



HAL
open science

A quantitative approach to microvariation: negative marking in central Romance

Diego Pescarini

► **To cite this version:**

Diego Pescarini. A quantitative approach to microvariation: negative marking in central Romance. Languages, In press. hal-03480594

HAL Id: hal-03480594

<https://hal.science/hal-03480594>

Submitted on 14 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.

A quantitative approach to microvariation: negative marking in central Romance

- First draft -

Diego Pescarini (CNRS, Université Côte d'Azur, BCL)

1. Introduction

Languages exhibit a natural tendency to evolve from preverbal to postverbal negative marking through a stage of discontinuous negation in which pre- and postverbal markers co-occur (Jespersen 1917). To avoid a misleading terminology, the descriptive terms *preverbal* and *postverbal* negation are here replaced by the conventional labels N1 and N2, which are defined as follows:

- 1) N1 usually derives from Latin *non* (Fr. *ne*, It. *non*) and co-occurs with Negative Polarity Items (NPI), giving rise to negative concord configurations; it is usually placed preverbally and, in finite clauses, it is proclitic to the finite verb (only other clitics can occur between N1 and the verb);
- 2) N2 derives from various kinds of elements such as nouns denoting a minimal quantity (dubbed *minimizers*, e.g. *pas* 'step', *point*, *mie/mica/brisa* 'crumb', etc.) or, to a lesser extent, negative quantifiers and polarity particles; it sometimes co-occurs with NPIs (cf. Fr. *je ne mange (*pas) rien* 'I eat nothing') and it normally occurs postverbally in finite clauses.

The position of N2 is subject to cross-linguistic variation and cross-contextual variation. N2 may occur preverbally in infinitives as in (1), imperatives as in (2), or can be focus-fronted under specific pragmatic conditions, in particular in those languages like standard Italian in (3), in which N2 is in an early stage of grammaticalization (Pescarini and Penello 2012).

- (1) a. Je ne bouge pas mon bras. (Fr.)
I= N1 move.1SG N2 my arm
'I cannot move my arm.'
- b. pour ne pas bouger mon bras.
to N1 N2 move.INF my arm
'(in order to) not to move my arm.'
- (2) a. eu ki'pe:f bek (AIS map 1658; Ems, Rhaeto-Romance)
I understand.1SG N2
'I do not understand.'
- b. 'be:ke sə'mwəntə (AIS map 1647; Ems, Rhaeto-Romance)
N2 move.2PL
'don't move'
- (3) a. Non lo mangio mica. (Italian)
N1 it= eat.1SG N2
- b. Mica lo mangio!
N2 it= eat.1SG
'I do not eat it/I am not eating it.'

In postverbal position, N2 can be preceded the past participle and various classes of aspectual adverbs that in Romance usually follow the inflected verb (Cinque 1999). The number and type of adverbs that precede negation is subject to cross-linguistic variation (Zanuttini 1997; Manzini & Savoia 2005, 2012). Most N2s precede complements, except the negator corresponding to the polarity particle 'no', which in northern Italian dialects such as Milanese always occurs in sentence-final position (e.g. Mil. ['dɔrmi nɔ] 'I do not sleep'). Variation in the syntax (and morphology) of N2s will be addressed in a separate work.

The present article focuses on the presence/absence of N1/N2 regardless of the position of the latter in the structure of the clause. The goal of this paper is twofold:

- a) it intends to assess the feasibility of multivariate analysis of syntactic variables on a sample of data collected almost a century ago for dialectological purposes;
- b) to verify whether grammatical factors such as clause typing and negative concord play any role in the cross-linguistic and cross-contextual distribution of N1/N2.

A statistical analysis of microvariation can shed light on the internal mechanisms that rule the distribution of N1 and N2 across space, i.e. across linguistic systems, and time. Jespersen 1917 first discovered the tendency to replace N1 with N2—a tendency that has been lately dubbed ‘Jespersen’s cycle’—and argued that the mechanism is triggered by a *weakening* of N1, associated with a progressive grammaticalisation of NPIs (minimizers quantifiers, pro-sentences). Originally, the distribution of incipient N2s is constrained by pragmatic conditions: the presence of a double negator tends to trigger the *activation* of an inferential link between the negative clause and a discourse-old proposition (Schwenter 2005; Larrivé 2008). Given its pragmatic value, N2s *in statu nascendi* are often ruled out in factive sentences, which are already presuppositionally linked (Cinque 1991 [1976]). When this pragmatically marked value is lost, such elements becomes full-fledged N2s—i.e. they become part of the *default* negation (Larrivé 2008)—and eventually remain as the sole negation if N1 is lost (see van Gelderen 2008, Breitbarth 2013 for formal analyses).

This article does not deal with change across time, but it aims to investigate how the syntax of N1 and N2 varies across space. To do so, this article builds on a statistical analysis of data from the *Atlas linguistique de la France* (ALF; Gilliéron and Edmont 1902-1910) and the *Sprach- und Sachatlas Italiens und der Südschweiz* (AIS; Jaberg and Jud 1928-1940). These atlases, which have been (partially) digitized in recent times, contain parallel translations of sentences, phrases, and isolated words elicited in, respectively, 639 and 407 datapoints across France, Belgium, Italy, southern Switzerland, and nearby regions. All data were gathered by interviewing a single NORM speaker (Non-mobile, Old, Rural Men) per dialect following a written questionnaire. The ALF and AIS questionnaires include 46 negative sentences (24 in Italian and 22 French, respectively) that elicited over twenty thousand replies (more on this in §2).

The geolinguistic distribution of N1 and N2 is quite regular, as shown in Fig. 1, which overviews the distribution of dialects that exhibit N1 (purple dots), N1 and N2 (green dots), and N2 (yellow dots) in the 1.046 ALF/AIS datapoints.

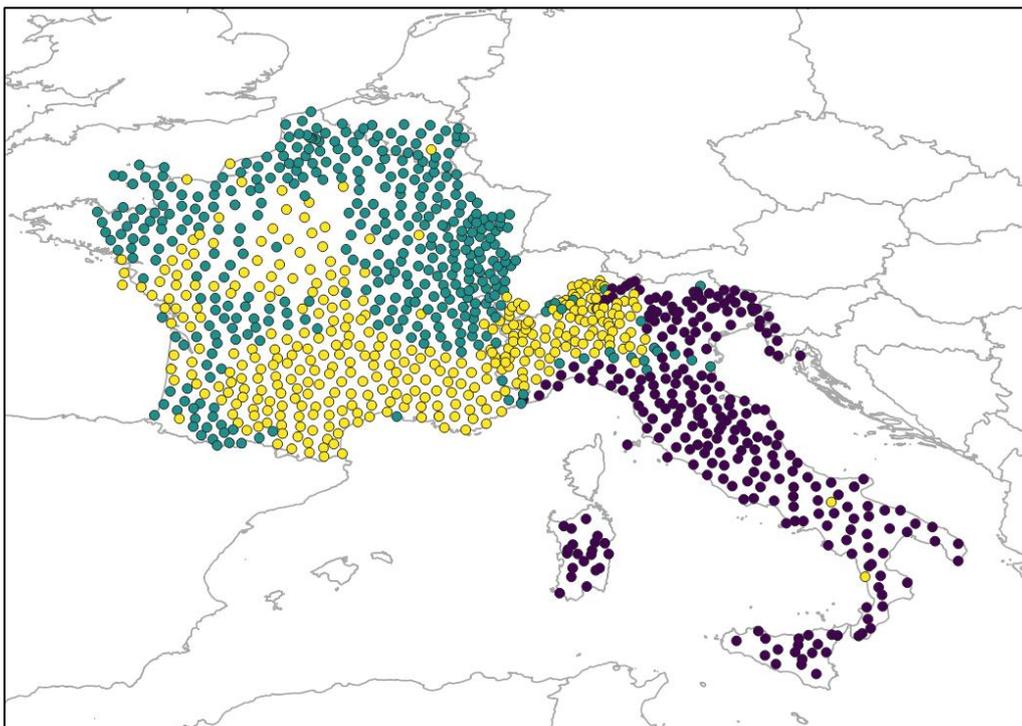


Figure 1. Distribution of N1 and N2 in declarative main clauses: dialects exhibiting N1 (purple dots), discontinuous negation (green dots) and N2 (yellow dots) in the 1.046 ALF/AIS datapoints.

