

Agreement syncretisation and the loss of null subjects:

Quantificational models for Medieval French

Alexandra Simonenko¹, Benoit Crabbé², and Sophie Prévost³

¹FWO/UGhent, alexandra.simonenko@ugent.be

²LLF, CNRS/Paris Diderot, USPC,

benoit.crabbe@linguist.univ-paris-diderot.fr

³LATTICE, CNRS/ENS/Paris Sorbonne Nouvelle, USPC & PSL,

sophie.prevost@ens.fr

November 24, 2017

Address:

Alexandra Simonenko

Ghent University

Faculty of Arts and Humanities

Department of General Linguistics

Blandijnberg 2, B-9000 Ghent, Belgium

Short title: Agreement syncretisation and the loss of null subjects

Abstract

Agreement syncretisation and the loss of null subjects:
quantificational models for Medieval French

This paper examines the nature of the dependency between the availability of null subjects and the “richness” of verbal subject agreement, known as Taraldsen’s Generalisation (Taraldsen 1980, Rizzi 1986, Adams 1987), from the point of view of grammar change in Medieval French based on corpus data. We present a corpus-based quantitative model of the syncretisation of verbal subject agreement spanning the whole Medieval French period and evaluate two hypotheses relating agreement and null subjects: one relating the two as reflexes of the same grammatical property and a variational learning-based hypothesis whereby phonology-driven ending syncretization creates a learning bias against the null subject grammar. We show that only the latter approach has the potential to reconcile the intuition behind Taraldsen’s Generalisation with the fact that it has proven non-trivial to formulate the notion of agreement richness in a way which would unequivocally predict whether a language has null subjects.

Acknowledgements: This work has been supported by the Labex EFL and Research Foundation Flanders. The authors are very thankful to the editors and two anonymous reviewers of *Language Variation Change* for extremely helpful comments, Yves Charles Morin, Henri Kauhanen, George Walkden, Paul Hirschbühler, Philippe Schlenker, and Hedde Zeijlstra for discussions and suggestions. The project has benefited from the feedback from the audiences at WCCFL 34, DiGS 2016, XLanS: Triggers of language change in the Language Sciences, workshops Linguistic Knowledge & Patterns of Variation, Texts, Tools, and Methods in Digital Classics and Medieval Studies and seminars at Institut Jean Nicod, Université Paris 7 Diderot, and the University of Manchester.

1 Introduction

This paper examines the nature of the relation between the availability of null subjects and the “richness” of verbal subject agreement, known as Taraldsen’s Generalisation (Adams 1987; Rizzi 1986; Taraldsen 1980), from the point of view of grammar change in Medieval French based on corpus data. We fill in a major gap in the empirical base by providing the first large-scale corpus-based study of agreement syncretisation, and use these data to test and compare a model which postulates a strict clause-level relation between subject expression and surface agreement and one which assumes a much looser connection between the two, which emerges in the process of variational learning.

The original generalisation based on synchronic observations states that a language having sufficiently discriminating, or non-syncretic, subject agreement entails the possibility of non-expression of subjects. This insight generated a series of null subject treatments which assume some sort of a grammatical dependency between the properties of a head responsible for verbal subject agreement and (non)expression of subjects (e.g. Alexiadou and Anagnostopoulou 1998; Poletto 2006; Roberts 2010; Sheehan 2016).

In terms of diachronic developments, it was argued, even earlier, that there is a causal relation between the loss of non-syncretic subject agreement and the emergence of obligatory subject pronouns (e.g. Ewert 1943; Vennemann 1975), the underlying intuition being that overt subjects take over the role of identifying the subject’s person which can no longer be fulfilled by verbal inflection due to its phonological erosion.¹ Haspelmath (1999, 14) says that “... in languages that are losing their rich subject agreement morphology on the verb ... speakers will increasingly tend to choose the option of using the personal pronoun, because the verbal agreement does not provide the information required for referent identification in a sufficiently robust way”.

This diachronic scenario, however, was questioned for Medieval French by Schøsler (2002) and Roberts (2014) on the grounds of an apparent temporal lag between the two changes. Paradoxically, however, these authors make opposite assumptions about the

¹E.g. Vennemann (1975, 298): “As *s* (verbal subject agreement – the authors) is rendered non-dependable by phonological change ... once-“emphatic” *S*-pronouns become obligatory in topical function (and as such unstressed) to avoid ambiguity.”

temporal sequence of changes. While Schøsler (2002) assumes that pro-drop was largely lost at the beginning of the Medieval French period, while verbal inflection was undergoing phonological restructuring for another several centuries, Roberts (2014) dates the total loss of the rich agreement by the XII c. and the completion of the loss of null subjects by the XVI c.

A major discrepancy in assumptions about the empirical base arose first of all due to the unavailability of a systematic quantitative study of syncretisation. We present a corpus-based quantitative model of the syncretisation of verbal subject agreement spanning the whole Medieval French period and use it, together with a model of the loss of null subjects, to evaluate two hypotheses relating these phenomena.² First, we test the predictions generated by the hypothesis that null subjects and non-syncretic agreement exponents are related at the clause level, both being dependent on the same functional head.

The second hypothesis we explore is based on Yang's (2002) variational learning model whereby the agreement exponents and subject expression are not strictly connected at the clause level. Instead, in the process of language learning (possibly over the speaker's lifespan) syncretic endings create a bias against the null subject grammar, which eventually drives it to extinction.

The paper is organized as follows. The next section gives a background on subject expression and verbal subject agreement in Medieval French. In section 3 we present the hypothesis about a clause-level relation between subject expression and agreement type and in section 4 we explore the predictions it makes. In section 5 we flesh out an approach which relates the two phenomena in the process of language learning. Section 6 concludes.

²A previous quantitative study of this topic by Ranson (2009) was done on samples extracted from 3 texts (the corpus we use is composed of 32 texts which amounts to approximately 1 million words). We will discuss her results in section 4.4.

2 Null subjects and subject agreement in French

The non-expression of referential subjects occurred in Medieval French, and Old French in particular, in contexts where their expression would be obligatory in Modern French (e.g. Foulet 1928). Observations by traditional grammarians have been supported by quantitative studies (Fontaine 1985; Hirschbühler 1992; Kaiser 2009; Prévost to appear). Below we replicate their results on largest parsed corpora of Medieval French currently available, MCVF (2010) and Penn Supplement to MCVF (2010).

2.1 Rise of overt pronominal subjects

Medieval French texts abound in examples of pro-drop impossible in Modern French, such as the following one.

- (1) Car bien \emptyset set qu' il nes porra pas retorner de ceste emprise.
 Since well (he) knows that he not.them will not divert from this plan
 'Since he knows well that he will not turn them away from this plan.'³

During the Medieval period such cases become more and more rare, which is illustrated in Fig. 1 separately for main and subordinate clauses. Our estimates are based on the data coming from the corpus of the project “Modéliser le changement : les voies du français” (MCVF) and Penn Supplement to MCVF (2010). The MCVF corpus together with Penn Supplement to MCVF (2010) include 35 syntactically parsed texts (total of \approx 1mln words).⁴ More information about corpus structure and query format, as well as the list of the corpus texts is given in Appendix B. We used all texts excluding only the three texts composed in Canada. On the assumption that null subjects correspond to phonologically null personal pronominal elements, observations about the emergence of overt subjects are given here as the estimated probability of overt personal pronominal subjects against null subjects, with demonstrative, nominal and other kinds of overt subjects being excluded from consideration.⁵

³(1225-QUESTÉ-P,22.562). For all examples from MCVF (2010) and Penn Supplement to MCVF (2010) we give their unique ID tag consisting of the corpus text name from which an example was taken, the date, and the sequential order number of the utterance in the text. By this tag an utterance can be unambiguously located in the corpus.

⁴We refer to a version of the corpus available as of May 2017.

⁵This assumption, based on a close distributional equivalence of null subjects in pro-drop languages

Our dataset includes all finite clauses with either an overt pronominal or null subject (total of 56,615), excluding imperatives, subject relatives and wh-questions targeting subjects because of their idiosyncratic subject syntax.⁶ We also excluded all coordinated clauses introduced by the coordinating conjunction *et* and the conjunctive adverb *si* since those license subject ellipsis throughout the Medieval period and, by hypothesis, could be generated either by a null subject grammar or by an overt subject grammar on the assumption that subject ellipsis and pro-drop are subject to different sets of grammatical constraints.⁷ Specifically, these connectives are very likely to indicate the presence of an antecedent required for subject ellipsis. These connectives are sometimes used even when there is no potential antecedent in preceding clause. This means that by excluding clauses with *et* and *si* indiscriminately we also may exclude some relevant contexts (i.e. those where true subject drop can be expected). We consider this methodological move harmless, however, given the nearly stable rate of subject omission with *et* and *si*, as Fig. 13 in Appendix B shows. We take this to mean that there are few true subject omission environments with these connectives.⁸ Subject ellipsis under coordination with *et* is still allowed in Modern French, as the example below shows, while *si* itself fell out of use as a conjunctive adverb.⁹

- (2) Pierre aime beaucoup ses amis et protège Natasha.
 ‘Pierre loves a lot his friends and protects Natasha.’

As has been noted before on multiple occasions, subordinate clauses favour overt subjects more than main clauses (Adams 1987; Franzén 1939; Foulet 1928; Hirschbühler and overt pronominal subjects in obligatory subject languages, was made in previous works on null subjects such as Hirschbühler (1992). The assumption is also warranted by the fact that the rate of non-pronominal overt subjects which are not personal pronouns stays the same throughout the Medieval period, whereas the rate of overt pronominal subject increases and the rate of null subjects decreases in a dramatic fashion.

⁶Null subjects correspond to a special tag in the MCVF annotation scheme. We considered overt personal subjects separately from demonstratives, grouping the latter together with nominal subjects on the basis of their distribution. We counted as nominal subjects all subject phrases headed by a noun, both animate and inanimate, as well as nominalized adjectives, numerals, quantifiers, and free-relatives, disregarding their syntactic position with respect to the finite verb.

⁷We thank an anonymous reviewer for bringing to our attention to the fact that both *et* and *si* license subject ellipsis.

⁸We thank an anonymous reviewer for raising this worry.

⁹There are a handful of other (less frequent) conjunctive adverbs, such as *puis*, as well as a disjunction *mais* which seem to license subject ellipsis in Modern French, and which we did not exclude from our dataset since this would require an exhaustive study of ellipsis licensing conditions in Medieval French, a topic which goes beyond the scope of this paper.

1992; Prévost to appear; Roberts 2014; Vance 1997; Zimmermann 2014, among others). In fact, initially Adams (1987) concluded that there is a qualitative asymmetry between main and subordinate clauses in that the true subordinates, that is, those which could not be analyzed as having a main clause structure, do not allow for null subjects. This proposal pioneered the analysis of null subjects in Old French as dependent on V-to-C movement, assumed to happen in main but not in subordinate clauses. Later the initial empirical generalisation was shown to be incorrect on a larger set of data, as null subjects can be found in all types of subordinates, though to a far lesser extent (Fontaine 1985; Hirschbühler and Junker 1988; Kaiser 2009; Prévost to appear; Roberts 1993).¹⁰ Fig. 1 illustrates the quantitative contrast based on the data from MCVF (2010) and Penn Supplement to MCVF (2010). In this work we will not try to provide an explanation for the quantitative lag.

Absolute numbers of null and overt pronominal subjects for each text are given in Table 18 in Appendix B and the estimated probabilities are plotted in Fig. 1.

¹⁰To capture the alleged qualitative asymmetry and also the historical fact that null subjects disappear more or less within the same timeframe as constructions where the finite verb is preceded by a non-extraposed constituent other than the subject (the so called V2 constructions), Adams (1987) proposed that syntactically null subject corresponded to a *pro* which can only be licensed if it is governed by the verb in C. On the assumption that C is occupied by the complementizer in subordinate clauses and that V2 orders involve a V-to-C movement, this proposal captured the relevant correlation. To rescue Adams' (1987) intuition about the relation between V2 and null subject disappearance, Roberts (1993, 101,125) proposed a layered Agr structure for subordinate and V1 main clauses that featured null subjects. On this proposal, even in configurations where the verb could not raise to C, it could raise to a position (Agr₁) governing a *pro* in Spec Agr₂. Roberts (1993) relates the disappearance of null subjects to the change in subject licensing parameter from “under government” to “under agreement” (i.e. in a Spec-Head relation).

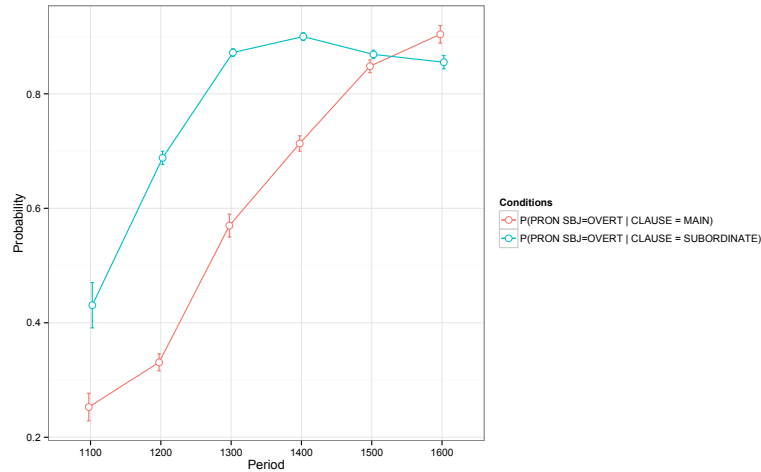


Figure 1: Overt pronominal subjects in main and subordinate clauses

2.2 Subject agreement syncretisation

French went from a language characterised by non-syncretic agreement inherited from Late Latin to a language with a largely syncretic agreement paradigm (Bettens 2015; Buridant 2000; Dees et al. 1980; Foulet 1935; De Jong 2006; Marchello-Nizia 1992; Morin 2001). As an example of Modern French verbal paradigms, consider the present tense of the verb aimer “to love”, representative of a large conjugation class with infinitives in -er which we will call Group I, and of the verb partir “to leave”, which represents Group II, where we will put all other conjugation types with the exception of être “to be” and avoir “to have” whose paradigms never became syncretic.

1ST SG	aime [ɛm]	1ST PL	aimons [ɛmɔ̃]
2ND SG	aimes [ɛm]	2ND PL	aimez [ɛme]
3RD SG	aime [ɛm]	3RD PL	aiment [ɛm]

Table 1: Modern French paradigm of a Group I verb aimer “to love”

1ST SG	pars [paʀ]	1ST PL	partons [partɔ̃]
2ND SG	pars [paʀ]	2ND PL	partez [parte]
3RD SG	part [paʀ]	3RD PL	partent [part]

Table 2: Modern French paradigm of a Group II verb partir “to leave”

Phonologically, verbal forms used in all contexts except for the 1st and 2nd person plural subjects in Group I and 1st, 2nd, and 3rd person plural in Group II are identical.

We can say that there is no systematic person marking on the verb in Modern French, and the only subject agreement feature present is number.¹¹ In contrast, as evidenced by the system of rhymes used in Old French versification to be discussed in section 4.5.1, verbal paradigms had a much less syncretic nature during that period (e.g. Bettens 2015).

The main changes in verbal agreement are shown in the tables below. We boldfaced changes which resulted in syncretism with respect to the person feature.

Group I verbs syncretise between 1st and 3rd person singular in present indicative and subjunctive, taking an ending -e.¹²

	PRESENT INDICATIVE	PRESENT SUBJUNCTIVE
1ST SG	aim > aime “love”	aim > aime
2ND SG	aimes	ains > aimes
3RD SG	aimet > aime	aint > aime

Table 3: Agreement changes in Group I present indicative and subjunctive

Another locus of syncretisation for this group is between 1st and 2nd person singular in imperfect and future conditional. **There is no syncretisation in the preterite and future in this time period. This will occur later when word-final -s ceases to be pronounced.**

	PRETERITE	FUTURE	IMPERFECT	FUTURE CONDITIONAL
1ST SG	amai	amerai	amoie > aimoies	ameroie > ameroies
2ND SG	amas	ameras	amoies	ameroies
3RD SG	amat > aima	amerat > aimera	amoiet	ameroit

Table 4: Agreement changes in Group I preterite, future, and imperfect

Group II verbs syncretise between the 1st and 2nd person singular in present indicative and preterite **with the addition of -s to the 1st person singular**, and between 1st and 3rd person singular in present subjunctive, as well as, identically to Group I, between the 1st and 2nd person singular in imperfect and future conditional, as Tables

¹¹Here we are setting aside the question of whether subject clitics in Modern French function as subject agreement markers (see De Cat (2005) for a discussion).

