Effects of text form on grammatical changes in Medieval French. A treebank-based diachronic study

Alexandra Simonenko[1], Benoit Crabbé[2], and Sophie Prévost[3]

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

[2]LLF/CNRS/Paris Diderot, USPC,

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

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

sophie.prevost@ens.fr

April 10, 2018

Address:

Alexandra Simonenko

Ghent University

Faculty of Arts and Humanities

Department of General Linguistics

Blandijnberg 2, B-9000 Ghent, Belgium

Abstract

This paper presents a corpus-based study of the effect the text form (prose vs. verse) has on the course of two grammatical changes in Medieval French: the loss of null subjects and the loss of OV word order. By means of statistical analysis, we demonstrate that naive estimates of the spread of overt subjects and VO orders give the impression that there is a significant difference between the rates of development in prose vs. verse. In contrast, estimates based on an abstract grammar competition model which distinguishes between grammar-ambiguous surface forms (overt personal subjects, null subjects in coordination contexts) and grammar-unambiguous surface forms (overt expletive subjects, null subjects in non-coordination contexts) show prose-verse parallelism, prose having an earlier change onset, in line with traditional intuitions. At a more general level, these results suggest that the product of the interaction of a particular grammar with universal pragmatic laws is constant, which can be observed if the factors responsible for variation in grammatical choices are controlled for.

Keywords: prose vs. verse in language change, Constant Rate Effect, null subjects, word order change, Medieval French

1 Introduction

This paper investigates the effects of the text form (prose vs. verse) on diachronic changes in Medieval French using the treebanks MCVF and the Penn Supplement to MCVF (≈1,5 million words, Penn scheme annotation).[1] Despite the common intuition that the prose is more "advanced" than the contemporary verse with respect to grammatical changes, in virtue of not being subject to the versification constraints, in the absence of statistical models based on large-scale corpora, the magnitude of the difference has remained unknown. Estimates for the decline of pro-drop based on smaller data samples strongly suggest that the distinction is indeed real (Prévost 2018). To estimate the prose-verse lag is especially important for studies modelling language evolution based on written sources. Grammatical factors influencing the speed of language change have to be disentangled from metagrammatical ones associated with conscious stylistic manipulations.

[1] Word counts are based on a version of the Penn Supplement available as of September 2017.

We estimate the prose-verse lag for different types of grammatical changes by means of statistical analysis. Specifically, we examine the trajectory of two changes: the decline of null subjects (morphosyntactic) and the shift from OV_{fin} to $V_{fin}O$ orders (syntactic) across text forms by modelling each change as evolution of a binary variable whose values correspond to competing grammars (Kroch (1989) and much subsequent work, see Pintzuk (2004)). That is, we estimate the effect of time on the probability that a finite clause has an overt pronominal subject (as opposed to a covert one), as well as the probability that a finite transitive clause with a nominal object exhibits VO (rather than OV) order.

The relevance of this work is threefold. First, it makes a methodological contribution to the study of language change by considering metagrammatical factors potentially affecting the rates of various grammatical changes. Estimating the rate of change has been central to a series of historical analyses pioneered by Kroch (1989), who first suggested that grammatical changes should be analysed not by directly comparing various data points, but by comparing the behaviour of well-understood mathematical functions fitted to relevant data sets. The Constant Rate Hypothesis of Kroch (1989) states that a grammatical change progresses at the same rate (or, more accurately, at not significantly different rates) in different *grammatical* contexts. The hypothesis relies on fitting logistic regression models to binary variables. It has been shown to hold for a number of grammatical changes across *grammatical* contexts and known as the Constant Rate Effect (see Pintzuk (2004) for an overview).[2] The hypothesis says nothing, however, about how changes spread across contexts which contrast in metagrammatical characteristics, such as prose vs. versified texts, and rightly so, since by definition such contrasts may be associated with conscious manipulations of linguistic features. This means that, to an extent, all bets are off as to what may happen to a given language change in text sources affected by such manipulations, such as versified textes. This study thus charts a new territory by means of a large-scale quantitative investigation of the effects of a metalinguistic distinction between prose and verse on the course of grammatical changes spanning the whole Medieval period. A major research question we addressed here is whether a grammatical change has the same trajectory across metalinguistically different environments. A statistical analysis relying on data from

[2]We are not aware of any counterexamples to the hypothesis, that is, developments of clearly the same nature proceeding at different rates in different grammatical contexts.

large (by the standards in historical linguistics) annotated corpora allows us to demonstrate that grammatical changes proceed in parallel ways in prose and verse, provided that strictly grammatical features are isolated from features susceptible to pragmatic/stylistic variation. In our case one such "volatile" feature is the use of subordinate clauses, which varies greatly (and in a temporally unstable way) between verse and prose. Our results are meant to be fully replicable: the full set of queries we used is given in the Appendix.