The map in Fig. 1 plots the data from a couple of AIS/ALF maps, both reporting translations of negative declarative clauses. Besides geolinguistic variation, however, N1/N2 vary along several other dimensions, both external (sociolinguistic) and internal.

Several works have dwelt on sociolinguistic variation in negation marking, especially in French. Ashby’s 1981 seminal paper, for instance, examined a corpus of over 100 interviews in colloquial French recorded in the Tour region, showing that the occurrence of N1 varies according to a complex set of linguistic, stylistic, and social factors, including discourse setting (formal/informal), age, social class, etc. The AIS/ALF dataset, however, is not amenable to sociolinguistic analysis of the kind that Ashby did because both AIS and ALF interviews were conducted with a single speaker per dialect and informants were all NORM speakers (Non-mobile, Old, Rural Men) in order to minimize diastatic and diaphasic variation. However, since various kinds of negative clauses are contained in both AIS and ALF questionnaire, we can try to compare the occurrence of N1 and N2 in different syntactic environments, including some of those that—according to previous studies such as Ashby 1981, Pescarini & Donzelli 2017—may affect the behavior of N1 and N2.

Ashby’s study showed that the incidence of N1 is affected by internal—i.e. grammatical—factors. Ashby shows that N1 is found less frequently in sentences containing a NPI, in embedded clauses (especially in subjunctive clauses), and when a subject clitic is present. Similar results are found when, instead of comparing speakers of a single variety, we compare data from nearby dialects. Pescarini & Donzelli 2017, for instance, found that variation in a sample of Ticinese dialects (Lombard-type dialects spoken in southern Switzerland) appears to be related to syntactic factors such as mood or the adjunct/argument status of NPIs, although their results need to be confirmed on a bigger dataset like the one examined in this study.

If variation across sentences—i.e. across AIS/ALF maps—is examined, cross-linguistic variation appears to be more nuanced, as shown in Fig. 2. While Fig. 1 shows the distribution of N1 and N2 in indicative clauses, Fig. 2 was generated by superimposing the data from various contexts, e.g. imperatives, questions, *if* clauses, etc.

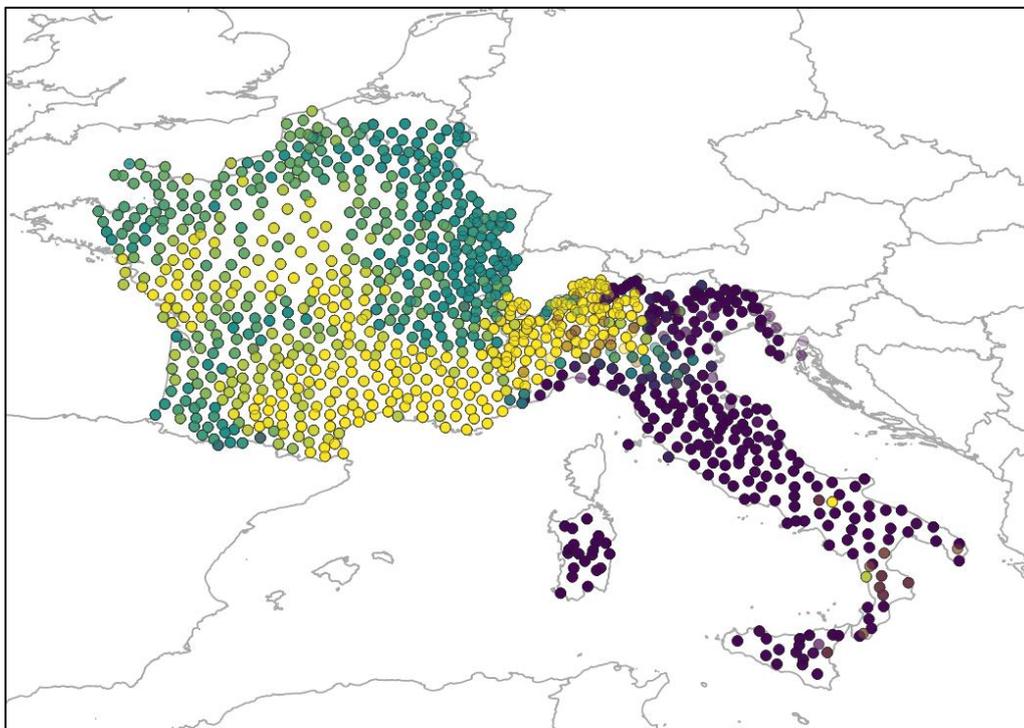


Figure 2. Distribution of N1/N2 in across syntactic contexts and across 1.046 ALF/AIS datapoints.

The light green and dark yellow points in Fig. 2 signal that, in certain datapoints, the distribution of N1 and N2 varies across syntactic environments. To find out whether cross-linguistic and cross-contextual variation correlates with grammatical factors, we need to perform a multivariate analysis of the AIS/ALF dataset with the intent of verifying whether the distribution of two dependent variables such as the incidence of N1 and N2 is linked to other syntactic factors or not. The main concern with this methodology is that syntactic factors are intertwined as sentences are often characterised by multiple grammatical factors at the same time. This holds particularly true for structured corpora such as linguistic atlases, which were not conceived for the purpose of syntactic analysis. Hence, since the inventory of input-sentences is limited while the spectrum of syntactic parameters is relatively wide, corpora such as the AIS/ALF tend to be heavily *collinear*. Due to collinearity, the role of each predictor is difficult to ascertain, even if the overall amount of tokens (i.e. dialect clauses) is relatively high.

Collinearity and other methodological issues will be discussed in the following sections, which are organized as follows: Section 2 overviews sources, data, metadata, methods, etc.; Sections 3 and 4 present some early results of a multivariate analysis on the AIS and ALF data, respectively; in Section 5 the results regarding the two areas are compared and then combined to obtain better results. Section 6 concludes.

2. Materials and Methods

Table 1 reports the number of negative sentences contained in each atlas:

Table 1. Overview of the dataset

Source	Negative sentences	Datapoints	Tokens
ALF	22	639	13185
AIS	24	407	9315
Total	46	1046	22500

In order to minimize random errors, the distribution of N1 was conducted on a subset of 19.432 sentences (see Appendix xxx). It is worth recalling that the primary data analysed in this work have been collected and transcribed almost a century ago. The mapping from primary data to metadata has been carried out manually in order to ensure accuracy, but some data remained unclear and were therefore excluded. For instance, in examining the distribution of N1, I excluded some sentences containing other preverbal nasal formatives such as first person plural subject clitics, impersonal clitics deriving from Latin HOMO ‘man’ and partitive/genitive clitics, e.g. ALF 97: *(Des pommes) nous n’en aurons (guère)*. All these elements have a nasal formative that can be easily mistaken for a preverbal negation marker and, if a single nasal segment occurs in the transcription, annotators were not always able to conclude whether that segment was a negation formative or not.

Table 2 shows how negation systems (N1, discontinuous: N1+N2, N2) are distributed in the two datasets (sentences containing NPIs are excluded because N2 is often in complementary distribution with NPIs). For the sake of consistency, I will focus on a subset of the AIS data, those spoken in ‘northwestern’ regions (Piedmont, Lombardy, Valle d’Aosta, Liguria, Emilia-Romagna, Trentino), thus excluding most dialects that exhibit only N1. Even if we focus on northwestern Italo-Romance dialects, yet the co-occurrence of N1 and N2 in the two datasets remains uneven because patterns of discontinuous negation are quite frequent in the ALF, as witnessed by the greenish dots in Fig. 2, but relatively rare in the AIS.

Table 2. Counts of tokens per negation system (excluding clauses containing NPIs).

Source	N1	N1+N2	N2
ALF	64	4404	3416
AIS: ‘northwestern’	559	190	1100
AIS: others	2050	5	18
Total	2673	4599	4534

The data reported in the AIS and the ALF have been collected following a well-established fieldwork methodology in dialectology: questionnaire-based interviews with NORM (Non-mobile, Old, Rural Men) informants, carried out in the late 19th and early 20th century (respectively, for the ALF and the AIS). Interviews were conducted by experienced linguists, who were specifically selected and trained for the purposes of the project. All data were transcribed on the spot by using ad hoc phonetic alphabets.