¹²In addition to ending changes, verb *aimer* “to love” had its root vowel grapheme vary between *a* and *ai*, the latter variant eventually winning. This alternation is not reflected in our tables.

5–6 illustrate.

	PRESENT INDICATIVE	PRESENT SUBJUNCTIVE
1ST SG	voi > vois “see”	voie
2ND SG	vois	voies
3RD SG	voit	voiet > voie

Table 5: Agreement changes in Group II present and imperfect subjunctive

	PRETERITE	FUTURE	IMPERFECT	FUTURE CONDITIONAL
1ST SG	vi/dormi > vis/dormis “see/” “sleep”	verrai	voioie > voioies	verroie > verroies
2ND SG	veïs/dormis	verras	voioies	verroies
3RD SG	vit/dormit > vi[t]/dormi[t]	verrat > verra	voioit	verroit

Table 6: Agreement changes for Group II preterite, future, and imperfect

As the tables show, in addition to the changes which led to an increase in syncretism, there were also changes which did not, such as the loss of final -t in preterite and future of the 3rd person singular of both groups of verbs.

Overall, there are three classes of changes that resulted in syncretism, namely, the drop of the final -t after vowels, e-insertion, and s-insertion. On the hypothesis of Dees et al. (1980) and van Reenen and Schøsler (1987) stating that e-insertion was a compensatory process “keeping” root consonants from the final position where they would have fallen, the first two changes can be seen as related. As we will see below, they are also much closer in time and in how they spread to each other than to the third one, s-insertion.

In this section we will be distinguishing ending change environments in terms of a) subject person and b) verbal group, and c) tense/aspect. We assume that the tense/aspect parameter is relevant only for its impact on the phonological environments. For instance, there are reasons to believe that glides patterned with consonants rather than with vowels in terms of the timing of a phonological change in the environment they create (Buridant 2000, 250). We therefore treat present and preterite apart from imperfect and future conditional since the morphology of the latter involves a diphthong ending in a glide. **The summary of the changes is given in (3).**

- (3) A. innovative final -e: 1st person, Group I, present indicative & subjunctive;

- B. innovative final -e (i.e. becoming final as a result of the drop of -t): 3rd person, Group I, present indicative & subjunctive;
- C. innovative final -e: 3rd person, Group II, present subjunctive;
- D. innovative final -a: 3rd person, Group I, preterite & future indicative;
- E. innovative final -a: 3rd person, Group II future indicative;
- F. innovative final ∅: 3rd person, Group II, preterite;¹³
- G. innovative final -s: 1st person, Group II, present and preterite indicative;
- H. innovative final -s: 1st person, Group I, imperfect and future conditional;
- I. innovative final -s: 1st person, Group II, imperfect and future conditional.

In Appendix A we illustrate variation in ending realization with corpus examples.

There are strong reasons to assume that as far as verbal endings and their syncretisation is concerned, the changes we have just presented exhaust the relevant phenomena and can be used to model phonological changes at least until the XIV c. One of the strongest arguments in support of the spelling reliability for phonological reconstruction is the novel observation based on quantitative data that the dropping of the final -t in verbs with stems ending in u/i is abruptly arrested and reversed just after the mid-XIV, when the French Royal Chancellerie is known (first mention 1342) to have introduced exams for the scribes requiring them to adhere to the standardised spelling rules (De Jong 2006, 25). While spelling unification had been taking place already for several decades, De Jong (2006) observes a sharp increase in what she calls “parasitic consonants” after around 1340, which she attributes to the prescriptions of the official examiners. Consequently, after that point we can only estimate verbal syncretism based on the change trajectory in the manuscripts written before that date. We will discuss this matter at length in section 4.5.1 once we have presented quantitative data on ending changes.

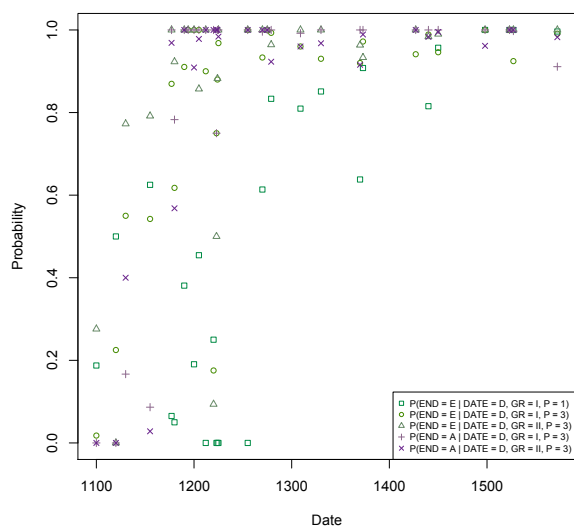
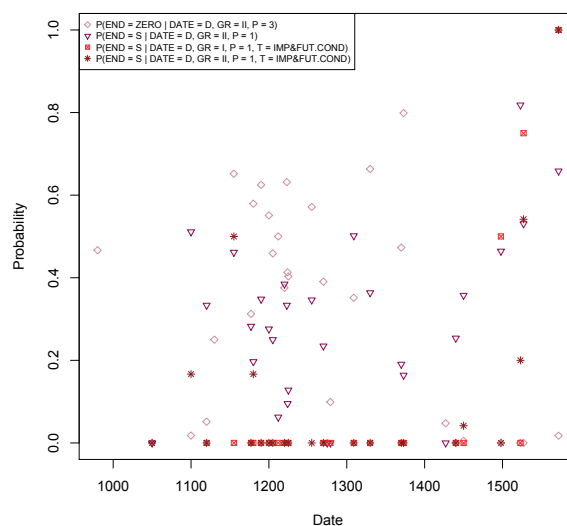
2.2.1 Quantifying the emergence of the new endings

In order to establish the temporal profile of the surface changes in verbal endings, we calculated the ratio of the “new” endings to the sum of the new and “old” endings for

¹³In preterite of Group II the fall of the final -t affected only verbs with non-root stress ending in -i or -u (e.g. Buridant 2000, 253-260).

each text in the corpus. In order to be able to identify the subject's person in an automated way, we limited ourselves to clauses with overt nominal or pronominal subjects. This means that we took a subset of all the cases of new ending appearing in the corpus. In order to determine whether considering only overt subjects skews the results, in section 4.3 we look at ending choice in a sample of clauses with null subjects manually annotated for subject person and conclude that there is no significant difference in the rate of new endings between null vs. overt subject contexts. This means that we can confidently estimate the rise of the new endings from a sample of clauses with overt subjects.

Fig. 2–3 show the rise of new endings divided into two major groups, namely, final -t deletion and e-insertion on the one hand, both of which resulted in an innovative -e ending, and s-insertion on the other.¹⁴ $P(\text{END} = \text{NEW} \mid \text{DATE} = D, \text{GR} = I, P = 1)$ stands for the estimated probability of a Group I verb to have a new ending (i.e. -e) in the context of the 1st person singular subject, etc.

Figure 2: Innovative -e and -a endingsFigure 3: Innovative zero and -s ending

Tables 20–28 in Appendix B give the observation numbers together with a proportion of the new endings in each text.

Comparing now Fig. 2–3 with Fig. 1, on the assumption that the spelling

¹⁴For now we leave aside those environments where final -t deletion did not lead to syncretism. We take them up in section 4.5.1.

innovations in question reflected changes in the verbal agreement phonology (in section 4.5.1 we present arguments supporting this assumption), there is no reason to assume that there was a temporal lag between the emergence of new syncretic endings and the rise of overt pronominal subjects. *However, we see that while the appearance of new -e and -a endings roughly parallels the emergence of overt subjects, innovative zero and -s follow a very different trend.* The next question is whether we can establish a non-accidental relation between the rise of new endings and overt subject.

3 Clause-level relation model

We will first explore quantitatively a classic line of analysis which relates null subjects and non-syncretic agreement via a certain structural property giving rise to both, let us call it Agr head. The two changes are thus viewed as a consequence of the loss of the grammar with Agr head. We show that an approach which maintains a clause-level relation between subject expression and the type of ending makes incorrect predictions about the rise of the new endings and overt pronominal subjects. We will then suggest a more flexible approach whereby syncretic endings, rather than being a direct manifestation of an alternative structure without Agr, are consequences of an independent phonological change which favours the alternative grammar. Thus, the second approach dissociates null subjects from a particular set of endings in terms of surface observations, but maintains that syncretisation eventually led to the disappearance of a grammar generating null subjects.

3.1 AgrP-Grammar

As part of the first model, we will assume that a null subject spells out pro, which, to be merged in the structure, needs to be identified by a verbal head specified with person features (e.g. Rizzi 1986; Adams 1987).¹⁵ We thus assume that the initial grammar,

¹⁵Since this model is meant to represent a family of approaches assuming a clause-level dependency between agreement type and subject expression, we are not committing ourselves to a particular representation. That is, instead of postulating an identifying relation between a pro and Agr, we could have adopted an approach which takes the verbal agreement head itself to have pronominal properties and

which allowed for null subjects, was characterized by the presence of a person feature-specified head Agr.¹⁶ We will call this configuration AgrP-Grammar. Let us also assume that person feature specification implies spellout rules for Agr which map every person feature to a unique exponent, to use the Distributed Morphology vocabulary. The endings -zero, -s, and -t are the outcomes of the following spellout rules of the initial grammar.¹⁷

- (4) a. $\emptyset \longleftrightarrow [_, 1P] / V + _$
 b. $s \longleftrightarrow [_, 2P] / V + _$
 c. $t \longleftrightarrow [_, 3P] / V + _$

The merge of Agr makes possible the merge of pro, as illustrated in Fig. 4.¹⁸

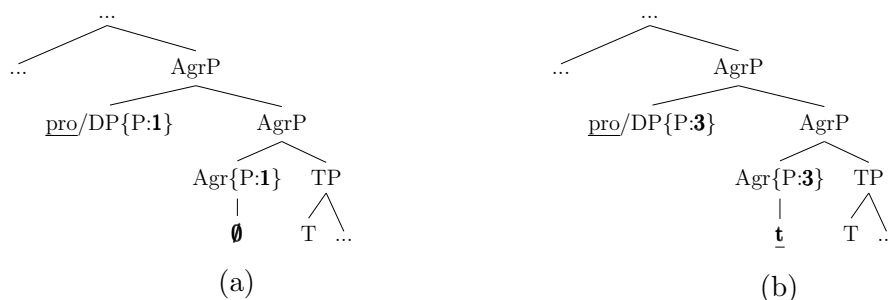


Figure 4: AgrP-Grammar spellout

Concerning the nature of the relation between Agr and pro, we will assume that person features introduce conditions on the meaning of a pronoun. A long semantic tradition ascribes to such features the status of presupposition triggers (e.g. Cooper 1983; Heim and Kratzer 1998; Kratzer 2009; Heim 2008; Sauerland 2008). Specifically, features have been claimed to introduce conditions on the properties of a potential referent of a pronoun, which, if not met in the discourse, make the utterance infelicitous. In addition to that, we will assume that the presence of such conditions is a prerequisite

thus serve to fill the verb's external argument slot (e.g. Poletto 2006).

¹⁶Importantly, we are not assuming that we can necessarily observe in existing sources the stage of French where the only grammar in use was the one we call here initial, that is, the null subject grammar without ending syncretism. This stage was likely left undocumented. This is why, we think, Old French is sometimes called a partial or no pro-drop language (e.g. Kaiser 2009; Zimmermann 2009). In terms of the grammar competition model of language change which we assume, the data which only partially conform to the criteria of a given grammatical type are modelled as a mix of outputs of two “pure” grammatical types.

¹⁷This is, of course, a coarse approximation, a more realistic spellout model would need to take into account information about tense, mood, conjugation type etc.

¹⁸For the sake of presentation clarity we abstract away from verbal movements through functional heads. We also stay non-committal as to whether pro should be governed by Agr.

for a felicitous use of a pronominal element. In other words, a pronoun needs to be accompanied by an element triggering a presupposition about its reference, whether it comes as part of the morphological form of the pronoun itself or as a verbal ending. That is, for a pronoun to be used, a person distinction has to be indicated.¹⁹ Taking the existence of the constraint for granted, we propose that person features on Agr introduce presuppositions about subject's reference. Specifically, in our model Agr introduces restrictions on the subject argument's domain of the function denoted by the TP. By assumption, an individual variable can only be interpreted if its domain has been narrowed down. In the absence of Agr therefore a pro will be left uninterpreted.²⁰

3.2 TP-Grammar

We model the replacement of null subjects with overt ones and of old endings with new ones as a passage from the initial AgrP-Grammar to an alternative grammar without Agr. Within the alternative grammar verbal endings correspond to the spellout of head T, unspecified for the person feature. The relevant spellout rule in this case is as follows.²¹ We will call this alternative grammar, sketched in Fig. 5, TP-Grammar.

$$(5) \quad e \longleftrightarrow [_, \text{present, singular}] / V + _$$

Since T does not carry person features, it does not introduce presuppositions necessary for a felicitous use of a pro. Therefore an overt DP, which itself carries the relevant information about subject's domain, has to be merged.

¹⁹In this relation we can invoke a long tradition, going back to Benveniste and supported by typological observations, of ascribing a universal status to the person distinctions in natural languages (e.g. Harley and Ritter 2002), Tvica 2017.

²⁰Null subjects have also been conceived of as minimal pronouns, in the sense of Kratzer (2009), which need to acquire ϕ -features from the verbal head which binds them (e.g. Alonso-Ovalle and D'Introno 2000).

²¹We assume that -ez and -ons endings, distinct from the rest and each other, are exponents of the feature [plural] in the context of 1st and 2nd person plural subjects.

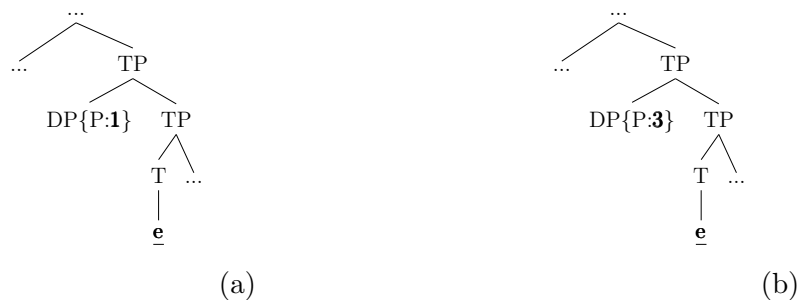


Figure 5: TP-Grammar spellout

Given our assumptions about AgrP-Grammar and TP-Grammar, if the latter replaces the former, null subjects will become unavailable. Assuming the Constant Rate Hypothesis of Kroch (1989), our model predicts that the rate of replacement of AgrP-Grammar by TP-Grammar should be the same whether it is measured as the rise of overt pronominal subjects or of new syncretic endings. For the general case, the Constant Rate Hypothesis (CRH) states that a grammatical change has the same rate of spreading in all grammatical environments, where the rate is taken to correspond to the slope coefficient of a logistic regression model. Over the last three decades CRH has acquired substantial empirical support in diachronic research on a number of languages (see Pintzuk (2004) for an overview). Kauhanen and Walkden (2017) provide theoretical foundations for the CRH and also discuss some problems with the standard way of testing the CRH. In particular, following up on the discussion in Paolillo (2011), they point out that the standard way of assessing statistical significance (Kroch 1989; Santorini 1993; Pintzuk 1995) of a putative Constant Rate effect is statistically unsound: “if the result is not statistically significant, then it is concluded that there is support for a [Constant Rate Effect]. However, it is not sound to treat a non-significant value as evidence for the null hypothesis, since it was assumed to begin with.” This problem also applies to our own methodology as described in section 4.1. We will maintain therefore that, whenever the result of an independence test on regression coefficients is non-significant, it does not contradict the CRH, rather than provides direct evidence for it.

Thus, we expect the rates of the emergence of overt pronominal subjects and of the new endings to be not significantly different. We will evaluate the two rates by

comparing the slope parameters of the corresponding logistic regression models.

