



1996

Subject Pronoun Variation in Central Romance

David Heap

Follow this and additional works at: <https://repository.upenn.edu/pwpl>

Recommended Citation

Heap, David (1996) "Subject Pronoun Variation in Central Romance," *University of Pennsylvania Working Papers in Linguistics*: Vol. 3 : Iss. 1 , Article 4.

Available at: <https://repository.upenn.edu/pwpl/vol3/iss1/4>

This paper is posted at ScholarlyCommons. <https://repository.upenn.edu/pwpl/vol3/iss1/4>
For more information, please contact repository@pobox.upenn.edu.

Subject Pronoun Variation in Central Romance

Subject Pronoun Variation in Central Romance

David Heap
University of Toronto and York University

1 Is Variation in Subject Pronouns Parametric?

A considerable amount of research in generative syntax over the last decade and a half has been based on the hypothesis that the settings of binary parameters, such as [\pm null subject], can exhaustively describe the full range of possible grammars of natural languages¹. Indeed, the prime example used by Chomky (1981) is the so-called "null subject parameter", which was supposed to correlate a number of seemingly unrelated properties, such as stylistic inversion and long-range *wh*-extraction, to the capacity of a grammar to have non-overt subjects, i.e. finite verbs with no phonetically realized argument in the structural subject position [NP, IP]. In this framework, the difference illustrated in (1) between languages like English and Standard French, with obligatory subjects, and those like Spanish and Standard Italian, where a verb can stand alone without an overt subject, represents an important bifurcation in the class of natural languages:

(1)

[-null subject] languages	[+null subject] languages
English: <i>I speak</i> French: <i>Je parle</i> etc...	Spanish: <i>Hablo</i> Italian: <i>Parlo</i> etc...

If this major division between possible natural languages were real, we would expect it to be reflected in geolinguistic terms by an abrupt change from one system to another: a grammatical system should have either one setting for this parameter or the other, and never both at once, and never partly one and partly another. For those interested only in standard varieties, this is just about true: if we go from Standard Italian to Standard French, we can in a sense draw a sharp isogloss between the [+null subject] and the [-null subject] zones. The only problem is that this isogloss would have to be drawn with a very broad brush in order to cover the transition zone which runs roughly from Florence and Venice to Grenoble and Nice. This transition zone is the object of a larger research project (Heap in progress) of which the findings presented here are just a small sample.

If we first consider some of the different linguistic factors which may condition subject pronoun use, and then examine the relative role of these factors using data drawn from a geolinguistic corpus, we arrive at some results which are difficult to fit into a binary parametric grammar.

¹ I wish to thank Terry Nadasdi, without whom this paper would never have happened (but who is much too nice a guy to blame its many shortcomings on). This research has been supported by SSHRCC doctoral fellowship 752-91-2167, by a QE II Ontario Scholarship, and by SSHRCC research grant (Roberge 410-91-1307).

2 Factors Affecting Subject Pronoun Realization

There has of course been a certain amount of attention devoted by generative syntacticians to these troublesome varieties from Northern Italy and South-Eastern France which would seem to mess up the neat binary divisions of parametric theory (Rizzi 1985, Vanelli, Rizzi and Benincà 1985). For example, Brandi & Cordin (1989) characterize varieties like Trentino and Fiorentino as [+null subject] varieties where the subject clitic is in fact part of the verbal morphology. This analysis depends crucially on the fact that these varieties (unlike Standard French) **require** a subject clitic even when there is an overt lexical subject, as in (2):

- | | | |
|-----|--------------------------------------|---|
| (2) | a. Mario e parla. (Fiorentino) | 'Mario <i>clitic</i> speaks.' |
| | b. *Mario parla. | 'Mario speaks.' |
| | c. El Mario el parla. (Trentino) | ' <i>article</i> Mario <i>clitic</i> speaks.' |
| | d. *El Mario parla. | ' <i>article</i> Mario speaks.' |
| | e. *Jean il parle. (Standard French) | 'Jean <i>clitic</i> speaks.' |
| | f. Jean parle. | 'Jean speaks.' |

Of course, we now have work by linguists such as Roberge & Vinet (1989), Auger (1994) and Nadasdi (1995) on nonstandard varieties of French (especially Non Acadian Canadian French) where sentences like those in both (2e) and (2f) are grammatical and widely attested. But concentrating on our Northern Italian transition zone, it seems clear that the compatibility of subject pronouns with overt lexical subjects is one of the factors that will help distinguish amongst the different grammars in this region. Benincà & Poletto (1991) contrast the cases of Friuli, where subject pronouns are obligatory even after overt subjects, and Venetian, where subject pronouns are optional in the same context, as in (3):

- | | | |
|-----|-----------------------------|-----------------------------|
| (3) | a. Meni al ven. (Friuli) | 'Dom <i>clitic</i> comes.' |
| | b. *Meni ven. | 'Dom comes.' |
| | c. Nane al vien. (Venetian) | 'John <i>clitic</i> comes.' |
| | d. Nane vien. | 'John comes.' |

There is also literature (Benincà & Poletto 1991, Nadasdi 1995) to justify a distinction between lexical subject which are definite (like those shown here) and those which are indefinite. So for the purposes of this study, the factor group "Subject" contained the factors "absence (of lexical subject)", "definite lexical subject" and "indefinite lexical subject".