The AIS/ALF questionnaires consist of thousands of items (words, phrases or sentences). Each interview took several hours, usually divided in multiple sessions across three days. Interviewers had therefore the opportunity to interact with their informant and with other members of the same community for a relatively long time span. By cross-checking the elicited material with spontaneous speech (and with data from the nearby dialects surveyed during the same campaign), AIS/ALF interviewers were able to provide a faithful picture of each dialects.

In principle, mixed methodologies (Poletto and Cornips 2005) would probably elicit better results, in particular with respect to a phenomenon such as negation marking that, as mentioned in Section 1, is subject to sociolinguistic, stylistic and contextual variation. However, with atlases such as the AIS or the ALF you must “make the best of your incomplete data” (Garzonio & Poletto 2018), while accepting the limits of tools designed more than a century ago. A quantitative perspective is, in my opinion, the safest way to revive the data contained in atlases such as the AIS and the ALF by limiting possible biases that are intrinsic to traditional dialectological enterprises.

The data contained in the above sources have been first mapped into discrete (binary) variables and factors, which have been organized in a single spreadsheet. Supervised annotators worked mainly on the original maps of the AIS and ALF, which had already been digitized and can be freely downloaded or consulted on line. By looking at areal distributions, annotators had better chances to understand the phenomena under study, provide a more precise segmentation of the transcribed material (sometimes phonetic transcriptions can blur morphosyntactic structure), and, in the end, ensure a more correct annotation of the data. Each annotation was crosschecked by another annotator and, eventually, by myself.

We obtained a single spreadsheet containing three kinds of metadata:

- 1) Source: AIS/ALF, sentence and datapoint identifiers, geographical coordinates, region/province;
- 2) Dependent variables: presence/absence of N1 and N2;
- 3) Independent variables, e.g. clause type, mood, presence and type of a co-occurring NPI.

Clause typing was coded following a scale of features: force > mood/modality > tense/aspect. Declarative, indicative, perfect clauses were considered baseline values; to tag the remaining sentences, annotators chose the highest feature in the scale that differs from the baseline (for instance, all interrogatives have been tagged as ‘question’ regardless of their tense/aspect/mood specifications). A screenshot of the dataset—which can be freely downloaded from **XXX**—is shown in Fig. 3:

primary.sourID	sentence	DatapointID	Region	Dialect(AIS).X	Y	NEG1	NEG2	Jespersen.TV clause	NPI.type	person	Doubling.N2		
AIS	1605 lui.non.corre	1	Grigioni	Coira	9,06137594	46,7695626	0	0	0	indicative	never	3sg	0
AIS	1605 lui.non.corre	3	Grigioni	Coira	9,21613271	46,7312311	0	0	0	indicative	never	3sg	0
AIS	1605 lui.non.corre	5	Grigioni	Coira	9,44735449	46,8336355	0	0	0	indicative	never	3sg	0
AIS	1605 lui.non.corre	7	Grigioni	Coira	10,2016903	46,7749028	1	0	1	indicative	never	3sg	0
AIS	1605 lui.non.corre	9	Grigioni	Coira	10,3163076	46,8127453	1	0	1	indicative	never	3sg	0
AIS	1605 lui.non.corre	10	Grigioni	Coira	8,76550828	46,6795379	0	0	0	indicative	never	3sg	0
AIS	1605 lui.non.corre	11	Grigioni	Coira	8,93381139	46,7258434	0	0	0	indicative	never	3sg	0
AIS	1605 lui.non.corre	13	Grigioni	Coira	9,09660479	46,6560321	0	0	0	indicative	never	3sg	0
AIS	1605 lui.non.corre	14	Grigioni	Coira	9,4045762	46,7356424	0	0	0	indicative	never	3sg	0
AIS	1605 lui.non.corre	15	Grigioni	Coira	9,41415836	46,63619	1	0	1	indicative	never	3sg	0
AIS	1605 lui.non.corre	16	Grigioni	Coira	9,45957667	46,7169288	0	0	0	indicative	never	3sg	0
AIS	1605 lui.non.corre	17	Grigioni	Coira	9,56327316	46,6858661	0	0	0	indicative	never	3sg	0
AIS	1605 lui.non.corre	19	Grigioni	Coira	10,0949082	46,7004906	1	0	1	indicative	never	3sg	0
AIS	1605 lui.non.corre	22	Ticino	Lugano	8,94334185	46,5303956	0	0	0	indicative	never	3sg	0
AIS	1605 lui.non.corre	25	Grigioni	Coira	9,58170001	46,60927	0	0	0	indicative	never	3sg	0
AIS	1605 lui.non.corre	27	Grigioni	Coira	9,751944	46,63611	1	0	1	indicative	never	3sg	0
AIS	1605 lui.non.corre	28	Grigioni	Coira	9,95999329	46,6029209	1	0	1	indicative	never	3sg	0
AIS	1605 lui.non.corre	29	Grigioni	Coira	10,4238335	46,6019983	1	0	1	indicative	never	3sg	0
AIS	1605 lui.non.corre	31	Ticino	Lugano	8,78241937	46,4931899	0	0	0	indicative	never	3sg	0
ALF	1605 lui.non.corre	33	Ticino	Lugano	8,94361771	46,4711876	0	0	0	indicative	never	3sg	0

Figure 3. Dataset: each row correspond to an observation (i.e. a reply to the questionnaire).

The statistical analysis has been carried out by using the r package *Rbrul* (Johnson 2009). The software performs multiple logistic regression in order to assess how independent variables affect the distribution of dependent variables, i.e. the incidence of N1 and N2. *Rbrul* is the last descendant of the family of variable rule programs, which sociolinguists have been using extensively since the early 70’s to evaluate the effects of social and linguistic factors on a binary variable. *Rbrul*, like its predecessors, identifies which groups of factors

significantly affect the chosen dependent variable and, for each group of factors (e.g. age, clause type, phonological context), it weights to what degree specific factors affect the distribution of the independent variable. Since Rbrul cannot handle poliotomous variables, each possible negation pattern (N1, N1&N2, N2) was reduced to a combination of two binary variables (10, 11, 01).

In sections 3-5 I will examine the effect of two types of factors on the distribution of each dependent variable (N1 or N2): an external factor such as “Region” (i.e. where a dialect is spoken) and one or more grammatical factor(s). With the help of the Rbrul package, I will first ascertain whether models with multiple factors (geographic + grammatical) fit the data better than models with Region as a single factor. Second, I will try to establish, for each group of grammatical factors, which ones correlate better with the presence of N1 or N2.

As for grammatical factors, Romance dialects provide insightful data about how negation marking interacts with other syntactic properties (Zanuttini 1997; Manzini & Savoia 2005, vol. III: 130-338; Garzonio 2008; Garzonio & Poletto 2010; Dagnac 2015; Palasis 2015; Pescarini & Donzelli 2017; Poletto & Olivieri 2018; Guillot & Becerra-Zita 2019)., but, despite many researches have been carried out on the topic in recent times, “there are still too many unidentified factors that might play a role in the doubling mechanism” (Poletto 2016:837).

The present work, like most of the previous literature, concentrates on two main factors:

- 1) *Clause typing*. As previously mentioned in Section 1, most N2s derive from NPIs such as minimizers and were originally confined to specific pragmatic contexts. The syntax of N2 (and its interaction with N1) in present-day dialects may therefore reflect the original conditions in which N2 underwent grammaticalisation. For instance, N2 (like minimizers in general) is expected to occur less frequently in factive environments (Cinque 1991 [1976]) or, like minimizers, it is expected to occur without N1 in *nonveridical* contexts such as questions, imperatives, if-clauses, etc. (Giannakidou 1998).
- 2) *Negative concord*, i.e. the interaction between N1/N2 and NPIs such as negative quantifiers (‘anything/nothing’), adverbs (‘(n)ever, yet’), coordinators (Fr. *ni...ni*). N2 is often in complementary distribution with NPIs, but in some datapoints we find instances of negative concord between N2 and a NPI (Dagnac and Burnett 2016); we need to verify whether this correlates with the loss of N1 or not.