One caveat of the prediction we have just made is that even stable null subject grammars allow for overt subjects, usually under some pragmatically defined conditions (e.g. Bates 1976; Otheguy et al. 2007). This makes it impossible to classify a given overt pronominal subject as an instance of AgrP-Grammar or TP-Grammar, since both of them are expected to generate overt pronominal subjects. The only context which sets the two apart clearly are expletive subjects, which are consistently null in null subject languages (e.g. Jaeggli and Safir 1989).²² We therefore will compare the rise of overt expletive subjects with the rise of the new endings. The relevant variation in subject expression is illustrated in (6)–(7).

- (6) Et \emptyset semble que li darrien soient Ausi com li premier estoient.
 and (it) seems that the last are also like the first were
 ‘And it seems that the last would be like the first (ones) were.’ (1190-BORON-R,13.185)
- (7) meis **il** nous semble Qu’ il ne pourroit pas estre vis
 but it us seems that he not could not be alive
 ‘But it seems to us that he could not be alive.’ (1190-BORON-R,69.1084)

There are at least three other immediate predictions. First, this model predicts that the rise of the new endings should proceed at the same rate in different contexts: if the emergence of the new endings reflects the disappearance of Agr, on the CRH we do not expect this change to proceed differently depending on the verb type or the subject person.

Another prediction is that there should be no increase in the frequency of null subjects in the contexts of new syncretic endings. This is so because the AgrP-Grammar which, by hypothesis, is the only grammar that can license null subjects, is associated with spellout rules which do not output syncretic endings, such as -e in the context of the 1st and 3rd person subjects, overt or null.

Finally, there is a prediction that there should be no increase in subject expression with old, non-syncretic endings: although AgrP-Grammar, associated with non-syncretic endings, does sometimes generate overt subjects, their distribution is

²²As mentioned above, the stage where only AgrP-Grammar is operative is not attested; we find overt expletives in the earliest French documents, even though at a very low rate.

governed by pragmatic rules which produce the same rate of subject expression during the course of existence of grammar AgrP.²³

4 Performance of the clause-level relation model

4.1 Testing the main hypothesis

In order to evaluate the hypothesis that the emergence of overt expletive subjects and syncretic verbal endings are two manifestations of the disappearance of the grammar with a person feature-specified Agr head, we fitted the data on the appearance of overt expletives and the new endings to logistic regression models plotted in Fig. 6. To test the hypothesis we compare two logistic models. First the model ENDING is a model that predicts whether the verbal ending, Y , is *new* (or syncretic) by contrast with an *old* or non syncretic verbal ending as a function of time. This model has the following form:

$$(8) \quad P(Y = \text{new} \mid \text{Date} = d) = \frac{e^{\alpha + \beta t}}{1 + e^{\alpha + \beta t}}$$

where α is the intercept and β the slope. The intercept is interpreted as an abstract indicator of when the change takes place in time and the slope is interpreted as the rate of change. The steeper the slope, the faster the change progresses.

We compare this model with an EXPLETIVE SUBJECT model which predicts whether the expletive subject realisation, Y , is *new* (or overt) by contrast with an *old* realisation where the pronominal subject is null. The second model has the same form:

$$(9) \quad P(Y = \text{new} \mid \text{Date} = d) = \frac{e^{\alpha + \beta t}}{1 + e^{\alpha + \beta t}}$$

but this time Y represents expletive subject realisation instead of verbal syncretism. In order to illustrate both models we first fitted them separately to the data. For the sake of comparison, we also plotted the data on the overt personal pronominal subjects. By assumptions, both competitors, Agr- and TP-Grammar, can generate overt pronominal subjects, the former having an upper threshold for their frequency determined by pragmatics. The logistic curves are provided in Fig. 6 and parameter estimates are as

²³Empirically, it appears that the rate of such pragmatically conditioned subject expression in pro-drop languages is around 30–40%, according to Bates (1976), Otheguy et al. (2007), Erker and Guy (2012).

follows:

MODEL	COEFFICIENT	STD. ERROR	Z VALUE	$\Pr(> z)$	ACCURACY
ENDING	0.0062	0.0001	42.75	$< 2 \times 10^{-16}$	0.76
EXPLETIVE SUBJECT	0.0055	0.0002	25.34	$< 2 \times 10^{-16}$	0.61
PERSONAL SUBJECT	0.0063	0.0001	60.87	$< 2 \times 10^{-16}$	0.81

Table 7: Logistic regression estimates for the new endings and overt pronominal subjects

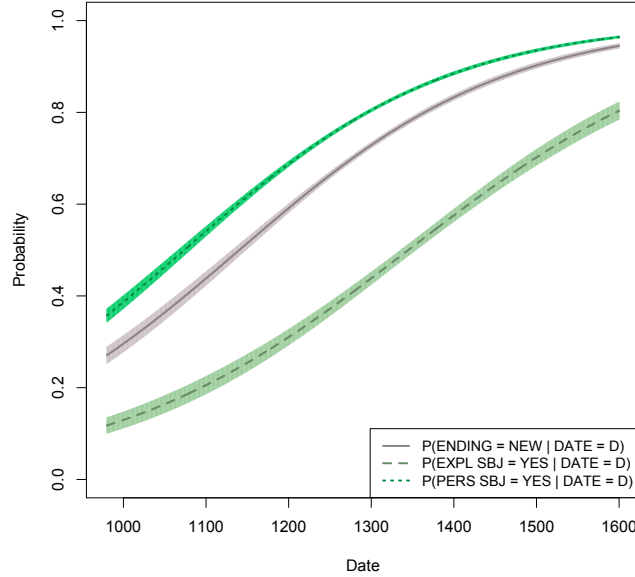


Figure 6: Spread of new endings and overt pronominal subjects

Numbers of observations of null and overt expletive and personal pronominal subjects are given in Table 19 in Appendix B. The coefficients are not very different from each other, but not identical either. We therefore further explicitly test the CRH. To do that we test for the contribution of the slope by comparing two mixed effect models. The first model amounts to predicting the new form Y , whether it is an overt expletive subject or a syncretic verbal ending, by contrast with an old form, that is a null subject or a non syncretic verbal ending. The prediction is still a function of time but we also add a random intercept α_c for each context c : either a morphological context or a subject context. This yields the following model:

$$(10) \quad P(Y = \text{new} \mid \text{Time} = t, \text{Context} = c) = \frac{e^{\alpha + \alpha_c + \beta t}}{1 + e^{\alpha + \alpha_c + \beta t}}$$

Informally this model means that the global model intercept may be further parametrized for each specific context but the slope is constrained to be identical for both contexts. We compare this model to an extended version, where this time we add a

random slope β_c , thus allowing the slope to vary for each context:

$$(11) \quad P(Y = \text{new} \mid \text{Time} = t, \text{Context} = c) = \frac{e^{\alpha + \alpha_c + (\beta + \beta_c)t}}{1 + e^{\alpha + \alpha_c + (\beta + \beta_c)t}}$$

Since the slope models the rate of change, this second model allows the rate of change to differ for each context. We test whether the slope introduces a significant difference between the two models (10) and (11) with a log likelihood ratio test which is χ^2 distributed ($df = 2$). The test has $p = 0.04$, and so we can reject the null hypothesis at the 0.05 confidence level. We must conclude that the introduction of the slope does better predict the data, and thus, on the CRH, these results are not compatible with the analysis of the two diachronic phenomena as stemming from the same grammatical change, which we identified as a passage from a grammar with Agr head to a grammar without. In the remainder of this section we will explore three other predictions made by the clause-level relation model and show that none of them is borne out.

4.2 Syncretisation in different contexts

The model for agreement syncretisation defined in (8) merges nine different syncretisation patterns described in section 2.2 and summarised in (3). If syncretisation is a consequence of the TP-Grammar associated with the new spellout rules winning over the old AgrP-Grammar, then these developments are expected to have the same rate. In order to test this, we modelled them separately, as illustrated in Fig. 7.

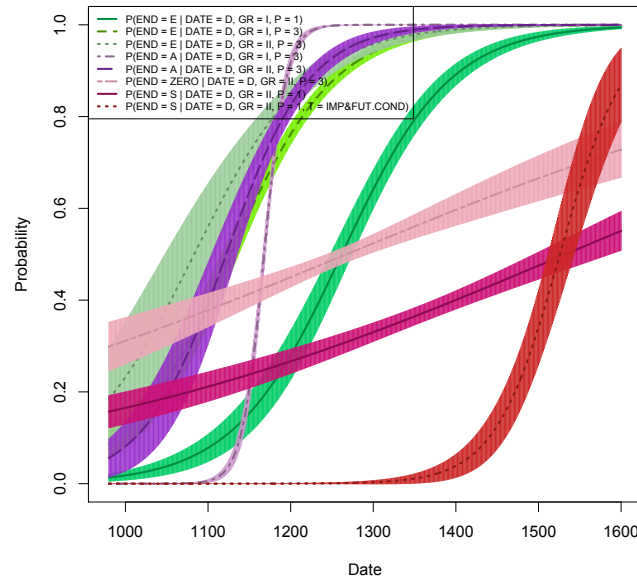


Figure 7: Logistic regression models of the emergence of the new endings

These models have the following estimates.²⁴

MODEL	COEFFICIENT	STD. ERROR	Z VALUE	Pr(> z)	ACCURACY
E, GROUP I, 1ST	0.0153	0.0009	16.77	$< 2 \times 10^{-16}$	0.9
E, GROUP I, 3RD	0.0160	0.0008	18.81	$< 2 \times 10^{-16}$	0.95
E, GROUP II, 3RD	0.0144	0.0017	8.32	$< 2 \times 10^{-16}$	0.9
A, GROUP I, 3RD	0.0752	0.0037	20.10	$< 2 \times 10^{-16}$	1
A, GROUP II, 3RD	0.0199	0.0019	10.46	$< 2 \times 10^{-16}$	0.93
ZERO, GROUP II	0.0030	0.0004	6.63	$< 2 \times 10^{-16}$	0.5
S, GROUP II, 1ST	0.0030	0.0003	9.05	$< 2 \times 10^{-16}$	0.65
S, GROUP I, 1ST, IMPERF. & FUT. COND.	undefined				
S, GROUP II, 1ST, IMPERF. & FUT. COND.	0.0255	0.0030	8.42	$< 2 \times 10^{-16}$	0.95

Table 8: Regression estimates for the emergence of the new endings

The observation numbers underlying these models are given in Tables 20–28 in Appendix B. Upon visual inspection, we see that the spread of the new ending -e has more or less the same profile in all of its contexts. In contrast, it differs from the spread of -a and -s, contrary to what was predicted by the clause-level relation model. Thus, individual endings spread at different rates, and the innovations seem to group into classes in terms of their phonological environments.²⁵

²⁴For the alternation between zero and -s with Group I 1st person imperfect and future conditional, we observe that the date is a perfect predictor for endings. In such situations, one cannot reliably estimate the parameters of a logistic model.

²⁵One way to interpret these findings is to assume that first there is a fall of the ending-final stops after -e- and a compensatory /e/-insertion following root-final stops, which appears as an innovative -e

4.3 Spread of the new endings with null subjects

Another prediction made by the clause-level relation model is that there should be no increase in the new endings in the context of null subjects. First, we do find occurrences of -e in the context of the 1st or 3rd person singular null subjects. Table 29 in Appendix B shows observation numbers for finite clauses with Group I verbs with zero or -e ending with 1st person null subjects (manually annotated). The second observation is that such occurrences of new endings with null subjects are not frequent: at all times they stay below 20 per text. One way to explain away their occurrence is to analyse them as etymological vowels which create noise in the passage from the old to new endings. However, if that is indeed noise, we expect it not to become stronger with time. To test this expectation we fit the data on the appearance of -e in the context of the 1st person singular overt and null subjects to a logistic regression. As Fig. 8 shows, the trend is the same.

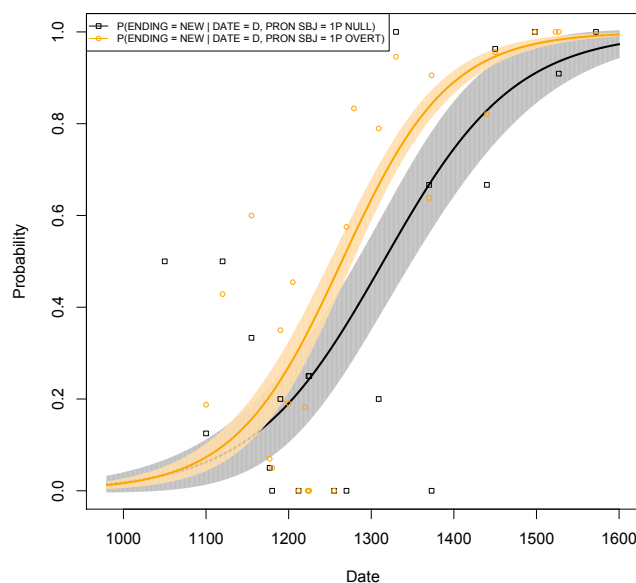


Figure 8: Rise of -e with Group I verbs in the context of the 1st person singular subjects

This result is unexpected if -e with null subjects is just an etymological residue.

ending. Second, there is a fall of the ending-final stops after -a-, which appears as an innovative -a ending. Third, there is a fall of the ending-final stops after -i/-u (Group II 3rd person preterite) and /s/-insertion after -i/-u (Group II 1st person present and preterite), which results in innovative zero and -s endings respectively. The latter process can arguably be considered as compensatory in relation to the former in order to keep the 1/3 person distinction. Finally, there is an innovative /s/-insertion after diphthongs (Group I and II first person imperfect and future conditional).

MODEL	ESTIMATE	STD. ERROR	Z VALUE	PR(> z)
NULL 1SG SUBJECT	0.0146	0.0018	8.05	$< 2 \times 10^{-16}$
OVERT 1SG SUBJECT	0.0153	0.0009	16.77	$< 2 \times 10^{-16}$

Table 9: The emergence of the new endings with null and overt subjects

Rather, the observation that the new ending spreads at similar rates in the context of null and overt subjects suggests that we are witnessing one and the same (phonological) change operating in different contexts. In other words, the choice of ending is independent of the expression of the subject, contrary to what is predicted by the structural model relating subject expression and ending type as manifestations of a particular grammar. Note that we do not need to check the spread of different types of new endings with overt and null subjects since the clause-level relation model predicts that no new endings increase with null subjects and is therefore falsified even by one case of the contrary.

4.4 Spread of overt subjects with old endings

The final prediction that we derive from the clause-level relation model is that there should be no increase in subject expression in the context of verbs with old, non-syncretic endings, because this is the context of the AgrP-Grammar, which is supposed to express subjects at a pragmatically determined stable rate. We compared the rate of subject expression in the contexts of verbs with the old non-syncretic endings -t and zero on the one hand and new syncretic endings -e, -s on the other. We estimated the probability of having an overt pronominal subject for finite clauses with verbs ending in -e (Group I & II), -t (Group I & II), -s (Group II), and zero (Group I & II) endings. We fit the seven datasets to the following logistic regression model:

$$P(\text{PRON SBJ} = \underline{yes} | \text{DATE} = \underline{d}) = \frac{e^{\alpha + \beta * d}}{1 + e^{\alpha + \beta * d}}.$$

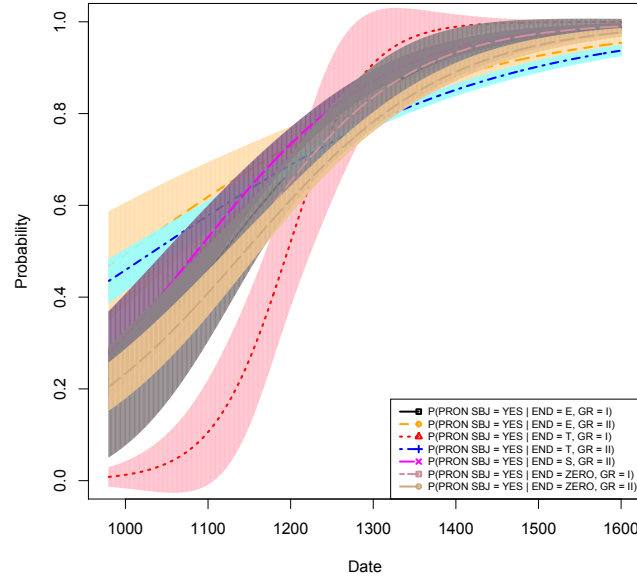


Figure 9: Pronominal subject expression with old and new endings

MODEL	COEFFICIENT	STD. ERROR	Z VALUE	PR(> z)	AMBIGUITY
e, Group I	0.0091	0.0006	14.39	$< 2 \times 10^{-16}$	yes
e, Group II	0.0051	0.0008	6.69	2.17×10^{-11}	yes
t, Group I	0.0222	0.0059	3.75	0.0002	no
t, Group II	0.0048	0.0003	16.02	$< 2 \times 10^{-16}$	no
s, Group II	0.0089	0.0008	10.55	$< 2 \times 10^{-16}$	yes
zero, Group I	0.0100	0.0020	5.08	3.77×10^{-7}	no
zero, Group II	0.0082	0.0006	14.72	$< 2 \times 10^{-16}$	no

Table 10: Logistic regression estimates for ambiguous and non-ambiguous endings

Clearly, subject expression rate grows over time for the non-ambiguous endings, a result not predicted by an approach which sees them as strictly part of the null subject licensing grammar which has a constant pragmatically determined rate of subject expression.²⁶

Relatedly, Ranson (2009) concludes, based on the three texts she examines, that ending ambiguity is not a good predictor of subject expression. Our findings support this conclusion. If we compare ambiguous -e (both groups combined) with unambiguous -zero (both groups combined as well) for any given text, there is no significant effect of ending choice on subject expression. For instance, for La Chanson de Roland, the model

²⁶One can see that overt pronominal subjects spread at very similar rates with all endings except for -t with Group I verbs. A plausible explanation is that -t with Group I verbs virtually disappears after 1200 and what we see past that date is essentially noise which skews the model.