Another factor which correlates with a considerable amount of variation in subject pronouns is grammatical person (Heap forthcoming): in the Trentino case mentioned above, the clitic subject pronoun is only obligatory in the 2nd, 3rd, and 6th persons; in the 1st, 4th and 5th person, a verb can appear without a subject pronoun (Brandi & Cordin 1989:113). Person variation in subject pronouns across this region is surveyed in Renzi &

Vanelli (1983), who propose a hierarchy of possible subject pronoun paradigms, ranging from just one clitic subject (always the 2nd person *tu*) to five clitics (all persons except the 1st). In order to test the hypothesis that this factor conditions the appearance of clitic pronouns, the tokens for this study were coded from 1 to 6 according to grammatical person (using the convention whereby persons 4 to 6 correspond to 1st, 2nd and 3rd persons plural).

There is also evidence, for example from Poletto (1993), showing that the auxiliary verbs used in perfect tenses sometimes pattern differently with respect to subject pronouns. Furthermore, the interaction of verb tense/mood morphology with subject marking through pronouns could of course be functionally motivated: even in Standard Italian, some verb forms present more syncretism than others, and in these cases normative grammars suggest using the (normally emphatic) strong subject pronoun "to prevent ambiguity". This functional hypothesis was examined quantitatively in another "null-subject language", Spanish, by Hochberg (1986), who found that subject pronoun usage increases in tenses with more syncretism in the verbal morphology. For coding purposes, Hochberg was able to reduce the various tenses and moods of Spanish to just three classes according to the level of syncretism. For the purposes of this study, however, I began with each of the verb forms in question coded separately: present indicative, present perfect, future, present subjunctive, and conditional.

The other factor groups taken into account here because of their demonstrated relevance in other studies were: a) the number of object clitics intervening between a subject and its verb (0 to 3), and b) verb type (transitive, unaccusative, unergative, and auxiliaries like *to be* and *to have*).

3 Corpus Studied

Most of the studies of subject pronouns in Northern Italian varieties rely on data from only one or two geolcts, e.g. the Brandi & Cordin (1989) study of Trentino and Fiorentino. Renzi & Vanelli's (1983) survey is the broadest-based, with elicited data from some 30 varieties. The spread of standard varieties and dialect restriction of course make it increasingly difficult to obtain new data for some of the varieties in question, but linguistic atlases can provide us with comparable data from a large number of points, which were collected at a time when local varieties were still commonly spoken. While primarily concerned with lexical and phonetico-phonological variation, Gilliéron & Edmont's (1902-1910) *Atlas Linguistique de la France* (ALF) and Jaberg & Jud's (1928-1940) *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS) contain between them hundreds of maps showing forms with inflected verbs i.e. potential subject pronoun environments.

While by no means all the points (there are over 1000 between the two atlases) fall within the transition zone which interests us here, there is nonetheless a dense enough network of points in Northern Italy and S-E France to give us a fine-grained portrait of this grammatical transition zone. For this paper a subsample was selected, consisting of the last 100 points in the ALF and the first 100 points in the AIS. This sample covers contiguous (in fact, slightly overlapping) areas in S-E France and N-W Italy, as you can see on the maps at the bottom of page 1:

ALF points #861-#992 (in Gard, Bouches-du-Rhône, Var, Alpes-Maritimes, Basses Alpes, Hautes Alpes, Vaucluse, Drôme, Isère, Savoie, Haute Savoie, Rhône, Ain, Saône-et-Loire, Jura, 8 points in Italy & 12 in Switzerland).

AIS points 1-222 (in Milan, Genoa, Turin & 36 in Switerland)

For the purposes of this paper, a subsample was drawn from the hundreds of maps bearing potentially relevant data, focussing on just seven pairs of maps with identical (or near-identical) forms in both of the atlases, as shown in (4):

(4)

ALF Maps	AIS Maps
99 Vous auriez (dù voir)	1250 Avreste dovuto (vedere)
509 Ils sont morts	76 Sono morti
514 (Quand mon fils) sera grand.	10 (Quando il mio figlio) sarà grande
803 Je suis malade...	1013 Sono digiuno
806 Si nous ne mangeons pas...	1278 Se non mangiamo ...
1103 Puisque tu as faim...	1015 Poichè hai fame...
1679 J' ai posé (ça là)	887 Ho messo...