Other factors seem to play a role in the distribution of N1 and N2, but for feasibility reasons they cannot be addressed properly here. Poletto 2016:842-843 argues that (inner) aspect may play a role in the diachronic change that turned negative quantifiers into N2, as in Piedmontese dialects; in fact, in other northern Italo-Romance dialects, where this change has not happened yet, negative quantifiers can function as ‘emphatic’ N2 only with activity verbs. Focus might have played a role in the grammaticalisation of N2 deriving from polar particles. Other factors that may correlate with the distribution of N1 and N2 are the presence of subject clitics (Ashby 1981 a.o.), the presence of suppletive imperatives, or indefinite objects introduced by the preposition *de*. Zanuttini 1997 claims that negated imperatives are suppletive in dialects lacking N2. As for indefinite objects, Garzonio and Poletto (2018:13) hypothesize that indefinites are introduced by *de* in dialects exhibiting N2 (see also Manzini & Savoia 2005: 280-285). Unfortunately, issues other than those listed in 1 and 2 cannot be addressed properly in this work. As for subject clitics, the topic has so many ramifications that a separate paper is needed (in this work I will limit myself to verify whether person and number agreement has some effect on the distribution of N1 in the AIS). Imperatives are still awaiting to be encoded in the dataset. As for indefinite and partitive objects, the dataset does not contain enough data.

For reasons of feasibility, then, only one external factor (Region) and two grammatical factors (Clause typing and Type of NPI) will be tested throughout Sections 3-5. The effect of Person/number will be briefly discussed in Section 3. The list of factors and example of specific tags/values are given in Table 3.

Table 3. Independent variables

Independent variables	Tags
Region	e.g. ‘Lombardia’ (AIS), Bourgogne (ALF)

Clause typing	‘subjunctive’, ‘question’, ‘if-clause’, etc.
Type of NPI	‘never’, ‘anymore’, ‘nobody’, etc.
Person.number	1sg, 2sg, etc.

3. Negation marking in Italo-Romance (AIS)

As shown in Fig. 1, N2 is attested in northwestern regions such as Valle d’Aosta, Piedmont, Lombardy, southern Switzerland (where both Lombard and Rhaeto-Romance dialects are spoken), Emilia-Romagna, and in two southern datapoints where northern communities emigrated in the Middle Ages. Some instances of N2 are attested in Trentino and Liguria, whereas the eastern regions such as Friuli and Veneto—like all central and southern varieties—have a (predominant) N1 system. I therefore focused on the 191 AIS datapoints that belong to the former group of regions in order to obtain more robust statistical results (the relevant subset of dialects can be selected by using the ‘northwestern’ tag in the dataset). The two following subsections focus on N1 and N2, respectively.

3.1. N1 (AIS)

As previously mentioned in Section 2, northern Italo-Romance is not the most appropriate test bed to study the evolution across the various stages of Jespersen’s cycle because the sample of dialects exhibiting discontinuous negation is quite narrow.

A second issue with the AIS dataset (in fact, an issue with all linguistic atlases) is *collinearity*, i.e. too many tokens exhibit the same independent variables. To avoid collinearity (or, at least, to limit its effect), let us focus on Table 4, which illustrates the structure of the dataset with respect to our two main grammatical factors: Clause and NPI.

Table 4. Number of tokens per sentence/NPI type (AIS ‘western’ dataset).

Clause	NPI (Type)	No NPI
Subjunctive	0	382
Imperfect	191 (Anymore)	0
Conditional	0	191
Indicative	199 (anything)	193
	191 (Any-where.PP)	
	945 (never)	
	191 (yet)	
Future	0	191
Question	0	191
Modal	4 (Anymore)	187
	191 (Neither...nor)	
Imperative	0	764

Table 4 shows that most indicative sentences contain NPIs, whereas non-indicative clauses either do not contain a NPI or contain a type of NPI that is not present in indicative clauses. If I tried to test models including both factors—NPI Type and Clause—the statistical analysis would return unfaithful results because the data are not well balanced. I therefore decided to test two separate subsets of data (corresponding to the grey cells in Table 4): a) indicative clauses containing NPIs and b) all clauses containing no NPI.

For each subset of data, I tested mixed models containing both geolinguistic and grammatical factors to find out whether Region alone accounts for the data or whether mixed models including external and grammatical factors have a better predictive power. I found that Region is always a significant factor, as expected in a dataset containing data from tightly related languages. For this reason, the results concerning the factor Region will be systematically omitted from now on. In general, I found that Region is seldom sufficient to

account for the distribution of N1 and N2 and mixed models including syntactic factors usually fit the data best.

First of all, I tested whether factors Region and Clause are significant predictors of the distribution of N1 in the subset of tokens that do not contain NPIs. Rbrul found that both factors are statistically significant (Region $p=2.3e-249$; Clause $p=0.00053$). The weight of each value of the factor Clause is given in Table 5. All the following tables are organized in the same way: for each group of factors, tables report the number of Tokens (i.e. the number of negative clauses per type), the relative frequency of N1 in each type of clause (the number of clauses in which N1 is present divided by the number of clauses with available data), while the last column reports an index, ranging from 0 and 1, that shows how probably N1 will occur in a given type of sentence. Notice that probability may differ from frequency because the former is calculated by taking into account all type of factors that are included in the model, e.g. Clause *and* Region. By excluding and including factors, Rbrul tests different scenarios until it finds the model that fits the data best (the winning model can be the one without factors!) and, eventually, weights the value of each value/factor in the distribution of the dependent variable.

Table 5. Frequency/Probability of N1 in different types of clauses in a model with factors Region and Clause (AIS ‘western’ dataset, **excluding sentences containing NPIs**).

Clause	Tokens	Frequency of N1	Rbrul probability
Subjunctive	331	.49	.64
Indicative	193	.44	.54
Question	168	.42	.50
Future	187	.42	.50
Conditional	174	.41	.49
Modal	161	.37	.42
Imperative	641	.35	.41

Table 5 shows that (embedded) subjunctive clauses are the environment in which N1 is found most probably, whereas imperatives are the contexts with the lowest incidence of N1. Imperatives, however, deserve further attention as the way in which negative imperatives are syntactically encoded is subject to a great degree of variation. In most dialects, negative imperatives are not obtained by adding a negative marker to the positive imperative form, but instead they are expressed by a periphrasis with the verb ‘stay’ (e.g. Ven. *no sta partir* ‘do not leave’, lit. ‘not stay to leave’), a subjunctive form or an infinitive. Zanuttini (1997: 105-107) claims that suppletive imperatives are found in dialects without N2, while in dialects in which N2 is available, negative and positive imperative may have the same form (see Garzonio & Poletto 2018:5 for apparent counterexamples and discussion). At present, I cannot verify whether Zanuttini’s generalization is confirmed or not by the AIS/ALF dataset because the data on imperatives still await to be coded in our spreadsheet. However, without a clear indication about the incidence of suppletive imperatives, it seems to me that the comparison between imperatives *vs* other clauses is not trustworthy.

Another issue with imperatives is that they cannot exhibit subject clitics, which—according to previous corpus studies on French—may play a role in the retention of N1 (Ashby 1981). Northern Italo-Romance, as well as northern Occitan, is a promising area to investigate the relationship between subject clitics and negation. In these areas, subject clitics are mandatory (even if a DP subject occurs), but inventories of subject clitics are often defective (Poletto 2000; see Pescarini 2019 for a quantitative overview). In principle, one can therefore verify whether the probability of finding N1 increases or decreases in the contexts and dialects in which subject clitics are missing. This kind of study, however, requires a painstaking reconstruction of each clitic system, a goal that goes beyond the limits of the present article. I therefore limited myself to check how N1 varies depending on Person (and number), a factor that might be in turn related—albeit indirectly—to the syntax of clitics. I found that Person is a significant factor in a model containing factors Region, Clause, and Person (Region $p=7.7e-169$; Person $p=0.0066$; Clause is not significant). The data in Table 6 therefore provide some first indications regarding the interaction between subject clitics and N1: the frequency of N1 is lower at the 2/3sg and higher at the 3pl and 1sg (sentences containing 1pl subjects were omitted as 1pl clitics are

easily mistaken for N1). The data in Table 6 point towards various avenues of research. Phonologically, 1sg and 3pl clitics in many dialects have a vocalic formative that provides a nucleus on which the N1 marker *-n* can syllabify. The higher incidence of N1 may therefore receive a morphophonological explanation. Alternatively, one may argue that the ranking in Table 6 correlates with syntactic factors as vocalic clitics often occupy a higher syntactic position than other clitics (Poletto 2000) and therefore tend to precede N1, whereas the other clitics occur between N1 and V. As previously mentioned, these issues will remain open until a fine-grained analysis of subject clitics in the AIS datapoints will be carried out.