$P(\text{PRON SBJ} = \underline{yes} | \text{ENDING} = \underline{x}) = \frac{e^{\alpha + \beta * d}}{1 + e^{\alpha + \beta * d}}$ where x ranges over ambiguous $\underline{-e}$ and unambiguous zero has the following estimates.

MODEL	ESTIMATE	STD. ERROR	Z VALUE	Pr(> z)
Ending (zero)	-0.154	1.2864	-0.12	0.9

Table 11: Ending ambiguity as a predictor of subject expression

4.5 Discussion

We have shown that a number of predictions generated by a model which assumes that subject expression and agreement type are related at the clause level, via a certain functional head, are not borne out. Namely, new endings spread at different rates depending on the ending type, which is unexpected if both are generated by a new grammar supposed to spread at the same rate on the CRH. In addition, new endings spread both with overt and null subjects, contrary to the model's assumption that null subjects are generated only by the old AgrP-Grammar where the Agr head spells out as old, non-syncretic endings. Finally, the expectation that there would be no increase in overt subjects in the context of old, non-syncretic endings, which, by hypothesis, are generated by the AgrP-Grammar producing overt subjects at a constant (relatively low) rate, is also not borne out. The overall conclusion is that a model which assumes a strict dependency at the clause level between what type of endings is used and whether or not pronominal subject is expressed is not supported by the diachronic data.

However, given that overall the new endings and overt pronominal subjects (whether personal or expletive) spread at almost the same rates, illustrated in Fig. 6, it would likewise be counterintuitive to conclude that we should give up altogether on all the models which assume a non-accidental relation between the two changes. This sort of tension is not astonishing given that despite the seeming cross-linguistic correlation between non-syncretic agreement and pro-drop, it has proven difficult to give a formal characterisation of the notion of the (morphological) agreement non-syncretism in such a way that it would effectively predict whether a language allows for null subjects (e.g. the discussions in Vance (1997, 214–225), Alexiadou and Anagnostopoulou (1998,

522–523), Camacho (2013, 111)).

What is needed is therefore a model which would dissociate subject expression and the ending shape at the clause level, but would relate them closely in the process of language evolution. In what follows we suggest a model which relates null subjects and verbal subject agreement via the process of grammar competition, building on Yang (2002, 2010). Within this model, syncretic endings, whose appearance is presumably governed by an independent phonological mechanism, create a negative bias for a null subject-licensing grammar, which ends up driving the latter out of use in speakers' community. This model makes no assumptions about other reasons for which null subjects can be absent from a language, and thus does not a priori rule out non-syncretic agreement languages with obligatory overt subjects, as for instance Icelandic is claimed to be.

Before proceeding to elaborate such a model, we need to deal with another possible explanation for why we do not find a complete parallelism between ending syncretisation and pro-drop disappearance, namely, that the verb ending changes registered in the written texts are not reflective of the phonological reality and therefore cannot be used to evaluate a clause-level relation hypothesis.

4.5.1 Spelling–pronunciation problem

A writing system being a conventional rendition of the phonological reality, there is no priory reason to assume that the convention is at all times a strict correspondence between pronunciation and spelling. For the purposes of the present study the problem entails two independent questions. The first one is whether the spelling innovations, such as the disappearance of final -t and appearance of new -e and -s, had phonological substance, that is, whether they reflected phonological innovations. Only answering this question positively makes our tests for the underlying structural relatedness between the observed changes in the paradigm and the change in subject expression meaningful. Below we provide a number of arguments in support of the affirmative answer. The second question about the spelling-pronunciation relation is concerned with the

emergence of phonological innovations behind conservative orthography. The state of Modern French witnesses to the fall of all the stops and sibilants (at least in an isolated pronunciation) that used to correspond to the present-day word-final consonantal graphemes, not just the final -t whose disappearance we tracked above. Again, judging from the Modern French spelling-pronunciation correspondence, this change is mostly not reflected in spelling. For the second part of our study, where we attempt to estimate the general level of syncretism in the system, it is important to know until what point in time we can equate presence in the spelling with phonological presence. Fortunately, it seems that we can estimate this date with a great deal of precision due to the co-occurrence of two independently attested facts. First, there exists a historical record of the first centralised spelling standardisation in the mid-XIV c. Second, our data show that the disappearance of the final -t with Group II verbs with unstressed roots ending in -i/-u, which, if it had followed a statistically expected trajectory, would have reached its completion around that time, was stopped and reversed in the late XIV c., presumably showing the effect of the spelling standardisation which marked the end of the strict spelling-pronunciation correspondence. The abrupt reversal of the change is illustrated on Fig. 10.

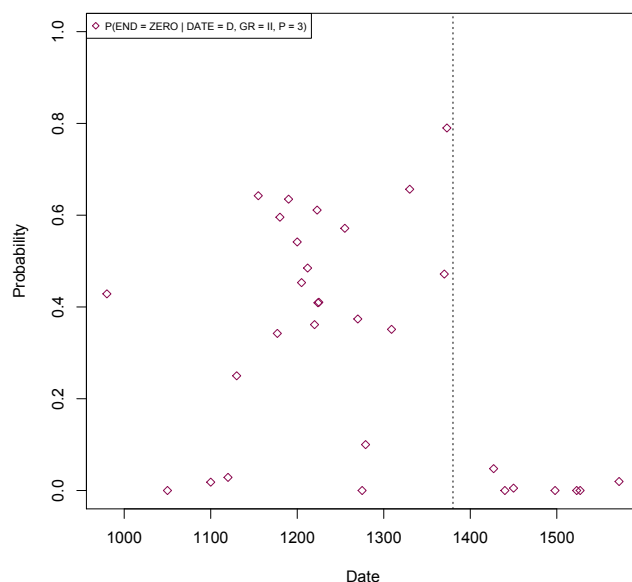


Figure 10: Change reversal for Group II verbs in preterite with 3rd person subject

There seems to be a consensus that the rise of the new endings reflected the phonological reality rather than simply a change in orthographical conventions (Dees

et al. 1980; Fouché 1931; Goyette 1993; Marchello-Nizia 1992; Morin 2001; van Reenen and Schøsler 1987). The only disputed point is the driving force behind these phonological innovations, namely, whether it was some sort of analogy induced by other forms in the paradigm or by the verbs whose stems etymologically ended in -e or -s (see Goyette (1993, 58–76) and Morin (2001) for -e and -s respectively) or else whether, at least in the case of the final -e, it was epenthesis which kept the preceding consonant from the final position at the time when final consonants were becoming unstable (Dees et al. 1980; van Reenen and Schøsler 1987). Arguments against a possible claim that what we observe in texts is just variation in writing conventions can be divided into the following groups.

First, as we saw in section 4.2, the emergence of *-e* as a final grapheme in the context of Group I verbs with a 3rd person singular subject in present indicative and subjunctive follows a logistic curve whose slope is indistinguishable from the slope of the curve modelling the emergence of *-e* with Group I verbs in the context of 1st person singular subjects in present indicative and subjunctive.²⁷ These results fit well with the hypothesis of Dees et al. 1980 and van Reenen and Schøsler 1987 about /e/-insertion being a compensatory process triggered by the instability of the final stops to preserve the integrity of the root. The appearance of *-e* as a final grapheme with the 3rd person singular subjects on this view results from the fall of the final /t/ (e.g. *aimet* > *aime* ‘(he) loves’ and *aint* > *aime* ‘(he) would love’), whereas its appearance with the 1st person singular subjects results from a compensatory /e/-insertion to keep the root final consonants from not being pronounced (e.g. *aim* > *aime* ‘(I) love’ and ‘(I) would love’). Although the quasi-identity of slopes is only indicative, this is expected on the hypothesis that this is a paradigm-wide morphophonological process. That is, given the CRH, it is entirely expected for a morphophonologically conditioned change to proceed at the same rate in different environments (cf. Fruehwald et al. 2009).

Second, according to our estimates, in the context of the 3rd person subjects syncretisation happened earlier than with the 1st person singular subjects, which makes

²⁷We used the same approach as for the comparison between the slopes of expletive subject and new ending emergence models, with $p = 0.8$ this time.

sense if the fall of the final stops which were not part of the root (again, *aimet* > *aime* ‘(he) loves’ and *aint* > *aime* ‘(he) would love’) preceded the emergence of a “compensatory” /e/ following root final consonants. In contrast, on the hypothesis that what we observe are changes in spelling conventions, although not theoretically impossible, it would look like a series of strange coincidences if, first, different spelling conventions in different contexts were changing at very same rates, and, second, if they first changed in the context of the 1st and then of the 3rd person subjects.

Third, according to Fouché (1931, 180) and Marchello-Nizia (1992, 201), in the context of the 1st person singular subjects in indicative and subjunctive, the *-e* grapheme first spread in the context of consonant-final and only later vowel-final roots (e.g. *cri-er* “to shout” where in the context of the 1st person singular subjects *cri* was replaced by *crie*). Again, this fits a phonology-based account of the change since the sequence of spreading across contexts can be described in terms of phonologically natural classes, whereas this appears as a mysterious plotting of the scribes on the spelling convention-based account.

Lastly, a phonologically motivated change affecting final vowels has precedence in the history of Late Latin, where all the final vowels ended up falling except for those cases where their fall would have led to an unacceptable consonant cluster (see the discussion in Goyette (1993) and references therein). In Old French reflexes of this process are the so called “etymological *e*”, that is, root-final *-e* following certain consonant clusters, as in *siffle-r* (from Latin *sibila-re* > *sifila-re* > *sifla-re*) (cf. *don-er* from Latin *dona-re* which lost its root final /a/ in Late Latin, unlike *siflare*). It is not so surprising then to see another round of “compensatory” root final /e/, this time as an epenthetic process meant to keep the root final consonants from falling.

In view of these arguments, none of which supports an account of the new endings in terms of spelling convention changes, we conclude that our results based on written source can be plausibly projected onto the phonological reality and thus used to test a model structurally relating syncretism introducing changes and pro-drop

disappearance.²⁸ Similarly, Fruehwald et al. (2009) analyse data on the loss of final fortition in (Bavarian) Early New High German, observable in orthographic variation of the period, e.g. tak vs. tag “day (acc. sg)”, rat vs. rad “counsel (acc.sg)”, and argue that this variation clearly represents a phonological change in progress rather than shifting scribal tradition.

When it comes to determining at what point the phonological reality behind conservative spellings changed, the reconstructions of the timing of the fall of the final consonants rely mostly on the analysis of rhymes (matching vs. non-matching), hypercorrections (insertions of etymologically absent consonants), omissions of etymologically present consonants, commentaries in the grammars of the time, and analyses of the borrowing from French into other languages which likely reflected the spelling of the time of borrowing. The dating question is important to us in as much as we want to take into account final consonant instability when evaluating the overall degree of syncretism or ambiguity in the verbal system. This means that we are mostly concerned with the pronunciation of the word final -s and -t, which will become especially relevant in section 5 where we evaluate the general syncretism of the verbal system.

De Jong (2006) undertakes a statistical analysis of the rhymes in three texts written in the Parisian dialect in the XIII–XIV cc.²⁹ She looks at the frequency of the non-matching rhymes for a given grapheme (e.g. escript “text” – (je) pris “I take”) compared with that of the matching rhymes (e.g. moult “many” – (je) doubt “I doubt”), taking higher-than-chance frequencies to be indicative of the grapheme non-pronunciation.

One of the general conclusions of De Jong (2006, 176) is that the non-pronunciation of the final consonantal graphemes increases dramatically in the XIV c. Although the numbers of the mismatching rhymes do not always warrant a solid statistical analysis,

²⁸We would like to remind the reader that, to the best of our knowledge, no one has advanced an account in terms of spelling convention changes.

²⁹This study of the rhyme system is based on three texts known to be written for the Parisian public. These are Oeuvres Complètes de Rutebeuf (XIII c.) La Chronique métrique attribuée à Geoffroy de Paris (first third of the XIV c.), and Les Miracles de Notre Dame de Paris (from 1340 to 1380).

this is the period when the mismatching rhymes, including mismatches involving our consonants of interest, begin to be observed in her corpus. Very similar observations were made earlier by Foulet (1935).³⁰

Importantly, De Jong (2006, 174) links the emerging mismatch between spelling and pronunciation with a particular historical event, namely, the introduction by the Royal Chancellery in Paris of the standard exams for the scribes in 1342. We found a rather dramatic argument in favour of this hypothesis in form of the reversal of the final *-t* disappearance in the preterite forms of certain Group II verbs (Fig. 10). We cannot conceive of any plausible explanation of this development in phonological terms. Rather, it seems to result precisely from an artificially introduced norm.

5 Change as a variational learning outcome

The disappearance of null subjects in the model we explored in the previous section was not related to the disappearance of non-syncretic agreement by causality, but rather by sameness of origin. In this model, non-syncretic endings and null subjects were considered manifestations of the same grammatical property, namely, a functional head responsible for subject agreement. Since this model does not make any statements about causality, it cannot be used to cash out the long-standing intuition going back to at least Foulet (1928) that it was the impoverishment of the verbal endings that triggered the loss of null subjects. As was demonstrated above, this model also makes incorrect empirical predictions.

Sprouse and Vance (1999) identified the cause of the disappearance of null subjects with the processing difficulty associated with their parsing. Namely, they proposed the first, to our knowledge, reinforcement learning model to explain the loss of null subjects.³¹ In this model, a null subject has a greater chance of inducing a parsing

³⁰Foulet (1935), however, also assumed, as a matter of pure intuition, that even before non-pronunciation started to be observable as consonant omissions and hypercorrections, scribes had not been pronouncing the final consonants for at least two centuries. De Jong (2006) and Bettens (2015) see no reason for making this assumption.

³¹Sprouse and Vance (1999) call it “Differential Parsing Model of Change”, which is in turn based on the model proposed in Klima (1965).

failure than its competitor, an overt pronominal subject. Since, by the authors' assumption, speakers tend to produce grammatical forms at frequencies at which they have encountered them in their speech community, failures to parse null subjects will lead to the decrease in the frequency of null subjects in the output of the speakers, which in turn will reduce the ambient frequency of null subjects on the next cycle. The cycle repeats until null subjects vanish from the speech community. In addition to the inherent parsing disadvantage of null forms in comparison with overt pronouns, Sprouse and Vance (1999, 264) make allowance for a secondary factor in the form of ambiguous verbal endings. Specifically, they predict a different (presumably higher) rate of overt subjects in the context of verbal forms with ambiguous endings.³² This model, however, did not receive an implementation which would make concrete predictions concerning the diachronic profile of the competition.

Below we suggest a model of the loss of null subjects which builds on the variational learning model proposed in Yang (2002, 2010). Based on the quantitative data from MCVF (2010) and Penn Supplement to MCVF (2010), we will assign concrete values to our model's parameters which will enable it to generate predictions concerning the course of the change. Contra Sprouse and Vance (1999), ambiguous endings are considered within this model as the main factor that creates a parsing difficulty for the null subject grammar. We will first present the general layout of Yang's model and then show how it can be adapted to our problematic.

