- Vanelli, Laura 1998. *I dialetti italiani settentrionali nel panorama romanzo. Studi di sintassi e morfologia*. Roma: Bulzoni.
- Vedovato, Diana. 2009. Weak pronouns in Italian: Instances of a broken cycle? In Elly van Gelderen (ed.), *Cyclical Change*. Amsterdam: Benjamins, 133-156.