Thus a theoretical total of almost 1400 tokens were coded for 5 factor groups: Tense/mood, Verb type, Subject type, Number of object clitics, Person, and of course for the variable itself, absence or presence of a subject pronoun. The actual total number of 1173 falls short of 1400 because of a number of points for which no data was available on a given map; the value "not applicable" or "/" was assigned where it was not possible to determine a given factor with certainty e.g. the verb tense of a given form.

4 Results

Three of the coded factor groups – Number of object clitics, Verb type, and Subject type – were eliminated by the GoldVarb 2 analysis as not significant. This may well reflect the poverty of the data in this subsample more than anything else, and we should expect to perhaps see significant effects from these factors once larger numbers of tokens are coded from maps with more object clitics, as well as a greater range of verb types and subject types.

The GoldVarb 2 analysis retained two factor groups as significant: Tense/Mood and Grammatical Person, as shown in (5) and (6).

In the first group, Tense/Mood, the factor "present" in fact represents the collapsing of present, present perfect and conditional tenses, as opposed to the future and the subjunctive. Thus this result confirms the tendency noted by Hochberg: the future verbal morphology most clearly marks person distinctions and this same tense (with a factor weight of just 0.192) strongly disfavors the appearance of a subject pronoun. Conversely, the present subjunctive has the highest degree of syncretism amongst the tenses considered here, and, not surprisingly, this is correlated with an effect which favors subject clitics (as shown by the factor weight of 0.690). The present indicative is grouped here with the present perfect (which contains an auxiliary verb inflected for present tense) and the conditional, all three of which fall somewhere between these two extremes: these tenses typically make fewer desinential distinctions between grammatical persons than the simple future but more than the subjunctive, and they predictably favor subject pronouns but not as strongly as the most syncretic case does, weighing in at just 0.547.

(5) GoldVarb 2 Factor Weights, Ns and %:

Group		Apps	Non-apps	Total	%	Factor Weights

Tense/Mood:						
present	N	631	386	1017	87	0.547
	%	62	38			
future	N	30	112	142	12	0.192
	%	21	79			
subjunctive	N	10	4	14	1	0.690
	%	71	29			
Total	N	67	502	1173		
	%	57	43			

Log likelihood = -746.111 Significance = 0.000

(6) GoldVarb 2 Factor Weights, Ns and %:

Group		Apps	Non-apps	Total	%	Factor Weights

Grammatical Person						
2nd	N	161	51	212	18	0.660
	%	76	24			
1st	N	193	143	336	28	0.453
	%	57	43			
6th	N	118	57	175	15	0.559
	%	67	33			
3rd	N	58	129	187	16	0.456
	%	31	69			
4th	N	67	72	139	12	0.363
	%	48	52			
5th	N	82	53	135	11	0.487
	%	61	39			
Total	N	679	505	1184		
	%	57	43			

Log likelihood = -746.111 Significance = 0.000

This result, like those of Hochberg (1986), supports the functional hypothesis whereby the marking of a particular grammatical category such as person by one means

reduces the likelihood of the same information being marked by another means. By the same token, this result would seem to contradict the well-documented trend towards the preservation of parallel structures (Labov 1994:Ch. 19). This difference may have to do with an inherent difference in the variables: the category of number (which can be treated univalently as simply "plural" or its absence) has been shown to appear redundantly in many Romance varieties, while grammatical person agreement in verbs (a six-way distinction) does not seem to lend itself to repetitive marking, at least in this case.

It should be noted however that the actual numbers represented by these cases, as shown in (6), suggest that the factor weighting for the subjunctive (representing just 1% of the data) should be treated with considerable caution until more data can be analysed along these lines. The actual numbers for the Grammatical Person group are more robust and will likely be borne out by more data.

Turning now to the results for second group, Grammatical Person, we see that the 2nd and 6th persons favor the presence of subject pronouns (at 0.660 and 0.559 respectively), while the 1st and 4th persons disfavor it (at 0.453 and 0.363 respectively). The 3rd and 5th persons are in the middle with slightly unfavorable factor weightings of 0.456 and 0.487. The resulting cluster of grammatical persons from most to least favorable to subject pronouns, as in (7):

(7) Subject Pronouns ranked by Grammatical Person factor weightings:

Most favorable		Least favorable
2, 6	3, 5	1, 4

may appear to contain no significant discernible pattern, but in fact it partially correlates with the implicational hierarchy of subject pronoun paradigms proposed by Renzi & Vanelli (1983:143):

(8) Person Hierarchy for Subject Pronouns

Most common	Least common
2, 3, 6	5, 4, 1

Specifically, the two rankings correlate most strongly at the extremes: in both studies, the 2nd person is the most likely to have a subject pronoun if there is to be one at all, and the 4th and 1st persons are the least likely; the relative differences in factor weights for the 3rd, 5th and 6th persons is likely not significant at these numbers, and will have to be revisited as more data become available.