5.1 General framework

Yang's (2002, 2010) model is based on the assumption that children have innate access to multiple grammatical systems and, in the course of language learning, use the input data to probabilistically evaluate the available options. They may either converge on a single grammar, or, as adults, they may end up with multiple grammars used at certain

³²Sprouse and Vance (1999, 262) assume that a prerequisite for a competition situation is the existence in a language of overt pronominal forms with referential properties identical to those of null pronouns. In case a language has only stressed overt pronouns which are not interchangeable with null variants, as they assume to be the case in Italian, it will manifest a diachronically stable rate of null subjects. On this view, what sets the system in motion is presumably the emergence of overt atonic pronominal forms.

probabilities, which corresponds to the case of synchronic variation. Depending on whether the next generation arrives at the same or different probability distribution, we get the case of diachronically stable variation or diachronic change respectively. Hypothesising what kind of data contributes to the probabilistic evaluation of the grammars, we can approximate the course of the competition based on corpus distributions of the relevant data.

Formally, we use Yang's (2002, 2010) model as a way to estimate the probabilities $P(\mathcal{G} = G_1)$ of using the grammar G_1 and $P(\mathcal{G} = G_2)$ of using the grammar G_2 from a data set $X = x_1 \dots x_n$ in which, for a specific example $x \in X$, we are not sure which of G_1 or G_2 actually generated x .

Informally, the estimation procedure is iterative and increases $P(\mathcal{G} = G_i)$ when G_i successfully parses an example x while it decreases $P(\mathcal{G} = G_j)$ ($i \neq j$). The iterative procedure runs as follows:

- Select randomly a clause x in the data set X
- Select randomly G_i in proportion to its probability
- Analyze x with G_i
 - If G_i succeeds in analyzing x , provide G_i a reward and G_j a penalty:
 $P(\mathcal{G} = G_i)$ increases and $P(\mathcal{G} = G_j)$ decreases.
 - If G_i fails in analyzing x , provide G_i a penalty and G_j a reward:
 $P(\mathcal{G} = G_i)$ decreases and $P(\mathcal{G} = G_j)$ increases.

Using the Linear Reward Penalty scheme of Bush and Mosteller (1958) to update the probabilities, this process could be run iteratively to get the actual estimation of the probabilities. Instead, Narendra and Thathachar (1989) prove a theorem on the asymptotic behaviour of this process that provides a simpler analytical estimation method.

Using the notation $G_i \not\vdash x$ to indicate that G_i fails to parse x , we can define the notion of penalty of a grammar G_i as $c_i = P(G_i \not\vdash x)$. That is, c_i is the probability that G_i fails to analyze an example in X . This quantity can be estimated simply by counting the proportion of grammar's failures in the data set. Given this notion, for the case

where we have two grammars G_1, G_2 with penalties c_1, c_2 , Narendra and Thathachar (1989) prove the following theorem:

$$(12) \quad \lim_{\tau \rightarrow \infty} P(\mathcal{G} = G_1 | T = \tau) = \frac{c_2}{c_1 + c_2}$$

$$(13) \quad \lim_{\tau \rightarrow \infty} P(\mathcal{G} = G_2 | T = \tau) = \frac{c_1}{c_1 + c_2}$$

In other words, the probability of using a grammar G_i is proportional to the number of observed failures of G_j in the data set ($i \neq j$). Specifically $P(\mathcal{G} = G_i) = 1$ when G_j always fails and $P(\mathcal{G} = G_i) = 0$ when G_j never fails.

5.2 Diachronic stability and change

The outcome of the learning process (possibly over the lifespan) may stay the same or it may change from one generation to another. In the model we are considering, the only reason why learning may not converge on grammar G_i is if its penalty probability c_i is greater than zero, that is, if there are some subset input data that G_i fails to parse.

Once c_i associated with G_i becomes greater than zero, a language may leave a diachronically stable state and enter a state of diachronic change. Moreover, an increase in the frequency of the data unparseable with G_i in the next generation will lead to the increase in c_i , and so on to the point when G_i gets completely demoted. Emergence of such data may have nothing to do with the grammatical options themselves and may stem from phonological changes as well as from a second language interference.

Applying this to the loss of null subjects in Medieval French, let us assume that the initial winning grammar is the one that licenses null pronominal subjects. We associate such a grammar with the presence of the Agr head which carries person features, as in Fig. 4. This structure licenses subject pro and therefore generates subject-less clauses as well as clauses with an overt subject (again, in pragmatically restricted environments). Its competitor, the TP-Grammar, where T head does not carry person feature specifications, does not license pro, and therefore only generates clauses with an overt subject. Notice that this model incorporates the Taraldsen/Rizzi's insight about a categorical, core grammar-based dependency between functional head features and null subjects; it only gives up on the assumption that the relevant person features

transparently translate to the spellout.

In order to model the competition between these two grammars, the crucial parameters are the penalty probabilities of the grammars. In order to determine these, we need to establish which types of data AgrP- and TP-Grammars fail to parse. By hypothesis, AgrP-Grammar fails each time the information about subject's reference cannot be retrieved from the verbal ending, which is the case whenever the ending is ambiguous. Such data are illustrated below with already familiar syncretic endings -e and -s.

- | | | |
|--|------|---|
| (14) a. $S_{P1}V\text{-}\underline{e}$ GROUP I VERBS
b. $S_{P3}V\text{-}\underline{e}$ | (15) | a. $S_{P1}V\text{-}\underline{s}$ GROUP II VERBS
b. $S_{P2}V\text{-}\underline{s}$ |
|--|------|---|

The distribution of syncretic endings is such that it cannot be captured by the rules making reference to subject's person, from which a person exposed to the data presumably infers that they do not carry person features. In other words, we assume that the speakers categorically classify endings as ambiguous or not based on the distributions in the data they have been exposed to. An ending is classified as ambiguous in case the speaker has been exposed to a data sample where the ending occurs in the context of overt subjects with various (more than one) person specifications. The distribution of the old endings is as in (16) and could therefore be captured with person-specific rules in (17) repeated from (4).

- | | | |
|---|------|--|
| (16) a. $S_{P1}V\text{-}\emptyset$
b. $S_{P2}V\text{-}\underline{s}$
c. $S_{P3}V\text{-}\underline{t}$ | (17) | a. $\emptyset \longleftrightarrow [_, 1P] / V + _$
b. $s \longleftrightarrow [_, 2P] / V + _$
c. $t \longleftrightarrow [_, 3P] / V + _$ |
|---|------|--|

In the case of ambiguous endings the Agr head cannot be projected during the parse since there is not enough information to give it semantic content.³³ In contrast, TP-Grammar fares well with all kinds of endings (as long as tense information can be read off of them), but fails when chosen to parse null subject clauses. In those cases in the absence of a subject DP providing presupposition triggering features, the domain of

³³Compare this with an early proposal of Yang (1999, 29) similar in spirit, "If an Italian grammar is chosen by an English child for grammatical analysis, it fails to check the Agr feature due to English's lack of Agr morphology, and is punished as a result", although there Yang does not specify what the lack of Agr morphology implies exactly in terms of empirical evidence, since any given ending, including null, can be part of a full-fledged agreement paradigm. As stated above, we identify it with ending ambiguity.

the external argument of the verb is left underspecified, and the composition does not converge. Now a diachronically stable null subject situation is predicted to obtain in case there are no problematic data of the kind described above, that is, there are no ambiguous endings and the penalty probability c_{Agr} is 0. This means that AgrP-Grammar never fails and in every generation ends up driving the competing TP-Grammar out since the latter cannot parse some of the AgrP-Grammar’s output, namely null subject clauses. In what follows we estimate the penalty probabilities c_{Agr} and c_{TP} in Medieval French based on corpus data in order to model the disappearance of the null subject grammar and its relation to endings’ ambiguity.

5.3 Estimating failure probabilities

To estimate c_{Agr} , we exhaustively classify verbal endings as ambiguous or unambiguous. We define an ending as ambiguous if it does not correspond to a unique combination of person and number features. Tables 12 and 14 in section A.2 of Appendix A give exhaustive lists of ending types for Groups I and II respectively (we took into account all spelling variants attested in the corpus). We coded every finite clause in the corpus (as usual, with the exclusion of subject *wh*-clauses and imperatives) as to whether the verbal ending is unambiguous, which determines whether the AgrP-Grammar can parse it, and as to whether the clause has an overt subject, which determines whether the TP-Grammar can parse it. In the case of endings which were ambiguous already in the earliest texts (those that have a “yes” in the Ambiguity column in Tables 12 and 14), all clauses with a finite verb having such an ending have been coded as ambiguous. In the case of endings classified as having emerging ambiguity, namely, those that became ambiguous later than in the earliest texts, we classified clauses dated before the first attested cases of ambiguity as having unambiguous predicates and those dated after the ambiguity emerged as having ambiguous predicates. By assumption, all clauses with ambiguous predicates make the AgrP-Grammar fail at parsing. The failure probability c_{Agr} is then estimated as the frequency of clauses with ambiguous predicates at a given date. In the case of TP-Grammar the estimate of c_{TP} is even more straightforward: it is

the frequency of null subject clauses, which, by assumption, TP-Grammar cannot parse.

The predicted value of P_{TP} given c_{Agr} and c_{TP} as estimated using the matrix above is plotted in Fig. 11. We give observation numbers which c_{TP} and c_{Agr} are based on in Tables 19 and 30 respectively in Appendix B.

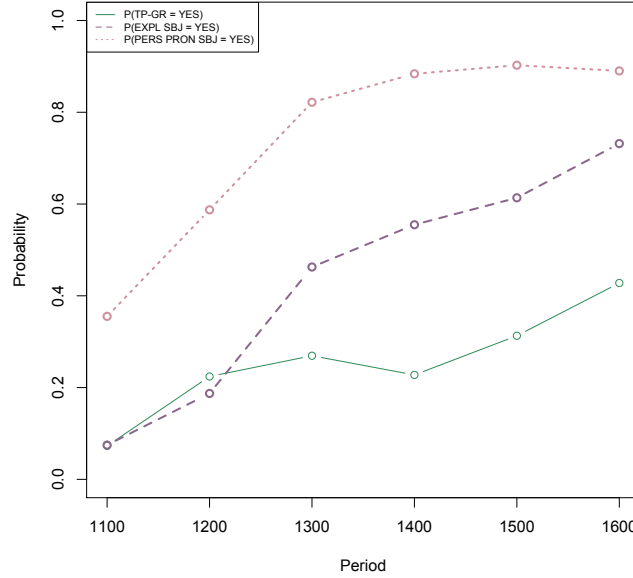


Figure 11: Parsing probability of the TP-Grammar

5.4 Discussion

We estimated the parsing probability of the TP-Grammar based on our estimates of the probabilities with which this grammar and its competitor, AgrP-Grammar, encounter data which they cannot parse. The parsing probability of the TP-Grammar grows steadily during the Medieval period.

Recall that the parsing probability of the TP-Grammar in the limit is the ratio of the probability of AgrP-Grammar to fail to the sum of the AgrP-Grammar and TP-Grammar probabilities to fail, (12). This means that the greater the probability of the AgrP-Grammar to fail, the greater, eventually, will be the probability of the TP-Grammar to be used. Given how we estimate the AgrP-Grammar's probability to fail, that is, as the frequency of ambiguous endings, it is clear that our estimate of the TP-Grammar probability to be used is dependent on the frequency of ambiguous endings. Thus this model, without assuming that ending ambiguity and subject expression depend on the same underlying factor, puts the two in a relation of direct

dependency. This is a welcome configuration given the desiderata expressed in Section 4.5, namely, finding a model which would dissociate the two phenomena at the clause level but would relate them in the course of language evolution. It is worth stressing here that a given ending in this model does not reveal which grammar was used to generate it: by assumption, the spread of syncretic endings is a phonological phenomenon which is “blind” to the syntactic origins of the string it is operating on.

As Fig. 12 shows, the curve corresponding to the parsing probability of the TP-Grammar is roughly parallel to the estimated probabilities of personal and pronominal subject expression. It is not immediately clear how to interpret this parallelism. Intuitively, the probability of the expletive subject expression corresponds to the probability of the TP-Grammar to be chosen for production, since this is the only grammar that can generate overt expletive subjects, expletives, by definition, not being able to meet the pragmatic conditions on overt subjects imposed by the AgrP-Grammar. What we observe, then, is the TP-Grammar parsing probability lagging significantly behind its production probability. One explanation for the apparent lag is that our estimate of the ending ambiguity, to which the TP-Grammar production probability is directly related, is overly conservative. We have already mentioned on several occasions that the spelling standardisation of the mid-XIV c. seems to have had a visible effect on the manuscripts our corpus is based on. The phenomenon we discussed is the reintroduction of the final -t for the Group II verbs with unstressed roots in -i/-u. Given the disappearance of the inflectional -t after vowels prior to 1300, the same should have happened to the -t following glides and stops shortly thereafter (cf. a suggestion in Buridant 2000, 250). And we know for a fact that eventually final inflectional stops did fall in all the environments. However, presumably due to the spelling standardisation, we do not observe these changes and therefore cannot take them into account in our estimates of ending ambiguity, which gives the impression that the latter lags seriously behind the production probability reflected in the rate of overt expletive subjects.

We can see what happens to the parsing probability of the TP-Grammar if we make a less conservative assumption about final consonant fall. That is, let us assume

that, in addition to the endings marked as ambiguous in tables 12 and 14 (-e, -s, zero following i/u, -eies), the following endings were ambiguous as well by virtue of effectively not being pronounced from 1400 on and thus resulting in verbal forms homophonous with either 1st or 2nd person singular forms: -t, -eiet, -it, -et, -at. In Fig. 12 we plotted the predicted parsing probability of the TP-Grammar. One can see that this less conservative estimate is almost identical to the estimated production probability in the form of overt expletives.

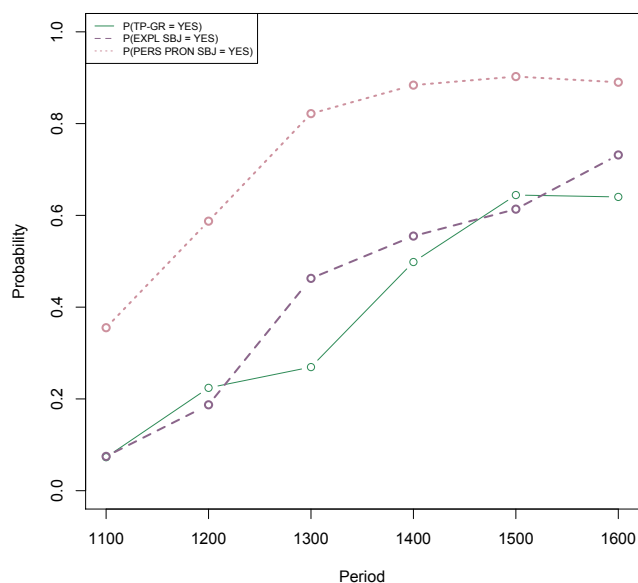


Figure 12: Parsing probability of the TP-Grammar

One may argue, however, that this parallelism in general is not a particularly interesting result since, in addition to the frequency of ambiguous endings, the TP-Grammar parsing probability in the limit inversely depends on the relative frequency of null subjects which, of course, decreases over time. Consequently, the question is whether ending ambiguity actually plays an important role in predicting the TP-Grammar parsing probability.