Table 6. Frequency of N1 w.r.t. Person (AIS, excluding sentences containing NPis).

Person.number	Tokens	Frequency of N1	Rbrul probability
3pl	331	.49	.62
1sg	371	.43	.49
2sg	168	.42	.47
3sg	344	.39	.42

I then tested how NPis affect the distribution of N1 in the AIS dataset. The data in Tab. 7 show the frequency and probability of N1 with respect to various classes of NPis and in negative sentences without NPis. Consider that the following probability rates were obtained in a model containing factors Region and NPI Type; only the subset of indicative clauses was taken into consideration to avoid collinearity (cf. Table 4). Both factors (Region and NPI Type) proved significant.

Table 7. Frequency/Probability of N1 w.r.t. NPis in a model containing factors Region and NPI Type (AIS dataset, ‘northwestern’ datapoints, **only declarative clauses**)

NPI Type	Tokens	Frequency of N1	Rbrul probability
Anywhere (PP)	175	.49	.60
Never	836	.47	.57
Absence of NPI	193	.44	.50
Anything (obj.)	179	.41	.47
Yet	174	.37	.36

If we take negative sentences lacking NPis as baseline, we find that, on average, NPis tend to favour the retention of N1, although the difference between clauses containing and not containing NPis is quite narrow. N1 is most probable when a NPI is embedded in an adjunct Prepositional Phrase, e.g. *in nessun luogo* ‘in any place’.

The syntax of the adverb ‘yet’, which seems to disfavour N1, needs elaboration. In most Italo-Romance dialects, ‘yet’ corresponds to a polysemous adverb (It. *ancora*) that conveys two aspectual values: repetitive (‘again’) and continuative (‘up till now’). Most Italo-Romance dialects do not display a polarity-sensitive alternation of the kind ‘still’/‘yet’, except for 25 datapoints that exhibit negative adverbs e.g. [na'mo], [naŋ'kora] that are negative counterparts of positive adverbs ['mo] ‘now’, [aŋ'kora] ‘again’. These negative adverbs occur predominantly in dialects without N1 and might be analysed as n-words (like Eng. *nothing*, *never*, etc.) that does not need to be licensed by N1. This explains why the occurrence of N1 in clauses containing ‘yet’ is less frequent/probable than in other negative clauses.

After having removed sentences containing ‘yet’, I grouped the values ‘anymore’, ‘anything’, ‘never’, and ‘anywhere PP’ together, thus obtaining a binary factor ‘presence’ vs ‘absence of NPI’, which proved not to be statistically significant in a model containing factors Region ($p=2.1e-170$) and NPI Type ($p=0.015$) (as above, the model was tested on the subset of indicative clauses). This amounts to saying that the presence of a generic NPis is not predictive of the distribution of N1 in the AIS dataset. Specific NPis, however, are better predictors of the behaviour of N1, which is more likely retained when the NPI is embedded in an adjunct PP.

Before addressing N2 (cf. Section 3.2), the remainder of this section elaborates on the occurrence of N1 in sentences with *N2-doubling*, i.e. in sentences in which N2 co-occurs with a NPI. In languages with discontinuous negation, N2 usually triggers a double negation reading when it co-occurs with another NPI, e.g. Fr. *Il n'a pas rien vu* 'It is not the case that he saw nothing'. However, in a few AIS datapoints we find examples of N2 + NPI combinations that, instead of triggering double negation effects, yield a pattern of bona fide negative concord. Theoretically, this may mean that in these varieties N2 has become a full-fledged negator that is able to license NPIs (a property that normally characterize N1-type negators, according to the preliminary typology given at the beginning of Section 1). One may therefore hypothesize that N2-doubling is allowed if and only if N1 is missing. In fact, however, I found a certain number of sentences in which both N1 and N2 co-occur with a NPI, mostly in examples in which the NPI is the argument of a preposition (as in the case of expressions such as It. *in nessun luogo* 'anywhere', but literary 'in no place') and in sentences containing a *neither... nor* coordination. If PPs and coordinations are removed, the number of cases of N2-doubling drops to 6.

N2-doubling proved to be a significant factor (along with Region; $p=0.00021$). Frequency and probability of N1 drop in sentences featuring N2-doubling, as shown in Table 8.

Table 8. Frequency/Probability of N1 w.r.t. N2-doubling in a model containing factors Region and N2-doubling (AIS dataset, 'northwestern').

N2-doubling	Tokens	Frequency of N1	Rbrul Probability
Absent	3414	.44	.64
Present	140	.13	.36

3.2. N2 (AIS)

In order to verify whether clause typing affects N2, I excluded from my sample all tokens containing NPIs, which are often in complementary distribution with N2 (the co-occurrence of N2 and NPIs will be addressed at the end of the present section). Imperatives and collinear values were removed for the reasons discussed in Section 3.1. Then I tested a model containing factors Region and Clause, both of which resulted significant (Region $p=4.9e-189$; Clause $p=5.8e-05$). Notice that core nonveridical environments such as 'Question', 'If clause', and 'Subjunctive' are the contexts in which N2 is less likely to occur. I will come back to this point later on, when commenting on the data in Section 5.

Table 9. Frequency/Probability of N2 w.r.t. Clause in a model with factors Region and Clause (AIS dataset, only 'northwestern' datapoints, excluding tokens containing NPIs and collinear values)

Clause	Tokens	Frequency of N2	Rbrul probability
Conditional	174	.75	.65
Future	187	.74	.63
Modal	161	.71	.53
Indicative	342	.68	.49
Question	168	.65	.42
If clause	174	.65	.41
Subjunctive	477	.62	.37

If Person is added to the model, Rbrul cannot weight all factors together because Clause and Person are highly collinear, as mentioned in Section 3.1.

I then tested NPIs, which, as previously mentioned are not free to co-occur with N2. The data in Table 10 confirm that not all NPIs are incompatible with N2: negative concord involving N2 is marginally allowed when NPIs are embedded in adjunct PPs or with a negative coordination of the type *neither... nor*.

Table 10. Frequency/Probability of N2 w.r.t. NPIs in a model containing factors Region and NPI Type (AIS dataset, only 'northwestern' datapoints, excluding 'yet' and collinear values)

Clause	Tokens	Frequency of N2	Rbrul probability
Anywhere (PP)	175	.36	.79
Neither... nor	160	.36	.78
Anything (obj.)	172	.09	.27
Never	836	.05	.17

3.3. Interim conclusion (AIS)

Mixed models including grammatical and geographical factors often perform better than models containing only geographical factors. As for N1, Clause, Person, NPI Type all proved significant, but I was not able to model all factors together due to collinearity.

As for Clause, (embedded) subjunctive clauses proved to be the environment in which N1 is more likely retained, but no clear distinction emerged between e.g. veridical/nonveridical contexts. The range of probability, in general, is quite narrow and the overall ranking of values is difficult to interpret under current analyses of negation marking. Conversely, I noticed that nonveridical clauses such as *if* clauses, questions, and embedded subjunctive clauses are the contexts in which N2 occurs less frequently.

Person seems to play a role in the distribution of N1 and I briefly commented on the possible morphological and syntactic reasons that might link Person and negation marking in dialects with subject clitics. The role of subject clitics, however, remains open to further research.

The absence vs presence of NPIs plays no significant role in the distribution of N1 in the AIS dataset, but N1 is retained more frequently in combinations with adjunct NPIs. Lastly, N2 and certain NPIs (adjunct PPs and negative coordinators) can marginally co-occur in a negative concord configuration. If NPIs are licensed by N2, N1 seldom occurs.

4. Negation marking in Gallo-Romance (ALF)

4.1. N1 (ALF)

Table 11 shows the structure of the ALF dataset with respect to factors Clause and NPI Type.

Table 11. Number of tokens per sentence/NPI type (ALF dataset).

Clause	NPI (Type)	No NPI
Conditional	0	638
Imperative	0	638
Imperfect	638 (Anymore)	0
Indicative	638 (Anymore)	3550
	359 (Nobody)	
	638 (Yet)	
Infinitive	1276	0
Question	0	638
Modal	638 (Neither...nor)	1276

As I did with the AIS dataset in Section 3.1, I first verified whether Clause is significant in determining the distribution of N1 in sentences not containing NPIs. Then I focused on the role of NPIs in indicative and modal clauses, which occur in the dataset with and without NPIs.