So while it may be not clear from these results exactly what the relative effect of the various persons is on subject pronouns, it is clear that grammatical person, along with the level of syncretism in the different verbal forms, both play a role in conditioning subject pronouns in these Central Romance varieties. And their role is relative and probabilistic, and as such cannot satisfactorily be reduced to any formalism which relies solely on binary parameters. In order to capture the combined effects of Tense/Mood and Grammatical person in a parametric framework, we would have to resort to a host of individual [\pm null subject] parameters which are set in an *ad hoc* fashion in each morphosyntactic environment – a solution which would obviate the original explanatory power of the parametric model. Alternatively, we could abandon the binary conception of parameters in favor of scalar or probabilistic parameters, but this too would clearly contradict the intent of Chomsky (1981). It remains to be seen whether the larger picture that emerges of this morphosyntactic transition zone (Heap, in progress) will prove more

or less compatible with the parametric model of grammar, but the early indications are that any strictly binary formalism is likely to be empirically inadequate faced with the facts of such a continuum.

5 Concluding Remarks

By way of conclusion, I would like to be the first to point out the most glaring flaw with this study, which is that the sample includes tokens from a range of varieties, some of which undoubtedly have quite distinct grammars. The data are skewed by the inclusion of points where subject pronouns are categorically absent, and others where this variable is categorically present. With this in mind, the factor weightings presented here should not be taken as corresponding to the likelihood of subject pronouns appearing in any one variety, but rather as a indication of overall trends. The preliminary step in analysing such a geolectal continuum will be to identify all those varieties where subject pronoun usage is categorical in a given morphosyntactic context. The remaining varieties, where subject pronouns appear variably, will then undergo a more comprehensive variable rule analysis, one which is capable of grouping large numbers of differing speakers according to whether they have the same grammar with the same constraints and factor weightings, or separate grammars with distinct constraints and factor weightings. Such an analysis, which necessarily implies a much larger and more varied database, is the next step in the study of the geolectal continuum which constitutes this morphosyntactic transition zone.

References

- Auger, Julie (1994). *Pronominal Clitics in Québec Colloquial French: A Morphological Analysis*. Unpublished Ph.D. Thesis, University of Pennsylvania.
- Benincà, Paola and Cecilia Poletto (1991). "Il modelo generativo e la dialettologia: un'indagine sintattica." *Rivista Italiana di Dialettologia* 15:77-97.
- Luciana Brandi & Patrizia Cordin (1989). "Two Italian Dialects and the Null Subject Parameter." in O. Jaeggli & K.J. Safir, eds., *The Null Subject Parameter*. (Dordrecht: Kluwer), 111-142.
- Chomsky, Noam (1981). *Lectures on Government and Binding*. (Dordrecht: Foris).
- Heap, David (forthcoming). "Subject Clitics in the Northern Italian Geolinguistic Continuum." to appear in *Quaderni d'italianistica*.
- Heap, David (in progress). *La variation grammaticale en géolinguistique: le cas des sujets clitiques*. Ph.D. Thesis, University of Toronto.
- Hochberg, Judith G. (1986). "Functional Compensation for /s/ Deletion in Puerto Rican Spanish." *Language* 62, 3:609-621.
- Gilliéron, Jules and Edmont Edmont (1902-1910). *Atlas linguistique de la France* (Paris:Champion).
- Jaberg, Karl and Jakob Jud (1928-1940). *Sprach- und Sachatlas Italiens und der Südschweiz*. (Ringier: Zofinger).
- Labov, William (1994). *Principles of Linguistic Change. v1: Internal Factors*. (Oxford & Cambridge: Blackwell).
- Nadasdi, Terry J. (1995). *La variation morphosyntaxique et langue minoritaire: le cas du français ontarien*. Ph.D. Thesis, University of Toronto.

- Renzi, Lorenzo and Laura Vanelli (1982). "I pronomi soggetto in alcune varietà romanze." *Scritti linguistici in onore di G.B. Pellegrini*. (Padova: Pacini), 121-45.
- Rizzi, Luigi (1986). "On the Status of Subject Clitics in Romance." in O.A. Jaeggli and C. Silva-Corvalán, eds., *Studies in Romance Linguistics*. (Dordrecht: Foris), 391-419.
- Roberge, Yves and Marie-Thérèse Vinet (1989). *La variation dialectale en grammaire universelle*. (Montréal: Presses de l'Université de Montréal).
- Vanelli, Laura, Lorenzo Renzi and Paloa Beninca (1983). "Typologie des pronoms sujets dans les langues romanes." in *Actes du XVIIeme Congres International de Linguistique et Philologie Romanes (Aix-en-Provence)*, 3:163-76.