One way to evaluate the role of the endings for the outcome of the grammar competition is to design a variational learning-based model in a way which would not make reference to them at all and to compare this “ending-less” model to the model that does take them into account, such as the one which have just considered. To this end, we use the measure of grammar fitness proposed in Yang (2000) and used, in particular, in

variational learning models of the loss of V-to-T raising in Scandinavian in Heycock and Wallenberg (2013) and the loss of OV in Latin in Danckaert (2017). Fitness of grammar G is defined as the proportion of clauses which only G generates out of all clauses that G generates. Put differently, fitness measures the proportion of unambiguous clauses in the output of G. For AgrP- and TP-Grammars we estimate fitness as follows:

$$(18) \quad \text{Fitness}(\text{AgrP-Grammar}) = \frac{\# \text{null subjects}}{\# \text{null or overt personal pronominal subjects}}$$

$$(19) \quad \text{Fitness}(\text{TP-Grammar}) = \frac{\# \text{overt expletive subjects}}{\# \text{overt expletive or personal pronominal subjects}}$$

Importantly, ending frequencies do not enter into the equations here since, by assumption, they play no role in identifying the grammar that generated the clause.³⁴

Fitness of the AgrP-Grammar cannot be straightforwardly estimated from our data since, by hypothesis, all the attested stages of historical French correspond to mixed grammar states, that is, to the outputs of the two competing grammars. This follows from the assumption that whenever we find overt expletive subjects, the TP-Grammar must have been at work and from the fact that expletive subjects are found in the earliest attested texts (e.g. Zimmermann 2014; Prévost to appear). Instead, we can approximate the fitness measure of the AgrP-Grammar on the basis of a language that is currently in a “pure” pro-drop state. By (18), AgrP-Grammar fitness corresponds to the estimated probability of null subjects. This parameter has been argued to be around 0.7 for pro-drop languages such as Italian or Spanish (e.g. Bates 1976; Nagy et al. 2011; Otheguy et al. 2007). That is, by assumption, the AgrP-Grammar produces unambiguous clauses with the probability of 0.7. Now the fitness of TP-Grammar corresponds to the estimated probability of expletive subjects in a no-pro-drop language, such as English, which obviously cannot be anywhere near 0.7.³⁵ Given these

³⁴Somewhat paradoxically, this also means that the speaker can use the AgrP-Grammar to generate a clause which the hearer will fail to parse using the same grammar because of a phonological process which renders the ending ambiguous, e.g. *done à la dame* “(I/she/he) give/gives to the lady”, where the verb received a final *-e* syncretic between 1st and 3rd person singular. Notice also that in the absence of such independent intervening processes the fitness measures are directly related to the penalty probabilities we used before: the relative frequency of clauses unambiguously generated by G₁ equals the relative frequency of clauses which G₂ fails to parse. We suggest, however, that independent phonological interventions may disturb this equivalence.

³⁵According to Chocholoušová (2009), in sentence initial position in English, texts dummy subjects occur at the frequency of 0.25% per 10,000 words. The author does not give a frequency in terms of sentences, but if we roughly estimate an average English sentence as consisting of 20 words, this gives us

approximations, $\text{Fitness}(\text{AgrP-Grammar}) \gg \text{Fitness}(\text{TP-Grammar})$. By the Fundamental Theorem of Language Change (Yang 2000, 239), which states that the winner in the long run is always the grammar with a greater fitness, this model predicts that AgrP-Grammar wins hands down, contrary to the historical facts. We thus conclude that a model which factors in ending ambiguity fares better than a model which does not, which supports the assumption that there is a causal relation between ending syncretisation and null subject disappearance.

6 Conclusions

The goal of this paper was to bring parsed corpus data and statistical modelling to bear on the old-standing puzzle of the relation between the disappearance of null subjects and verbal subject agreement syncretisation in the historical development of French. We engaged the Constant Rate Hypothesis in order to explore a model which relates the two changes as reflexes of one underlying structural shift and showed that it generates predictions which are not supported by the data. The key feature of the failed predictions is the unexpected (in the model in question) independence of the two developments. Specifically, we found that the increase in overt personal pronominal subjects was uniform across old and new endings and, likewise, that the spread of the new syncretic endings was uniform across clauses with null and overt subjects.

A second model which we explored related the two changes via the step of language learning (possibly over the lifespan) whereby one change (ending syncretisation) promotes the appearance of sentences which disadvantage a grammar with a null subject option and thus automatically favour the overt subject grammar, thus causing null subject disappearance. We approached this model from two perspectives.

First, we focused on what we assumed to be the parsing capacities of the two competing grammars and used the linear reward-penalty theorem to estimate the evolution of the probabilities of the two grammars over time by estimating their failure

a frequency of 4.8%. That is, 0.05 can be used as a (very rough) approximation of the probability of the unambiguous output by the TP-Grammar.

probabilities at each time point. Crucially, the failure probability of a null subject grammar is taken to be directly related to the frequency of ambiguous endings in the data. This was the core assumption meant to capture the intuitive link between agreement quality and subject expression in the process of language change. Another crucial assumption, which made the model compatible with the facts concerning the surface-level independence of subject and ending types, was that the type of ending is determined by a phonological process which is entirely independent of which grammar is picked by the speaker to generate a given sentence.

As a second possibility, we estimated fitness of null and overt subject grammars based on the data from pure state languages. This time the measure did not rely on ending ambiguity. Estimated this way, fitness gives an advantage to the null subject grammar, and on Yang's Fundamental Theorem it is expected to win, contrary to the historical facts. We conclude thus that so far the best model of null subject disappearance is one which factors in the increase in ending ambiguity without assuming a categorical clause-level dependency between the two phenomena.

This is where this study of diachronic change can contribute to the general theory of language variation. Synchronic studies of variation in null subject expression in Romance languages, to the best of our knowledge, fail to establish verbal ending ambiguity as a relevant factor, and thus leave us with a puzzle as to the nature of Taraldsen's generalisation. For instance, Nagy and Heap (1998) report that whether an ending is ambiguous is not a good predictor of subject expression in Francoprovençal. The same conclusion is reached in Carvalho and Child (2011) based on Spanish material. This agrees with Ranson's (2009) conclusions and our own observations concerning the diachronic French data. Our work thus supplements synchronic variationist studies in that we offer a diachronic model which can capture the relation without postulating a clause-level dependency. This suggests that in some cases the study of natural language variation must include the temporal dimension, otherwise some potentially highly relevant factors will remain "invisible".

It has to be noted that our conclusions do not rule out in principle a clause-level

dependency between surface forms. Such an outcome could, for instance, be the result of a competition between Agr- and TP-Grammars whereby by the end of a variational learning cycle they end up in a complementary distribution with respect to tense/aspect environments. An analysis along these lines would need to be worked out for the systems where the only contexts disavouring subject omission are certain tense/aspectual forms syncretic with respect to subject person, such as, for instance, some Northern Italian, Franco-Provençal, and Occitan dialects (Manzini and Savoia 2005), Hebrew (Shlonsky 2009), Finnish (Koenenman 2006), Irish (Speas 1995), Russian (Bizzarri 2015). We leave the search for the best analysis of those to future research.

This study is part of a more general agenda of using diachronic material for the study of interfaces, that is, formal relations between syntax, morphology, phonology, and semantics/pragmatics. Another prominent group of what seems to be parallel and potentially related changes is the disappearance of nominal case marking and word order changes. Simonenko et al. (2015) show that the remnants of the case opposition in Medieval French disappear within approximately the same timeframe as the possibility of having an OV order. It remains to be seen in further research if any of the models explored above can be used to explore the nature of the relation between these two changes.

References

- Adams, Marianne (1987). From Old French to the theory of pro-drop. *Natural Language & Linguistic Theory* 5:1–32.
- Alexiadou, Artemis, and Elena Anagnostopoulou (1998). Parametrizing AGR: Word order, V-movement and EPP-checking. *Natural Language & Linguistic Theory* 16:491–539.
- Alonso-Ovalle, Luis, and Francesco D’Introno (2000). Full and null pronouns in Spanish: the zero pronoun hypothesis. In *Hispanic Linguistics at the Turn of the Millenium: Papers from the 3rd Hispanic Linguistics Symposium*, ed. Héctor Campos, Elena

- Herburger, Alfonso Morales-Front, and Thomas J. Walsh. Somerville, MA: Cascadilla Press.
- Bates, Elizabeth (1976). *Language and context: The acquisition of pragmatics*, volume 13. New York: Academic Press.
- Bettens, Olivier (2015). Chantez-vous français ? Remarques curieuses sur le français chanté de Moyen Âge à la période baroque. URL <http://virga.org/cvf/>.
- Bizzarri, Camilla (2015). Russian as a partial pro-drop language: Data and analysis from a new study. *Annali di Ca' Foscari. Serie occidentale* 49.
- Buridant, Claude (2000). *Nouvelle grammaire de l'ancien français*. Paris: Sedes.
- Bush, Robert R., and Frederick Mosteller (1958). *Stochastic models for learning*. John Wiley & Sons, Inc.
- Camacho, José (2013). *Null subjects*. Cambridge University Press.
- Carvalho, Ana M., and Michael Child (2011). Subject pronoun expression in a variety of Spanish in contact with Portuguese. In *Selected Proceedings of the 5th Workshop on Spanish Sociolinguistics*, 14–25.
- Chocholoušová, Bohumila (2009). Dummy subjects in English, Norwegian and German. A parallel corpus study. Doctoral Dissertation, Masarykova univerzita.
- Cooper, Robin (1983). *Quantification and syntactic theory*. Dordrecht: Reidel.
- Danckaert, Lieven (2017). The loss of latin *ov*: Steps towards an analysis. In *Elements of comparative syntax: Theory and description*, volume 127, 401. Walter de Gruyter GmbH & Co KG.
- De Cat, Cécile (2005). French subject clitics are not agreement markers. *Lingua* 115:1195–1219.
- De Jong, Thera (2006). *La prononciation des consonnes dans le français de Paris aux 13ème et 14ème siècles*. Utrecht: LOT/Netherlands Graduate School of Linguistics.

- Dees, Anthonij, S. Meilink, Karin van Reenen-Stein, and Pieter van Reenen (1980). Un cas d'analogie: l'introduction de *-e* à la première personne du singulier de l'indicatif présent des verbes en *-er* en ancien français. *Rapport/Het Franse Boek* 50:105–110.
- Erker, Daniel, and Gregory R. Guy (2012). The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88:526–557.
- Ewert, Alfred (1943). *The French Language*. London: Faber & Faber.
- Fontaine, Carmen (1985). Application de méthodes quantitatives en diachronie: L'inversion du sujet en français. Master's thesis. Université du Québec à Montréal.
- Fouché, Pierre (1931). *Le Verbe français: étude morphologique*. Paris: Société d'édition "Les Belles Lettres".
- Foulet, Lucien (1928). *Petite syntaxe de l'ancien français*. Paris: Champion, troisième édition revue. Réédition 1982.
- Foulet, Lucien (1935). L'extension de la forme oblique du pronom personnel en ancien français. *Romania* 61–62:257–315.
- Franzén, Torsten (1939). Étude sur la syntaxe des pronoms personnels sujets en ancien français. Doctoral Dissertation, Uppsala.
- Fruehwald, Josef, Jonathan Gress-Wright, and Joel Wallenberg (2009). Phonological rule change: The constant rate effect. In *Proceedings of the 40th Annual Meeting of the North East Linguistic Society*, ed. Seda Kan, Claire Moore-Cantwell, and Robert Staubs, volume 1, 219–230. GLSA Publications.
- Goyette, Stéphane (1993). Le système verbal de Jacques Peletier du Mans, XVI^e siècle. Master's thesis, Université de Montréal.
- Harley, Heidi, and Elizabeth Ritter (2002). Person and number in pronouns: A feature-geometric analysis. *Language* 78:482–526.

- Haspelmath, Martin (1999). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft* 18:180–205.
- Heim, Irene (2008). Features on bound pronouns. In *Phi Theory: Phi-features across Modules and Interfaces*, ed. David Harbour, David Adger, and Susana Bejar, 35–52. Oxford: Oxford University Press.
- Heim, Irene, and Angelika Kratzer (1998). *Semantics in Generative Grammar*. Oxford: Blackwell.
- Heycock, Caroline, and Joel Wallenberg (2013). How variational acquisition drives syntactic change. *The Journal of Comparative Germanic Linguistics* 16:127–157.
- Hirschbühler, Paul (1992). L’omission du sujet dans les subordonnées V1: les CNN de Vigneulles et les CNN anonymes. *Travaux de linguistique* 24:25–46.
- Hirschbühler, Paul, and Marie-Odile Junker (1988). Remarques sur les sujets nuls en subordonnées en ancien et en moyen français. *Revue québécoise de linguistique théorique et appliquée* 7:63–84.
- Ingason, Anton Karl (2016). PaCQL: A new type of treebank search for the digital humanities. *Italian Journal of Computational Linguistics* 2:51–66.
- Jaeggli, Osvaldo, and Kenneth J. Safir (1989). The null subject parameter and parametric theory. In *The null subject parameter*, ed. Osvaldo Jaeggli and Kenneth J. Safir, 1–44. Springer.
- Kaiser, Georg A. (2009). Losing the null subject. A contrastive study of (Brazilian) Portuguese and (Medieval) French. In *Proceedings of the Workshop “Null-subjects, expletives, and locatives in Romance”*, 131–156.
- Kauhanen, Henri, and George Walkden (2017). Deriving the Constant Rate Effect. *Natural Language and Linguistic Theory* .
- Klima, Edward Stephan (1965). Studies in diachronic transformational syntax. Doctoral Dissertation, Harvard University.

- Koenenman, Olaf (2006). Deriving the difference between full and partial pro-drop. In *Arguments and agreement*, ed. Peter Ackema, Patrick Brandt, Maaïke Schoorlemmer, and Fred Weerman, 76–100. Oxford: OUP.
- Kratzer, Angelika (2009). Making a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry* 40:187–237.
- Kroch, Anthony (1989). Reflexes of grammar in patterns of language change. *Language variation and change* 1:199–244.
- Manzini, Maria Rita, and Leonardo Maria Savoia (2005). *I dialetti italiani e romanci: morfosintassi generativa*, volume 1. Edizioni dell’Orso.
- Marchello-Nizia, Christiane (1992). *Histoire de la langue française aux XIVe et XVe siècles*. Paris: Dunod.
- MCVF (2010). Corpus MCVF annoté syntaxiquement, dirigé par France Martineau, avec Paul Hirschbühler, Anthony Kroch et Yves Charles Morin. URL http://www.voies.uottawa.ca/corpus_pg_en.html.
- Morin, Yves Charles (2001). La troncation des radicaux verbaux en français depuis le Moyen Age. Études diachroniques. *Recherches linguistiques de Vincennes* 63–85.
- Nagy, Naomi, and David Heap (1998). Francoprovençal null subjects and constraint interaction. *CLS* 34:151–166.
- Nagy, Naomi G., Nina Aghdasi, Derek Denis, and Alexandra Motut (2011). Null subjects in heritage languages: Contact effects in a cross-linguistic context. *University of Pennsylvania Working Papers in Linguistics* 17:16.
- Narendra, Kumpati S., and Mandayam A.L. Thathachar (1989). *Learning Automata: An Introduction*. Englewood Cliffs, NJ: Prentice-Hall.
- Otheguy, Ricardo, Ana Celia Zentella, and David Livert (2007). Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language* 770–802.

- Paolillo, John C. (2011). Independence claims in linguistics. *Language Variation and Change* 23:257–274.
- Penn Supplement to MCVF (2010). Penn Supplement to the MCVF Corpus by Anthony Kroch and Beatrice Santorini.
- Pintzuk, Susan (1995). Variation and change in Old English clause structure. *Language Variation and Change* 7:229–260.
- Pintzuk, Susan (2004). Variationist approaches to syntactic change. In *The handbook of historical linguistics*, ed. Brian D. Joseph and Richard D. Janda, 509–528. Wiley Online Library.
- Poletto, Cecilia (2006). Asymmetrical pro-drop in northern Italian dialects. In *Arguments and agreement*, ed. Peter Ackema, Patrick Brandt, Maaïke Schoorlemmer, and Fred Weerman, 159–191. Oxford: Oxford University Press.
- Prévost, Sophie (to appear). Increase of Pronominal Subjects in Old French: Evidence for a Starting-point in Late Latin. In A. Carlier et C. Guillot (éds) *Latin tardif – français ancien: continuités et ruptures*, coll. *Beihefte zur Zeitschrift für romanische Philologie*, Berlin: De Gruyter.
- Ranson, Diana L. (2009). Variable subject expression in Old and Middle French prose texts: The role of verbal ambiguity. *Romance Quarterly* 56:33–45.
- van Reenen, Pieter, and Lene Schøsler (1987). Le problème de la prolifération des explications. *Vrije Universiteit working papers in linguistics* 27.
- Rizzi, Luigi (1986). Null objects in Italian and the Theory of *pro*. *Linguistic Inquiry* 17:501–557.
- Roberts, Ian (1993). *Verbs and Diachronic Syntax: A Comparative History of English and French*. Dordrecht: Kluwer.