First of all, I tested whether clausal factors are significant in the distribution of N1. After having removed all tokens containing NPIs, I began testing models with factors Region and Clause. Both factors proved significant (Region $p \approx 0$; Clause $p = 3.3e-59$). Frequencies and probabilities of N1 with respect to Clause are reported in Table 12, which shows that indicative clauses are the context in which N1 is most disfavored (the usual caveats apply for imperatives, cf. Section 3.1).

Table 12. Frequency/Probability of N1 in different types of clauses in a model with factors Region and Clause (ALF dataset, excluding sentences containing NPIs).

Clause	Tokens	Frequency of N1	Rbrul probability
Imperative	568	.79	.80
Conditional	637	.66	.57
Infinitive	1275	.58	.44
Modal	1265	.56	.40
Question	635	.54	.38
Indicative	3537	.52	.37

Then I checked the role of NPIs in indicative clauses. Rbrul showed that factors Region and NPI Type are both significant in the distribution of N1 (Region $p=1.4e-260$; NPI Type $p=8.9e-165$). The data in Table 13 show that N1 is found more probably in sentences containing NPIs. This holds particularly true for sentences with a preverbal negative quantifier (factor: Nobody subj.): in this context the occurrence of N1 is almost at ceiling. In this respect, it is worth noting that the condition cannot be checked in the AIS dataset, which does not contain a comparable sentence. However, even if the AIS had had cases of preverbal quantifiers, the comparison would have been far from straightforward because most Gallo-Romance languages are *strict negative concord languages* in which a preverbal NPI must co-occur with N1, while most Italo-Romance dialects are *weak negative concord languages*. Unfortunately, we cannot ascertain the distribution of strong/weak negative concord systems in the AIS.

Table 13. Frequency/Probability of N1 w.r.t. NPI Type in a model with factors Region and NPI Type (ALF dataset, only indicative clauses)

NPI	Tokens	Frequency of N1	Rbrul probability
Nobody (subj.)	359	.96	.93
Yet	637	.78	.45
Anymore	633	.71	.33
Absence of NPI	3537	.52	.14

In terms of frequency and probability, the data in Table 15 show that N1 is often retained when a negative quantifier is in preverbal subject position (although for this condition the ALF does not provide data for all datapoints). Furthermore, N1 is likely to occur with the adverb ‘yet’. Unlike northern Italo-Romance (Section 3.1), Gallo-Romance dialects do not exhibit any n-word for ‘yet’. However, the incidence of N1 is relatively higher with ‘yet’ than in other negative clauses that contain N2 or postverbal NPIs. This may indicate that ‘yet’/‘still’ is not entirely polarity-neutral.

Eventually, I tested a mixed model with factors Region, NPI Type, and Clause. The model was tested on indicative and modal clauses, which occur in the ALF with and without NPIs (see Table 11). The structure of the relevant subset is given in Table 14. The NPI Type factor was simplified by eliminating the value ‘yet’ and by reducing the remaining values to a binary choice: presence vs absence of NPI. Notice, however, that the subset in Table 14 is not well-balanced because we already know that the incidence of N1 varies depending on the type of NPI it co-occurs with. Hence, the \pm NPI condition is not uniform across clausal contexts.

Bearing in mind the above caveat, I tested a model with factors Region, Clause (modal vs indicative) and NPI (presence vs absence). Factors Region and NPI proved significant and the data in Table 15 confirm that N1 is more likely to occur in sentences containing a NPI. The significance of the factor Clause was too volatile: it varied depending on whether sentence including preverbal *Nobody* were included or not in the sample. In any case, the values of the probability index for the factor Clause (modal vs indicative) were always too close to reach any solid conclusion.

Table 14. Subset of ALF sentences.

Clause	NPI	No NPI
--------	-----	--------

Indicative	638 (Anymore) 359 (Nobody)	3550
Modal	638 (Neither...nor)	1276

Table 15. Frequency/probability of N1 w.r.t. factor \pm NPI in a model with factors Region, Clause, \pm NPI (ALF subset in Table 14).

NPI	Tokens	Frequency of N1	Rbrul probability
Present	1629	.72	.66
Absent	4802	.53	.34

The remainder of the section focuses on the occurrence of N1 in combination with N2 and a NPI. As we will see in Section 4.2, the phenomenon I dubbed N2-doubling is marginally allowed in the ALF dataset only in two contexts: with the adverb *plus* ‘anymore’ and with the *ni ... ni* coordination. As in the case of Italo-Romance, N2-doubling proved to be significant with respect to the retention/omission of N1 in various models. As predictable, Table 16 shows that N1 is disfavored in sentences displaying N2-doubling.

Table 16. Frequency/probability of N1 w.r.t. factor N2-doubling in a model with factors Region and N2-doubling (ALF dataset, only sentences containing a NPI).

N2-doubling	Tokens	Frequency of N1	Rbrul probability
Absent	2698	.74	.66
Present	206	.41	.34

4.2. N2 (ALF)

In the Gallo-Romance corpus, the incidence of N2 is at ceiling in almost all contexts, as shown in Table 17.

Table 17. Counts of sentences containing/not containing N2 (ALF dataset)

Clause	No N2	N2
Conditional	3	634
If clause	6	630
Imperative	5	561
Indicative	3	3501
Infinitive	2	3501
Modal	90	630
Question	5	630

The distribution of N2 in the ALF is quite homogeneous even if NPIs are introduced in the model. Cases of N2-doubling, in fact, are quite sporadic, although the data in Table 18 show some interesting tendencies that can be compared with those observed in the literature (see Section 5). In the ALF dataset, ‘Anymore’ and, to a lesser extent, negative coordinators are marginally found to co-occur with N2, whereas negative quantifiers (in subject position) and the adverb *guère* are in complementary distribution with N2. ‘Yet’ always co-occur with N2, thus confirming the intuition that the adverb encoding ‘yet’ is not a NPI, although the data on N1 (Section 4.1) suggested that the ‘yet’/‘still’ alternation is not entirely polarity-neutral.

Table 18. Incidence of N2 w.r.t. NPI Type (ALF dataset)

Clause	Tokens	Frequency of N2
Yet	636	.99
Anymore	1846	.14
Neither ... nor	632	.09
Nobody (subj.)	359	.01
<i>Guère</i>	343	.00

4.3 Interim conclusions (ALF)

Clause typing has some effect on the distribution of N1 in the ALF dataset: indicatives are the context in which N1 is less retained, but the ranking of the other clause environments is difficult to interpret in the light of current theorizing. N2, by contrast, is at ceiling in all clausal contexts.

Regarding the effect of NPIs, N1 is retained more frequently in clauses containing a NPI, especially a preverbal one. N2 sporadically co-occurs with non-argumental NPIs. If a NPI and N2 co-occur, N1 is often, but not always dropped.

5. Combining datasets and reducing factors

In doing my statistical analysis in Sections 3 and 4, I have maintained the data from the AIS and the ALF separated for practical and methodological reasons. Although they seem contiguous on maps (cf. Fig. 1 and Fig. 2), Gallo- and Italo-Romance systems are separated by the Alps and, therefore, it is reasonable to hypothesize that the conditions ruling the distribution of N1/N2 in the two areas are not necessarily alike. For the sake of clarity, Table 19 resumes the conclusions of Sections 3 and 4:

Table 19. Interim conclusions

Factor	Effects on variables
Clause Typing	N1 varies across contexts, but no clear trend has been found. N2 is disfavored in nonveridical clauses (AIS); N2 is at ceiling in the ALF.
Negative Concord	N1 and N2 are favored by non-argumental NPIs. N1 is mandatory with preverbal NPIs (ALF) N1 seldom occurs if N2 and a NPI are both present.

Despite some difficulties and limitations, the preliminary observations made in Sections 3 and 4 allowed us to reach some provisional conclusions regarding both methodological and theoretical aspects. First and foremost, we proved that mixed models including grammatical factors always perform better than models that are entirely based on geographical factors. Both type of factors, i.e. grammatical and geographical, can be improved and refined, but it seems reasonable to conclude that combined models fit the data best.