Second, this study paves the way for overcoming the issue of a text form/time correlation, which plagues historical research for empirical reasons. The empirical reality is such that for some periods verse may be the only or the dominant form in the available texts, which makes it crucial to understand its potential effects on the course of grammatical changes. For instance, the available body of Medieval French texts is characterised by the prevalence of versified texts until approximately the end of the XII c. It needs to be stressed that given the time/form correlation, the only way to estimate the effect of text form on linguistic changes is by means of statistical extrapolation, which, in turn, is only possible if we can estimate parameters of interest, such as the rate of null subjects, at time points for which we have data. Estimating those necessarily requires exhaustive annotation of text samples, which essentially amounts to using an existing annotated corpus or creating a new one. We do not see any other way which would allow us to make conclusions about the text form/time correlation.

Thirdly and finally, this project contributes to a better understanding of specific linguistic phenomena, that is, subject omission and word order, by examining their interaction with text forms. We get a better handle on factors governing these phenomena by relating them to the features which characterise a given text form.

In what follows we first consider the loss of null subjects, then we turn to the loss of the OV order (in finite clauses with a non-clitic direct object).

2 The loss of null subjects

We begin by considering the decline of subjectless finite clauses during the Medieval period across text forms. Early Medieval French (henceforth MF) is commonly recognised as being (at least partially) a pro-drop language, whereas late MF lost this

option completely except in cases of subject ellipsis under coordination. This change is well documented (Foulet 1928, Fontaine 1985, Hirschbühler 1992; Schøsler 2002; Kaiser 2009; Zimmermann 2014; Prévost 2018; Simonenko et al. 2018). We model it by estimating the distribution of the variable SUBJECT which takes the value *yes* if a clause has an overt personal pronominal subject and *no* otherwise in a sample including all finite clauses with either an overt personal pronominal or null subject (total of 76,150).[3] All clauses are tagged for the date of the manuscript they belong to. We fit these data to a logistic regression model $P(\text{Subject} = \text{yes} \mid \text{Date} = d) = \frac{e^{\alpha+\beta\, d}}{1+e^{\alpha+\beta\, d}}$ plotted in Figure 3.[4]. Parameter estimates of the model are given in tables 1 (prose), 2 (verse), and 3 (overall).

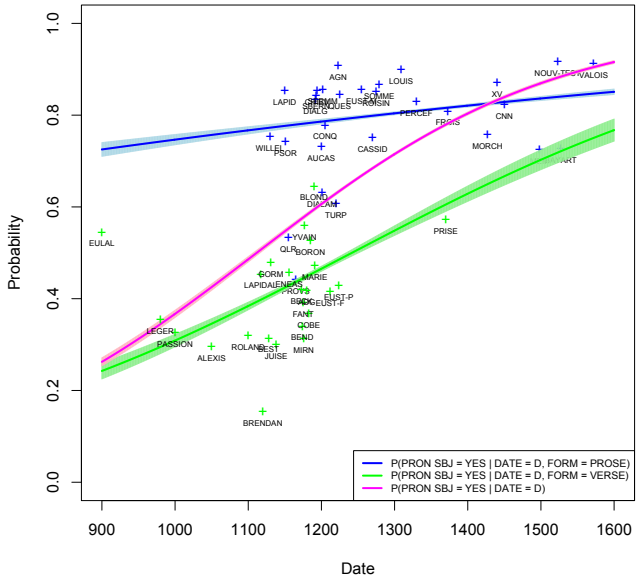

Figure 1: Overt subject emergence in prose and verse

	ESTIMATE	STD. ERROR	Z VALUE	PR(>|Z|)
INTERCEPT	−0.3562	0.1439	−2.474	0.01
COEFFICIENT	0.0016	0.0001	14.72	$<2\times10^{-16}$

Table 1: Logistic regression estimates for overt pronominal subjects in prose

[3]We excluded imperatives and wh-clauses targeting subjects because of their idiosyncratic subject syntax, as well as clauses introduced by connectives *et* ("and") and *si*. Connectives license subject ellipsis almost at a constant rate throughout the Medieval period as well as in Modern French, and therefore should not be considered as possible pro-drop environments. There is a handful of other conjunctive adverbs capable of licensing subject ellipsis, such as *puis*, but since those are much less frequent than *et* and *si*, we did not exclude them.