- Roberts, Ian (2010). A deletion analysis of null subjects. In *Parametric variation: Null subjects in Minimalist theory*, ed. Biberauer, Theresa and Holmberg, Anders and Roberts, Ian and Sheehan, Michelle, 58–87. Cambridge University Press.
- Roberts, Ian (2014). Taraldsen’s Generalization and Language Change: Two Ways to Lose Null Subjects. In *Functional Structure from Top to Toe: The Cartography of Syntactic Structures*, ed. Peter Svenonius, volume 9, 115–148. Oxford University Press.
- Santorini, Beatrice (1993). The rate of phrase structure change in the history of Yiddish. *Language Variation and Change* 5:257–283.
- Sauerland, Uli (2008). On the semantic markedness of phi-features. In *Phi Theory: Phi-features across Modules and Interfaces*, ed. David Harbour, David Adger, and Susana Bejar, 527–82. Oxford: Oxford University Press.
- Schøsler, Lene (2002). La variation linguistique: le cas de l’expression du sujet. In *Interpreting the History of French, A Festschrift for Peter Rickard on the occasion of his eightieth birthday*, ed. Rodney Sampson and Wendy Ayres-Bennett, 187–208. Amsterdam - New York, NY: Rodopi.
- Sheehan, Michelle (2016). Subjects, null-subjects and expletives in Romance. In *Manuals of Grammatical Interfaces in Romance*, ed. S. Fischer and S. Gabriel. Berlin: Mouton de Gruyter. URL <http://ling.auf.net/lingbuzz/002466>.
- Shlonsky, Ur (2009). Hebrew as a partial null-subject language. *Studia linguistica* 63:133–157.
- Simonenko, Alexandra, Benoit Crabbé, and Sophie Prévost (2015). Morphological triggers of syntactic changes: Treebank-based information theoretic approach. In *Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT14)*, ed. Markus Dickinson, Erhard Hinrichs, Agnieszka Patejuk, and Adam Przepiórkowski, 194–205. Warsaw: Institute of Computer Science, Polish Academy of Sciences.

Speas, Margaret (1995). Economy, agreement and the representation of null arguments. Ms. University of Massachusetts, Amherst.

Sprouse, Rex, and Barbara Vance (1999). An explanation for the decline of null pronouns in certain Germanic and Romance dialects. In *Language creation and language change: creolization, diachrony and development*, ed. Michel deGraff, 257–284. Cambridge, Mass.: MIT Press.

Taraldsen, Tarald (1980). *On the NIC, vacuous application and the that-trace filter*. Indiana University Linguistics Club.

Tvica, Seid (2017). On minimal person and number features. URL https://www.academia.edu/23278325/On_the_universality_of_person_and_number, unpublished manuscript.

Vance, Barbara (1997). *Syntactic change in medieval French*. Studies in Natural Language and Linguistic Theory. Dordrecht/Boston/London: Kluwer.

Vennemann, Theo (1975). An explanation of drift. In *Word Order and Word Order Change*, ed. Charles N. Li, 269–305. Austin: University of Texas Press.

Yang, Charles (1999). The variational dynamics of natural language: acquisition and change. Ms. MIT Artificial Intelligence Laboratory, MIT, Cambridge, MA .

Yang, Charles (2000). Internal and external forces in language change. *Language Variation and Change* 12:231–250.

Yang, Charles (2002). *Knowledge and learning in natural language*. Oxford University Press.

Yang, Charles (2010). Three factors in language variation. *Lingua* 120:1160–1177.

Zimmermann, Michael (2009). On the evolution of expletive subject pronouns in Old French. In *Proceedings of the workshop “null-subjects, expletives, and locatives in romance”*. *arbeitspapier 123*, ed. E.-M. Remberger G.A. Kaiser, 63–92. Fachbereich Sprachwissenschaft, Universität Konstanz.

Zimmermann, Michael (2014). *Expletive and referential subject pronouns in medieval french*, volume 556. Walter de Gruyter GmbH.

A Illustrations of spelling variation and paradigms

A.1 Syncretisation as reflected in spelling

Below we provide some illustrations for the variation accompanying the rise of the new endings. Paragraph letters correspond to the changes listed in (3). **A.** The use of the ending -e instead of zero for the 1st person singular subjects with Group I verbs, according to Marchello-Nizia (1992, 200), began in the XII c. and by the beginning of the XV c. generalized onto the roots ending in a consonant, the zero ending lingering for longer with stems ending in a vowel. This alternation is illustrated by the two examples below. A handful of verbs whose stems etymologically ended in -e, such as monstre-r “to show” were not affected by this change.

- (20) A lui **lais** jo mes honors e mes fieus.
 to him leave.1P I my lands and my fiefs
 ‘I leave to him my lands and my fiefs.’ (1100-ROLAND-V,23.290)

- (21) Je vous **laiss-e** ma fenme la roine enchainte.
 I you leave-1/3P my wife the queen pregnant
 ‘I leave to you my wife, the queen, pregnant.’ (1370-FROISSART-2-P,174.2966)

B. The emergence of final -e as a consequence of the disappearance of the final -t in the context of the 3rd person singular subjects (either null or overt), illustrated in (22)–(23), is generally considered to predate the changes in the 1st person singular contexts.

- (22) Il en **apele-t** e ses dux e ses cuntes...
 he of.there call-3P and his dukes and his counts
 ‘He calls both his dukes and his counts...’ (1100-ROLAND-V,2.15)

- (23) Nuls ne l’ **apele** fors le rei.
 No.one not him accuse.1/3P except the king
 ‘No one accuses him but the king.’ (116X-MARIE-DE-FRANCE-R,86.1762)

C. The alternation between -et and an innovative -e as the endings of the 3rd person singular present subjunctive in Group II is illustrated in (24) and (25) respectively.

- (24) ne.s poet guarder que mals ne l' i **ateign-et**.
 neg.refl can prevent that evil not him there attain-3P
 "He can not prevent that evil reaches him there." (1100-ROLAND-V,1.9)

- (25) E Deus le nos **cunsent-e**!
 and God it to.us agree-1/3P
 "May God give this to us!" (1100-ROLAND-V,216.2994)

F.³⁶ In Group II the ending -t alternated with zero in the context of the 3rd person singular in preterite. This case is special in that the innovative zero ending was on the rise up until the mid-XIV c., when it suddenly went into a sharp decline, the old ending reinstalling itself completely. In our discussion of the spelling-pronunciation correspondence in section 4.5.1 we take this fact to indicate that the mid-XIV c. was a cut-off point in spelling-pronunciation contiguity, and, therefore, it gives support to the assumption that until that point spelling and pronunciation went largely hand in hand.

- (26) ...cum il l' **audi-t**...
 when he him heard-3P
 "When he heard him." (0980-LEGER-V,XXXII.210)

- (27) Dunc **respundi** li evesche Hely...
 so responded.1/3P the high priest Eli
 "So the the high priest Eli responded..." (1150-QUATRELIVRE-P,5.62)

G. The variation between a new, syncretic ending -s and non-syncretic zero for the 1st person singular with Group II verbs is illustrated in (28)–(29). According to Marchello-Nizia (1992, 201), -s occurs very infrequently before the XIV c. We showed that this change is indeed delayed compared to the spread of -e. Marchello-Nizia (1992, 202) observes that in the case of the stems ending in a vowel it takes longer for the new variant to establish itself. There is also a limited number of verbs with stems ending in -s for etymological reasons (e.g. *finis* < Lat. *finisco* "to finish").

- (28) jol **sai** bien.
 I.it know.1P well
 'I know it well' (1100-ROLAND-V,59.728)

- (29) je **sai-s** qu' aves esté si transportée de joie...
 i know-1/2P that have been so elated by joy
 'I know that (you) have been so elated by joy...' (158X-VALOIS-AUTOGRAPH-P,99.408)

³⁶We omit illustrations of the D–E cases in the interest of space since those changes do not result in syncretism.

H–I. Finally, the following two examples illustrate the alternation related to the innovative use of -s in imperfect and future conditional with 1st person singular subjects. These examples involve Group I verbs, but the pattern, as mentioned above, extends onto both groups.

- (30) que j' ai anpris une autre afeire que je ne **lesseroie** pas.
 that I have undertaken an other deal that I not would.leave.1P not
 “...that I undertook another deal that I would not give up.” (1170-YVAIN-R,145.5051)
- (31) j' **aymeroi-s** mieulx mourir que vous presser de deshonneur.
 I would.like-1/2P better die that you torture of dishonour
 ‘I would prefer to die than to torture you by the dishonour.’ (1527-BAYART-P,49.929)

A.2 Ambiguity estimates

ENDING		PERSON & NUMBER (TENSE & MOOD)	AMBIGUITY
<u>-∅</u>		1st sg (present indicative & subjunctive)	no
<u>-e</u>	[e]	1st & 3rd sg (present indicative & subjunctive)	yes
<u>-es</u>	[es]	2nd sg (present indicative)	no
<u>-s</u>	[s]	2nd sg (present subjunctive)	no
<u>-t</u>	[t]	3rd sg (present indicative & subjunctive)	no
<u>-ai</u>	[ɛ]	1st sg (preterite; future)	no
(a) <u>-s</u>	[s]	2nd sg (preterite; future)	no
(a) <u>-∅</u>	∅	3rd sg (preterite; future)	no
(a) <u>-t</u>	[t]	3rd sg (preterite; future)	no
(oie) <u>-∅</u>		1st sg (imperfect; future conditional)	no
(oie) <u>-s</u>	[s]	1st & 2nd sg (imperfect; future conditional)	emerging
(oi) <u>-t</u>	[t]	3rd sg (imperfect; future conditional)	no
(ss) <u>-e</u>	[e]	1st sg (imperfect subjunctive)	no
(ss) <u>-es</u>	[es]	2nd sg (imperfect subjunctive)	no
(s) <u>-t</u>	[t]	3rd sg (preterite; imperfect subjunctive)	no
<u>-on(s)</u>	[ons] > [ō]	1st pl	no
<u>-ez</u>	[ets] > [es]	2nd pl	no
<u>-ent</u>	[ent]	3rd pl	no

Table 12: Person and number endings Group I

The distribution of the person/number endings is illustrated by the following paradigms, **taken mostly** from Buridant (2000).³⁷

³⁷We assume that the 2nd person plural ending in the context of all moods and tenses was never ambiguous. Even though, according to De Jong (2006, 66), the final -s in this ending was likely unstable around the XIII c., this would not make the ending ambiguous, as one can see from Table 12. We therefore do not list here its variants in various contexts. According to De Jong (2006), the passage

	1ST SG	2ND SG	3RD SG	1ST PL	2ND PL	3RD PL
PRESENT INDICATIVE	aim(e)	aines	aime	amons	amez	aiment
SUBJUNCTIVE	aim(e)	ains	aint/aime	amons	amez	aiment
PRETERITE	amai	amas	ama(t)	amames	amastes	amerent
IMPERFECT SUBJUNCTIVE	amasse	amasses	amast/se	amissons	amisseiz	amassent
FUTURE	amerai	ameras	amera(t)	amerons	amerez	ameront
IMPERFECT	amoie(s)	amoies	amoit	amiiens	amiiez	amoeent
FUTURE CONDITIONAL	ameroie	ameroies	ameroit	amerions	ameriez	ameroient

Table 13: Group I aimer “to love”

Ending		Person & Number features (Tense & Mood contexts)	Ambiguity
<u>-∅</u>		1st sg (present indicative)	no
<u>-s</u>	[s]	1st & 2nd sg (present indicative)	yes
<u>-t</u>	[t]	3rd sg (present indicative)	no
<u>-e</u>	[e]	1st & 3rd sg (present subjunctive)	yes
<u>-es</u>	[es]	2nd sg (present subjunctive)	no
<u>-et</u>	[et]	3rd sg (present subjunctive)	no
<u>(-i/-u)-∅</u>		1st & 3rd sg (preterite)	yes
<u>(-i/-u)-s</u>	[s]	2nd sg (preterite)	no
<u>(-i/-u)-t</u>	[t]	3rd sg (preterite)	no
<u>-ai</u>	[ɛ]	1st sg (future)	no
<u>(a)-s</u>	[s]	2nd sg (future)	no
<u>(a)-∅</u>		3rd sg (future)	no
<u>(a)-t</u>	[t]	3rd sg (future)	no
<u>(oie)-∅</u>		1st sg (imperfect; future conditional)	no
<u>(oie)-s</u>	[s]	1st & 2nd sg (imperfect; future conditional)	emerging
<u>(oi)-t</u>	[t]	3rd sg (imperfect; future conditional)	no
<u>(ss)-e</u>	[e]	1st sg (imperfect subjunctive)	no
<u>(ss)-es</u>	[es]	2nd sg (imperfect subjunctive)	no
<u>(s)-t</u>	[t]	3rd sg (preterite; imperfect subjunctive)	no
<u>-on(s)</u>	[ons] > [õ]	1st pl	no
<u>-ez</u>	[ets] > [es]	2nd pl	no
<u>-ent</u>	[ent]	3rd pl	no

Table 14: Person and number endings Group II

from an affricate to a sibilant in the pronunciation of z in 2nd person plural marker happened during the second half of the XIV c.

	1ST SG	2ND SG	3RD SG	1ST PL	2ND PL	3RD PL
PRESENT INDICATIVE	vien(s)	viens	vient	venons	venez	vienent
SUBJUNCTIVE	dorme	dormes	dorme	dormons	dormez	dorment
SUBJUNCTIVE			hace/hast			
IMPERFECT SUBJUNCTIVE	covrisse	covrisses	covrist/se	covrissons	covrissez	covrissent
PRETERITE	dormiØ	dormis	dormi(t)	dormimes	dormistes	dormirent
FUTURE	vaudrai	vaudras	vaudra(t)	vaudrons	vaudrez	vaudront
IMPERFECT	valoie(s)	valoies	valoit	valiions	valiiez	valoeient
FUTURE CONDITIONAL	vaudroie	vaudroies	vaudroit	vaudrions	vaudriez	vaudroient

Table 15: Group II venir “to come”; dormir “to sleep”, couvrir “to cover”; valoir “to be worth”; haïr “to hate”

B Corpora and observation numbers

B.1 MCVF (2010) and Penn Supplement to MCVF (2010)

The quantitative data used in this paper were extracted from Penn-scheme annotated corpora MCVF (2010) and Penn Supplement to MCVF (2010) using Parsed Corpora Query Language (PCQL) and CorpusSearch software.³⁸ The parsing scheme and the query language are thoroughly described in Ingason (2016). Lists of morphological and syntactic tags used specifically in the MCVF can be found at the following locations:

<http://gtrc.voies.uottawa.ca/manuel/manuel-morpho-fr/index.htm> and

<http://gtrc.voies.uottawa.ca/manuel/syntax-manual-fr/index.htm>.

CorpusSearch can code chunks of structure (of any size) for any number of annotation parameters and extract coding strings as datasets. As an example, consider two clauses in (32) and (34) with parses in (33) and (35) respectively which were coded for the variables *Date*, *Clause type*, *Subject type*, *Subject person*, *Subject position*, *Conjugation type*, *Text form* and *Ending type*.

- (32) de sanct Maxens abbas divint.
of Saint Maxence abbot became
“(He) became the abbot of Saint Maxence” (0980-LEGER-V,V.30)

³⁸<http://corpussearch.sourceforge.net/>

- (33) ((IP-MAT (CODING-IP-MAT **0980:mat:no:null:na:pre:second:verse:t**)
 (NP-PRD (PP (P de)
 (NP-PP (ADJ sanct) (NPRS Maxens)))
 (NCS abbas))
 (NP-SBJ *pro*)
 (VJ divint)
 (PONFP .))
 (ID 0980-LEGER-V,V.30))
- (34) par le divine volentét, il derables icel sul filz angendrat.
 by the divine will he desired this his son conceived
 “By the divine will, he conceived this much desired son of his.” (10XX-ALEXIS-V,0.5)
- (35) ((IP-MAT (CODING-IP-MAT **1050:mat:yes:pron:third:pre:first:verse:at**)
 (PP (P par)
 (NP-PP (D le) (ADJ divine) (NCS volentét)))
 (PON ,)
 (NP-SBJ (PRO il))
 (ADJP-SPR (ADJ derables))
 (NP-ACC (D icel) (ADJ sul) (NCS filz))
 (VJ angendrat)
 (PONFP .))
 (ID 10XX-ALEXIS-V,0.5))

These codings can be extracted and treated by a statistical software as a dataset such as the toy example below:

Date:	Clause:	Subject:	Type:	Person:	Position:	Conjugation:	Form:	Ending
1050:	mat:	yes:	pron:	third:	pre:	first:	verse:	at
0980:	mat:	no:	null:	na:	pre:	second:	verse:	t

Table 16: A corpus-based dataset example

The following table gives the available relevant information about corpus texts. In this table, some texts are assigned artificial *exact* dates for methodological reasons, since the actual date attribution for the manuscripts of the earliest periods is rarely precise and is usually done using a half or a third of a century intervals rather than exact dates.³⁹

³⁹More information about the manuscript editions used in the treebanks can be found at http://www.voies.uottawa.ca/corpus_pg_en.html and <http://www.ling.upenn.edu/~beatrice/corpus-ling/frenchTexts.html>.