Second, by examining the distribution of N1 and N2 in the two datasets, we noticed that the AIS and ALF are not homogeneous test beds, especially in the analysis of N1. Northwestern Italy provides no clear indication regarding the ‘loss’ of N1 as most AIS datapoints exhibit either N1 or N2, whereas the ALF is characterized by a relatively higher incidence of languages with discontinuous negation. Then, it seems to me that, to obtain better statistical results, the data from the AIS and the ALF should be merged together, rather than being compared. This holds particularly true for one group of factors (Clause Type) whose effects are not clear and not convergent in both Gallo- and Italo-Romance dialects (see Table 19). Furthermore, to strengthen statistical results, we may also want to focus, deductively, only on those specific factors that, according to the recent literature, are expected to affect the incidence of N1 and N2. For instance, since veridicality is often mentioned as factor affecting negation marking, we can test whether the incidence of N1 and N2 varies significantly in two groups of clauses—indicative (declarative) clauses *vs* interrogatives—without testing the whole dataset. However, even if all other types of clauses are removed and data from both the AIS and ALF

are merged together, I found that the distinction between indicative declaratives and interrogative was not significant with respect to N1 in a model with factors Clause and Region. Interestingly, however, clause proved statistically significant with respect to the distribution of N2 (0.015), which is more likely to occur in indicative than in questions, as shown in Table 20:

Table 20. Frequency/Probability of N2 w.r.t. in a model containing factors Clause and Region (AIS ‘western’ and ALF datasets, excluding sentences containing NPIs)

	Tokens	Frequency of N2	Rbrul probability
Indicative	3697	.98	.60
Question	803	.92	.40

The fact that N2 is disfavoured in interrogatives is in line with Cinque’s (1991 [1976]) analysis of incipient discontinuous negation. On the contrary, the fact that N1 is not particularly sensitive to clause typing does not support an analysis of discontinuous negation as an instantiation of negative concord: if N2 was analysed as a NPI (recall that most N2 etymologically derive from NPIs such as minimizers), one would expect N1 to be disfavoured in nonveridical contexts, where NPIs can be licensed even if N1 is missing. However, since no meaningful difference was found between declaratives and interrogatives, nonveridicality seems to play no role in the loss/retention of N1.

By combining the data from both atlases, we can also obtain a richer taxonomy and eventually draw a finer picture of the role of single factors. Negative concord is a case in point because the data from the AIS and ALF seem to converge (cf. Table 19), but both atlases provide us with incomplete indications. Conversely, by merging both, we may obtain a contingency table such as Table 21, reporting the frequency of each pattern of negation (absence of negation, N1, N2, discontinuous negation) per type of NPI.

Table 21. Incidence of N2-doubling w.r.t. NPI Types in clauses containing/not containing N1 (AIS ‘western’ and ALF dataset)

NPI	0	N1	N2	N1+N2
Anymore	0,29	0,59	0,07	0,05
Anywhere	0,22	0,42	0,29	0,07
Neither ... nor	0,29	0,57	0,13	0,02
Anything (postV)	0,51	0,41	0,07	0,01
Never	0,48	0,47	0,05	0,01
Nobody (preV)	0,04	0,95	0,00	0,01

The data in Table 21 are plotted in Figure 4.

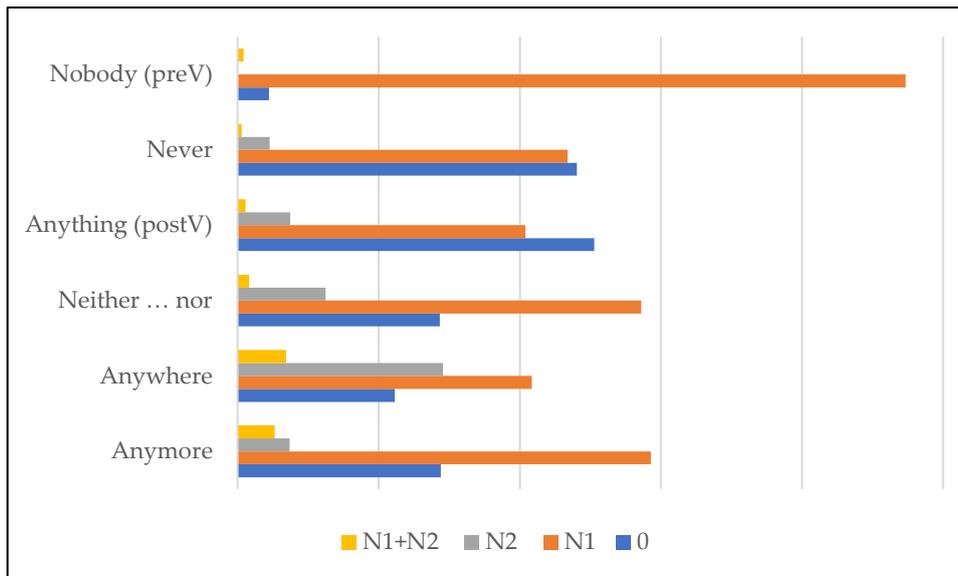


Figure 4. Patterns of negation marking per type of NPI (ALF+AIS dataset).

Figure 4 shows that negation marking varies significantly depending on the type of NPI. Preverbal quantifiers are in complementary distribution with N2 and, in strict negative concord languages, always require N1. Postverbal quantifiers (and the adverb *never*) do not trigger negative concord in around 50% of the sentences, whereas negative coordinators and the adverb *anymore* favor negative concord with N1. *Anywhere* stands out because it co-occurs with N2 quite frequently and it seldom co-occurs with both N1 and N2.

Dagnac and Burnett (2016 and references therein) reports similar data from Picard dialects and Montréal French. Notice that, while N1 is completely lost in Montréal French, Picard dialects still retain N1, but this seems completely orthogonal to the co-occurrence of N2 and NPIs. Table 22 shows that different types of NPIs allow, to various extents, negative concord with N2.

Table 22. Incidence of N2 with various types of NPI in Montréal French and Picard dialects (Dagnac and Burnett 2016)

Clause	Montréal	Picard
Anywhere	.83	.90
Anybody	.59	.33
Anything	.15	.13
Any	.11	-
Never	.1	.10
Anymore	*	.04

The AIS/ALF data in Table 21 and Dagnac and Burnett's (2016) data in Table 22 confirm that *Anywhere* is the type of NPI with which N2 co-occurs more frequently. Table 21 seems to differ from Table 22 with respect to the behaviour of *Anybody*, which in Montréal and Picard triggers negative concord with N2 with a relatively high frequency, whereas in the ALF *Anybody* never co-occurs with N2. However, recall that in the ALF corpus *Anybody* is always a preverbal subject, whereas Dagnac and Burnett reports aggregate data. They nonetheless notice that preverbal position blocks (in Montréal French) or disfavors (in Picard) negative concord with N2 (Dagnac and Burnett 2016: 9).

Another difference between Table 21 and Table 22 regards the relatively higher incidence of N2 doubling with *Anymore*. This, however, may be a specificity of certain geolinguistic areas. In fact, most sentences with N2-doubling in the AIS/ALF are concentrated in two specific areas, shown in Fig. 5, but, crucially, they are almost unattested in the Picard area studied by Dagnac and Burnett 2016.

The fact that NPIs and N2 can co-occur, in certain dialects and with certain NPIs, deserves further attention, but it indicates that N2s are marginally involved in negative concord configurations. This provides further evidence supporting the hypothesis that N2s are not polarity items, but full-fledged negative elements that concur to the licensing of NPIs under certain syntactic conditions, which need to be clarified.