[4]We use logistic regression as is traditional for modelling historical data since Kroch (1989) (see also Kauhanen & Walkden (2017))

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	-4.6863	0.2226	-21.04	$<2\times10^{-16}$
coefficient	0.0038	0.0002	20.30	$<2\times10^{-16}$

Table 2: Logistic regression estimates for overt pronominal subjects in verse

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	-6.976×10	0.2226	-21.04	$<2\times10^{-16}$
coefficient	6.223×10^{-3}	7.701×10^{-5}	80.81	$<2\times10^{-16}$

Table 3: Logistic regression estimates for overt pronominal subjects overall

The greater intercept for prose indicates that the change manifested itself first in this text form, in line with the traditional intuition. Looking at the coefficients, we see, first, that the trends are rather different in verse and in prose, and, second, that prose is more advanced than verse in terms of the probability of a pronominal subject being overt throughout the medieval period. This contrast is not surprising in itself given that the prose/verse distinction is of a metalinguistic nature, so we have no *a priori* reasons to expect to find the Constant Rate Effect here. However, investigating what it is about prose that makes it favour overt subjects can be a fruitful line of inquiry since it can shed light on the grammar of null subjects. Interestingly, according to Walkden & Rusten (2016), during the Old English period which features the tailing off of the null subject decline, it is also verse that favours null subjects. Walkden & Rusten (2016, 27–28) conclude that "null subjects in O[ld] E[nglish] can be seen mainly as a feature of the poetry".[5] They suggest that metrical requirements imposed on versified texts could have favoured deletion of unstressed monosyllabic pronominals. They also quote Mitchell (1985, 992–993) who suggests that null subjects help poetry "to achieve compression and to give the poetry its characteristic texture". As a matter of speculation, we can say that subject (non)omission is a parameter which can be engaged for metrical purposes (adding or subtracting a syllable whenever needed).[6] However, this topic will have to await a focused quantitative study which would test whether (non)omission of pronominal subjects in verse was aligned with metrical requirements in a non-random way.

[5]Walkden & Rusten (2016, 27) show that in the earliest Old English texts the share of null subjects in verse is about 12%, as opposed to ca. 2% in prose.

[6]Old French and Old English poetry were both based on qualitative metre, the most widespread metres being iamb and trochee.

2.1 Abstract grammar-based analysis

Before concluding, however, that the emergence of overt pronominal subjects was happening at significantly different rates in verse and in prose, let us consider what these surface patterns mean in terms of grammatical shifts. Assuming a model of diachronic variation in terms of grammar competition (between two or more grammars), let us say that the replacement of null personal pronominal subjects by overt ones corresponded to the replacement of a grammar which had a structural component licensing null subject, such as an Agr(eement) head (Jelinek 1984, Barbosa 1995, Alexiadou & Anagnostopoulou 1998) by a grammar without such a head. Specifically, the output of the first grammar, let us call it the AgrP-Grammar, contained both null and overt personal pronominal subjects, but only null expletive subjects (as it is the case in modern incontestable pro-drop languages such as Italian). The output of the second, let us call it the TP-Grammar, had only overt subjects (in contexts not licensing subject ellipsis), be they personal pronominal or expletive. Thus the only subject type which can unambiguously be classified as belonging to the output of one grammar or another are expletive subjects. A null expletive corresponds to the AgrP-Grammar, an overt one – to the TP-Grammar. Moreover, because both grammars are, by hypothesis, categorical as to whether expletives are overt or null, we can expect that the (non)expression of expletives is entirely a function of the probability of a given grammar to be used at a given point in time and is not something a given speaker has control of once (s)he has chosen a generating grammar for a given illocutionary act. This means that while the expression of some personal pronominal subjects in verse could have been the result of metrical adjustments or other stylistic factors, with expletives this possibility is eliminated. We therefore model the spread of overt expletives only across prose and verse.

We fit finite clauses with either null or overt expletive subjects (total of 11,495) to the model $\mathrm{P}(\mathrm{Subject} = \mathrm{yes} \mid \mathrm{Date} = \mathrm{d}) = \frac{e^{\alpha+\beta\, d}}{1+e^{\alpha+\beta\, d}}$ plotted in Figure 2.