Date	Title	Author	Form	Word-count
980	Saint Léger	na	verse	1,443
1050	La vie de Saint Alexis	na	verse	4,996
1100	La chanson de Roland	na	verse	29,818
1120	Le voyage de Saint Brendan	Benedeit	verse	11,018
1130	Leis Willelme	na	prose	
1155	Li quatre livre des reis	na	prose	44,199
1177	Yvain	Chrétien de Troyes	verse	42,423
1180	Les lais	Marie de France	verse	33,489
1190	Joseph d'Arimatee	Robert de Boron	prose	20,789
1194	Chartre de Chievres	na	prose	
1200	Aucassin et Nicollette	na	prose, verse	10,203
1205	La Conquête de Constantinople	Robert de Clari	prose	34,560
1212	La vie de Saint Eustache (ed. J.R. Fischer)	Pierre de Beauvais	verse	9,946
1220	Chronique de (pseudo)Turpin	na	prose	26,577
1223	Sermon sur Sainte-Agnès	na	prose	4,690
1224	La vie de Saint Eustache (ed. H. Petersen)	na	verse	11,884
1225	La Queste del Saint Graal	na	prose	36,909
1255	Saint Eustache (ed. J. Murray)	na	prose	7,434
1270	Cassidorus	na	prose	52,350
1275	Livre Roisin	Jean Roisin	prose	
1279	Somme des vices et des vertus	Frère Laurent	prose	27,945
1309	La vie de Saint Louis	Jean de Joinville	prose	76,806
1330	Roman de Perceforest	na	prose	53,461
1370	Prise d'Alexandrie	Guillaume de Machaut	verse	53,287
1373	Chroniques	Jean de Froissart	prose	221,129
1427	Formulaire de la chancellerie royale	Odart Morchesne	prose	
1440	XV Joyes de mariage	na	prose	154,242
1450	Les Cent Nouvelles Nouvelles	na	prose	154,242
1498	Mémoires	Philippe de Commynes	prose	44,626
1523	Le nouveau testament	transl. Jean Lefèvre D'Etaples	prose	34,784
1527	Histoire du Seigneur de Bayart	na	prose	57,961
1572	Correspondance	Marguerite de Valois	prose	30,028

Table 17: MCVF (2010) and Penn Supplement to MCVF (2010) texts

B.2 Observation numbers

	Matrix			Subordinate		
Date	overt	null	overt	overt	null	overt
980	24	14	0.37	40	21	0.34
1050	111	32	0.22	96	60	0.38
1100	665	220	0.25	264	221	0.46
1120	262	29	0.10	291	71	0.20
1130	6	5	0.45	15	98	0.87
1155	404	172	0.30	256	635	0.71
1177	541	291	0.35	487	1119	0.70
1180	920	315	0.26	395	634	0.62
1190	420	338	0.45	431	1359	0.76
1194	2	0	0	4	40	0.91
1200	116	174	0.60	44	291	0.87
1205	212	174	0.45	42	1114	0.96
1212	176	67	0.28	161	176	0.52
1220	144	102	0.41	122	415	0.77
1223	12	79	0.87	4	166	0.98
1224	165	46	0.22	218	227	0.51
1225	94	347	0.79	76	1477	0.95
1255	34	101	0.75	10	215	0.96
1270	503	804	0.62	290	2105	0.88
1275	7	11	0.61	23	152	0.87
1279	68	141	0.67	49	732	0.94
1309	196	753	0.79	90	2935	0.97
1330	175	481	0.73	133	1307	0.91
1370	468	303	0.39	505	1163	0.70
1373	366	1455	0.80	339	4143	0.92
1427	23	17	0.42	55	248	0.82
1440	103	639	0.86	95	1392	0.94
1450	384	2573	0.87	877	5028	0.85
1498	111	248	0.69	126	993	0.89
1523	14	698	0.98	76	855	0.92
1527	168	538	0.76	340	1373	0.80
1572	1	501	1	157	1176	0.88

Table 18: Overt and null pronominal subjects in main and subordinate clauses

	Personal			Expletive		
Date	null	overt	overt	null	overt	overt
980	60	35	0.37	4	0	0
1050	175	91	0.34	32	1	0.03
1100	767	426	0.36	162	15	0.08
1120	457	98	0.18	96	2	0.02
1130	18	101	0.85	3	2	0.40

1155	568	800	0.58	92	7	0.07
1177	842	1321	0.61	186	89	0.32
1180	1043	931	0.47	272	18	0.06
1190	726	1647	0.69	125	50	0.29
1194	5	37	0.88	1	3	0.75
1200	123	449	0.78	37	16	0.30
1205	200	1216	0.86	54	72	0.57
1212	284	234	0.45	53	9	0.15
1220	192	501	0.72	74	16	0.18
1223	14	244	0.95	2	1	0.33
1224	344	264	0.43	39	9	0.19
1225	107	1711	0.94	63	113	0.64
1255	25	304	0.92	19	12	0.39
1270	508	2647	0.84	285	262	0.48
1275	2	147	0.99	28	16	0.36
1279	78	818	0.91	39	55	0.59
1309	167	3438	0.95	119	250	0.68
1330	213	1628	0.88	95	160	0.63
1370	788	1323	0.63	185	143	0.44
1373	358	5221	0.94	347	377	0.52
1427	44	220	0.83	34	45	0.57
1440	87	1839	0.95	111	192	0.63
1450	831	6930	0.89	430	671	0.61
1498	123	1055	0.90	114	186	0.62
1523	64	1481	0.96	26	72	0.73
1527	366	1644	0.82	142	267	0.65
1572	143	1517	0.91	15	160	0.91

Table 19: Overt and null personal and expletive subjects

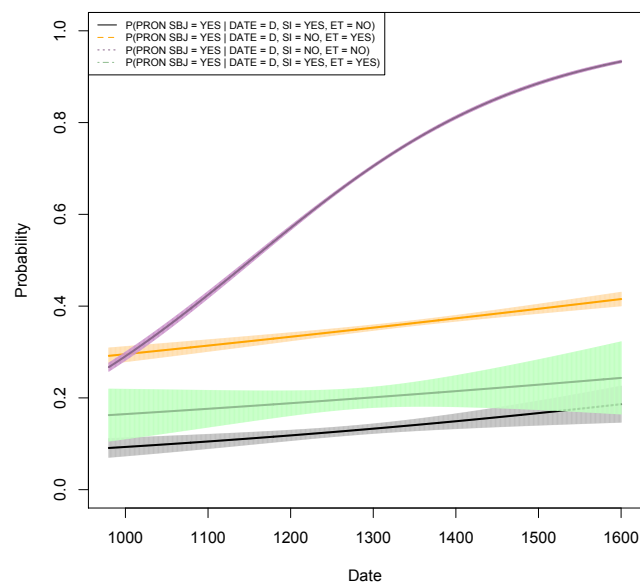


Figure 13: Overt pronominal subjects with and without connectives

	980	1050	1100	1120	1130	1155	1177	1180	1190	1194	1200	1205	1212	1220	1223	1224	1225	1255	1270	1275	1279	1309	1330	1370	1373	1427	1440	1450	1498	1523	1527	1572
zero	0	0	13	3	0	6	43	19	13	0	17	6	2	9	2	3	24	4	17	0	1	12	7	21	7	0	12	11	0	0	0	0
-e	0	0	3	3	0	10	3	1	8	0	4	5	0	3	0	0	0	0	26	0	5	51	40	37	69	0	53	244	16	51	68	26
-e	na	na	0.19	0.50	na	0.62	0.07	0.05	0.38	na	0.19	0.45	0	0.25	0	0	0	0	0.60	na	0.83	0.81	0.85	0.64	0.91	na	0.82	0.96	1	1	1	1

Table 20: Ending variation in Group I present indicative and subjunctive & 1st sg subject

	980	1050	1100	1120	1130	1155	1177	1180	1190	1194	1200	1205	1212	1220	1223	1224	1225	1255	1270	1275	1279	1309	1330	1370	1373	1427	1440	1450	1498	1523	1527	1572
-t	2	19	160	25	9	24	29	7	5	0	0	0	1	25	7	0	6	0	5	0	1	4	9	17	1	3	3	16	0	0	4	1
-e	1	0	3	8	11	29	191	47	59	1	29	16	17	10	21	21	204	1	64	15	266	92	116	193	134	45	248	285	25	74	48	101
-e	0.33	0	0.02	0.24	0.55	0.55	0.87	0.87	0.92	1	1	1	0.94	0.29	0.75	1	0.97	1	0.93	1	1	0.96	0.93	0.92	0.99	0.94	0.99	0.95	1	1	0.92	0.99

Table 21: Ending variation in Group I present indicative and subjunctive & 3rd sg subject

	980	1050	1100	1120	1130	1155	1177	1180	1190	1194	1200	1205	1212	1220	1223	1224	1225	1255	1270	1275	1279	1309	1330	1370	1373	1427	1440	1450	1498	1523	1527	1572
-et	1	4	22	4	5	10	0	2	0	0	0	0	0	29	2	2	0	0	0	0	1	0	0	3	2	0	1	1	0	0	0	0
-e	0	0	8	0	17	37	74	24	38	1	5	6	8	3	1	15	48	4	31	12	27	62	16	79	28	14	61	94	7	14	21	47
-e	0	0	0.27	0	0.77	0.79	1	0.92	1	1	1	0.86	1	0.09	0.33	0.88	1	1	1	1	0.96	1	1	0.96	0.93	1	0.98	0.99	1	1	1	1

Table 22: Ending variation in Group II imperfect and present subjunctive & 3rd sg subject

	980	1050	1100	1120	1130	1155	1177	1180	1190	1194	1200	1205	1212	1220	1223	1224	1225	1255	1270	1275	1279	1309	1330	1370	1373	1427	1440	1450	1498	1523	1527	1572
-at	10	9	42	14	9	397	0	26	0	0	0	0	0	0	4	0	0	0	1	0	0	4	0	0	0	0	0	0	0	0	1	4
-a	1	0	0	0	2	37	90	111	187	2	40	165	35	204	12	25	169	46	209	4	9	477	359	201	833	1	70	868	200	155	361	40
-a	0.09	0	0	0	0.18	0.09	1	0.81	1	1	1	1	1	1	0.75	1	1	1	1	1	1	0.99	1	1	1	1	1	1	1	1	1	0.91

Table 23: Ending variation in Group I preterite and future & 3rd sg subject

	980	1050	1100	1120	1130	1155	1177	1180	1190	1194	1200	1205	1212	1220	1223	1224	1225	1255	1270	1275	1279	1309	1330	1370	1373	1427	1440	1450	1498	1523	1527	1572
-at	3	0	31	11	6	62	2	13	0	0	1	1	0	0	0	0	1	0	0	1	2	3	6	1	0	2	1	1	1	0	0	1
-a	5	0	0	0	3	2	58	24	164	0	10	45	9	55	4	6	59	6	43	11	11	46	88	68	86	3	113	196	25	48	65	53
-a	0.62	na	0	0	0.33	0.03	0.97	0.65	1	na	0.91	0.98	1	1	1	1	0.98	1	1	1	0.92	0.96	0.97	0.92	0.99	1	0.98	0.99	0.96	1	1	0.98

Table 24: Ending variation in Group II future & 3rd sg subject

	980	1050	1100	1120	1130	1155	1177	1180	1190	1194	1200	1205	1212	1220	1223	1224	1225	1255	1270	1275	1279	1309	1330	1370	1373	1427	1440	1450	1498	1523	1527	1572
-t	8	6	54	34	3	123	73	55	46	0	22	99	17	136	7	26	69	18	134	1	99	314	68	84	93	20	129	1161	71	237	167	15
zero	6	0	1	1	1	221	38	81	80	0	26	82	16	77	11	18	48	24	80	0	11	170	130	75	350	1	0	6	0	0	0	0
zero	0.43	0	0.02	0.03	0.25	0.64	0.34	0.60	0.63	na	0.54	0.45	0.48	0.36	0.61	0.41	0.41	0.57	0.37	0	0.10	0.35	0.66	0.47	0.79	0.05	0	0.01	0	0	0	0

Table 25: Ending variation in Group II preterite & 3rd sg subject

	980	1050	1100	1120	1130	1155	1177	1180	1190	1194	1200	1205	1212	1220	1223	1224	1225	1255	1270	1275	1279	1309	1330	1370	1373	1427	1440	1450	1498	1523	1527	1572	
zero	0	2	22	2	0	21	94	49	73	0	21	3	15	16	2	19	75	17	186	2	16	160	63	85	102	1	100	313	30	22	46	54	
-s	0	0	23	1	0	18	37	12	39	0	8	1	1	10	1	2	11	9	57	0	0	161	36	20	20	0	34	174	26	99	52	106	
-s	na	na	0	0.51	0.33	na	0.46	0.28	0.20	0.35	na	0.28	0.25	0.06	0.38	0.33	0.10	0.13	0.35	0.23	0	0	0.50	0.36	0.19	0.16	0	0.25	0.36	0.46	0.82	0.53	0.66

Table 26: Ending variation in Group II present indicative and preterite & 1st sg subject

	980	1050	1100	1120	1130	1155	1177	1180	1190	1194	1200	1205	1212	1220	1223	1224	1225	1255	1270	1275	1279	1309	1330	1370	1373	1427	1440	1450	1498	1523	1527	1577
zero	0	0	0	1	0	3	7	3	14	0	8	8	1	1	0	0	6	0	14	0	1	39	8	8	17	0	14	39	1	1	2	0
-s	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	6	2
-s	na	na	na	0	na	0	0	0	0	na	0	0	0	0	na	na	0	na	0	na	0	0	0	0	0	na	0	0	0.50	0	0.75	0

Table 27: Ending variation in Group I imperfect and future conditional & 1st sg subject

	980	1050	1100	1120	1130	1155	1177	1180	1190	1194	1200	1205	1212	1220	1223	1224	1225	1255	1270	1275	1279	1309	1330	1370	1373	1427	1440	1450	1498	1523	1527	1577
zero	0	2	5	1	0	2	22	5	21	0	6	3	0	1	0	0	38	2	52	0	0	93	19	18	19	0	34	69	7	4	11	0
-s	0	0	1	0	0	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	1	13	4
-s	na	0	0.17	0	na	0.50	0	0.17	0	na	0	0	na	0	na	na	0	0	0	na	na	0	0	0	0	na	0	0.04	0	0.20	0.54	0

Table 28: Ending variation in Group II imperfect and future conditional & 1st sg subject

	980	1050	1100	1120	1130	1155	1177	1180	1190	1194	1200	1205	1212	1220	1223	1224	1225	1255	1270	1275	1279	1309	1330	1370	1373	1427	1440	1450	1498	1523	1527	1577
zero	1	1	18	2	0	3	24	45	8	0	0	0	3	1	0	3	3	3	4	0	0	4	1	6	4	0	1	1	0	0	1	0
-e	0	1	2	2	0	1	1	0	2	0	0	0	0	0	0	1	1	0	0	0	0	1	6	12	0	0	2	26	1	0	10	3
-e	0	0.50	0.10	0.50	na	0.25	0.04	0	0.20	na	na	na	0	0	na	0.25	0.25	0	0	na	na	0.20	0.86	0.67	0	0	0.67	0.96	1	na	0.91	0

Table 29: Ending variation in Group I present indicative and subjunctive & 1st sg **null** subject

	980	1050	1100	1120	1130	1155	1177	1180	1190	1194	1200	1205	1212	1220	1223	1224	1225	1255	1270	1275	1279	1309	1330	1370	1373	1427	1440	1450	1498	1523	1527	1577
no	173	483	2555	1086	174	3704	3243	2711	3701	70	853	2830	721	2140	306	756	2936	539	4333	235	1206	5941	3621	3411	13399	321	2265	10547	2276	2313	3292	8
yes	8	14	115	68	39	477	798	394	380	2	159	123	109	70	50	145	600	67	483	36	412	742	393	647	846	85	624	1232	112	358	276	6
yes	0.04	0.03	0.04	0.06	0.18	0.11	0.20	0.13	0.09	0.03	0.16	0.04	0.13	0.03	0.14	0.16	0.17	0.11	0.10	0.13	0.25	0.11	0.10	0.16	0.06	0.21	0.22	0.10	0.05	0.13	0.08	0

Table 30: Ending ambiguity counts