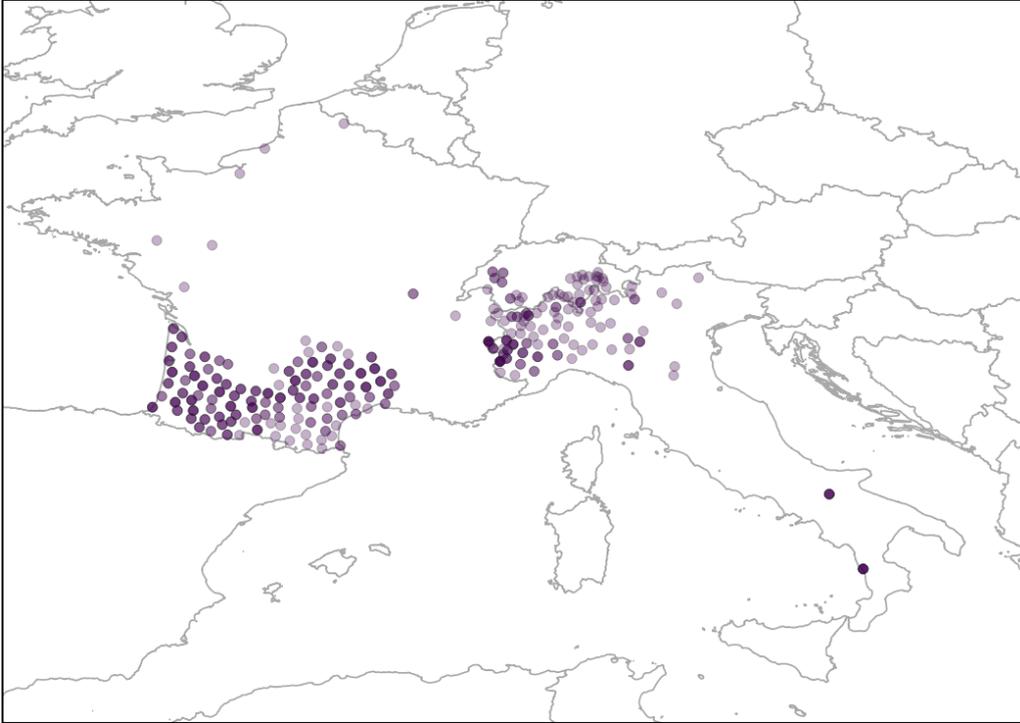


Figure 5. Distribution of N2-doubling patterns (ALF+AIS dataset).

6. Conclusions and open issues

The intent of this paper was twofold: methodological and theoretical. On the methodological side, this work aimed at exploring the feasibility of multiple linear regression on raw data collected by dialectological atlases. I showed that, to do statistical analysis, we need to adapt datasets, merge various primary sources, and aggregate factors in order to avoid collinearity and obtain significant statistical results.

On the theoretical side, I focused on the distribution of N1 and N2 across central Romance dialects and across syntactic contexts. I proved that models including both geolinguistic and grammatical factors always fit the data better than models in which grammatical variables are not taken into account.

In my study, I focused on two kinds of grammatical variables: i) clause typing (in particular, I tried to verify whether veridicality affects the distribution of N1/N2) and ii) negative concord, i.e. the interaction between N1/N2 and NPIs. My preliminary results indicate that the role of veridicality in negation marking is not proven, while factivity might still be a factor in the cross-contextual distribution of (statu nascendi) N2.

On average, the presence of NPIs correlates positively with the presence of N1, but not all NPIs favour N1. Certain NPIs (coordinators and adverbs) favor negative concord with N1. Preverbal quantifier disfavour negative concord with N2. Adverbial PPs allows concord with N2 and, marginally, with N1 and N2. These conclusions are in line with the literature on Gallo-Romance (Dagnac and Burnett 2016).

I mentioned in Sections 2 and 3.1 that many empirical and theoretical questions will remain open because we still lack detailed metadata concerning other grammatical phenomena that may interact with negation marking such as the morphosyntax of imperatives, the distribution of subject clitics, verb movement (e.g. in infinitives), the role of N2 in licensing indefinite objects introduced by the preposition *de*, etc. Some of the above properties are arguably related to the typology/etymon/shape of N2s, which needs to be encoded in our dataset. To shed light on these issues we need to integrate data from more recent dialectological enterprises such as the ASIt (Syntactic Atlas of Italy), the Thesoc (Thesaurus Occitan), or the wealth of data published in

reference works such as Manzini & Savoia 2005. I believe it is worth distinguishing the data elicited from monolingual speakers (such as most of those interviewed for the AIS/ALF in the late 19th or early 20th century) from the data gathered in the late 20th or early 21st century by speakers that, to various extents, are probably dialect/standard language bilingual. I will therefore conduct in a separate study a comparison between the two sets of data in order to verify whether and to what extent negation marking has changed over a century.

References

- (Breitbarth 2013) Breitbarth, Anne. 2013. Indefinites, Negation and Jespersen's Cycle in the History of Low German. *Diachronica* 30: 171—201. doi 10.1075/dia.30.1.02bre
- (Cinque 1991 [1976]) Cinque, Guglielmo. 1991 [1976]. Mica: note di sintassi e pragmatica. In *Teoria linguistica e sintassi italiana*, by Guglielmo Cinque. Bologna: Il Mulino, 311-323.
- (Cinque 1999) Cinque, Guglielmo. 1999. *Adverbs and Functional Heads: A Cross-Linguistic Perspective*. Oxford: Oxford University Press.
- (Dagnac and Burnett 2016) Dagnac, Anne and Burnett, Heather. 2016. Concordance négative optionnelle : contrastes forts et faibles entre picard et québécois. *Web of Conferences* 27: 13003.
- (Garzonio and Poletto 2018). Garzonio, Jacopo and Poletto, Cecilia. 2018. Exploiting microvariation: How to make the best of your incomplete data. *Glossa: A Journal of General Linguistics* 3(1): 112. DOI: <http://doi.org/10.5334/gjgl.556>
- (van Gelderen 2008) Gelderen, Elly van. 2008. Negative Cycles. *Linguistic Typology* 12: 195-243.
- (Giannakidou 1998) Giannakidou, Anastasia. 1998. *Polarity Sensitivity as (Non)Veridical Dependency*. Amsterdam: Benjamins.
- (Gilliéron and Edmont 1902-1910). Gilliéron, Jules and Edmont, Edmond. 1902-1910. *Atlas linguistique de la France*. Paris: Champion.
- (Jaberg and Jud 1928-1940). Jaberg, Karl and Jud, Jakob. 1928-1940. *Sprach- und Sachatlas Italiens und der Südschweiz*. Bern: Zofingen.
- (Johnson 2009). Johnson, Daniel Ezra. 2009. Getting off the GoldVarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis. *Language and Linguistics Compass* 3(1): 359–383.
- (Larrivée 2008). Larrivée, Pierre. 2008. The pragmatic motifs of the Jespersen cycle: Default, activation, and the history of negation in French. *Lingua* 120: 2240–2258
- (Manzini and Savoia 2005) Manzini, Maria Rita and Savoia, Leonardo. 2005. *I dialetti italiani e romanci. Morfosintassi generativa*. Alessandria: Edizioni dell'Orso.
- (Manzini and Savoia 2012) Manzini, Maria Rita and Savoia, Leonardo. 2012. Sentential negation in Piedmontese varieties. *Quaderni di lavoro ASIt* 13: 135-159.
- (Pescarini 2019). Pescarini, Diego. 2019. Microvariation and Microparameters. Some Quantitative Remarks. *Quaderni di Linguistica e Studi Orientali* 5: 255-277. DOI: <http://dx.doi.org/10.13128/QUISO-2421-7220-25968>
- (Pescarini and Donzelli 2017) Pescarini, Diego and Donzelli, Giulia. 2017. La negazione nei dialetti della Svizzera italiana. *Vox Romanica* 76: 74-96. doi: 10.2357/VOX-2017-004
- (Pescarini and Penello 2012) Pescarini, Diego and Penello, Nicoletta. 2012 L'avverbio mica fra widening semantico e restrizioni sintattiche' In *Linguaggio e cervello / Semantica* (Proceedings of the XLII annual meeting of the Società di Linguistica Italiana - Pisa, Scuola Normale Superiore, 25-27 september 2008), ed. by Pier Marco Bertinetto, Valentina Bambini e Irene Ricci. Rome: Bulzoni. ISBN: 978-88-7870-652-1
- (Poletto 2000). Poletto, Cecilia. 2000. *The higher functional field. Evidence from northern Italian dialects*. Oxford: Oxford University Press.
- (Poletto 2016). Poletto, Cecilia. 2016. Negation. In *Oxford guide to the Romance languages*, ed. by Martin Maiden and Adam Ledgeway. Oxford: Oxford University Press.
- (Poletto and Cornips 2005) Poletto, Cecilia and Cornips, Leonie. 2005. On standardising syntactic elicitation techniques. Part I. *Lingua*, 115: 939-957.
- (Schwenter 2005) Schwenter, Scott A. 2005. The pragmatics of negation in Brazilian Portuguese. *Lingua* 115: 1427–1456.

(Zanuttini 1997). Zanuttini, Raffaella. 1997. *Negation and Clausal Structure: A Comparative Study of Romance Languages*. Oxford: Oxford University Press.