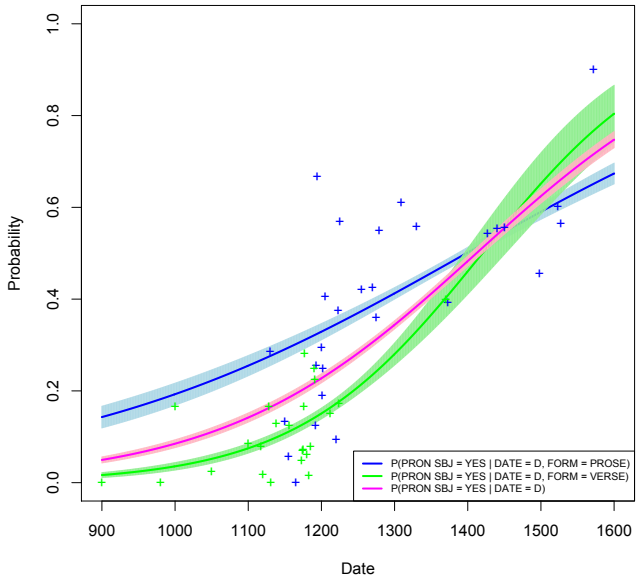

Figure 2: Overt expletive subject emergence in prose and verse

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	-5.0264	0.2842	-17.69	$<2\times10^{-16}$
coefficient	0.0036	0.0002	17.36	$<2\times10^{-16}$

Table 4: Logistic regression estimates for overt expletive pronominal subjects in prose

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	-1.115×10^{01}	6.377×10^{-01}	-17.48	$<2\times10^{-16}$
coefficient	7.851×10^{-03}	5.204×10^{-04}	15.09	$<2\times10^{-16}$

Table 5: Logistic regression estimates for overt expletive pronominal subjects in verse

What we observe in Figure 2 is a striking parallelism between verse and prose for the time period for which we have good confidence of estimation (approximately until around 1300). This confirms our grammar competition-based prediction that expletive expressions is "out of reach" for metalinguistic manipulations, since those presumably cannot override the boundaries of grammaticality. To quote Kroch (1989, 36), this shows "the controlling effect of abstract grammatical analyses on patterns in usage data". Specifically, an analysis in terms of grammatical options rather than in terms of direct surface forms allowed us to separate what appears to be a properly grammatical change from the effects of metalinguistic prose/verse distinction, even though the nature of the latter remains to be explained. We will see below that an abstract syntactic analysis has a similar clarifying effect for the disappearance of the OV order.

2.2 Direct vs. narrative discourse

Let us take yet another perspective and consider overt pronominal subject emergence in MF across discourse types, that is, direct vs. narrative. It is well established that these two registers differ quantitatively with respect to a number of grammatical characteristics (e.g. Dufter 2010 and references therein, Marchello-Nizia 2012, Lagorgette & Larrivée 2013, Guillot-Barbance et al. 2017, Glikman & Mazziotta 2013, Prévost 2018). The two types are illustrated in (1). Figure 3 visualises logistic regression models estimating the emergence of overt subjects (both personal and expletive) in direct discourse vs. narrative for verse and prose.

(1) Respondet l' altre: "Mal i diz."
responds the other bad there say
"The other one responds, 'You are wrong'." (1000-PASSION-BFM-P,113.216)

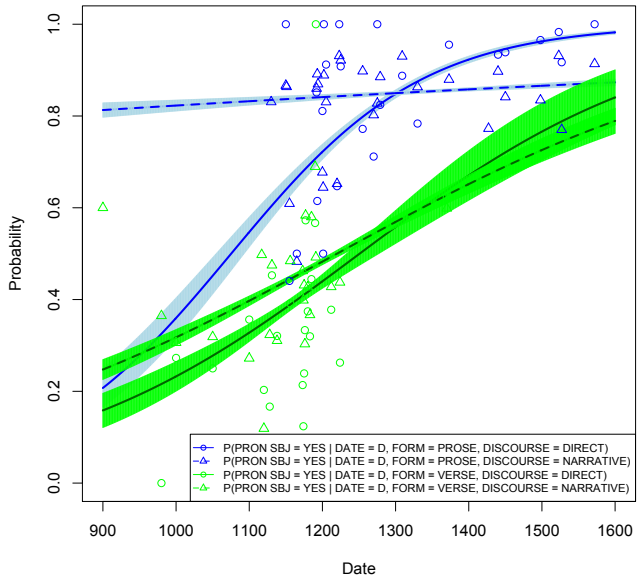

Figure 3: Overt subject emergence in prose and verse & direct and narrative discourse

The estimates of the logistic regression models are given in tables 6–9.

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	−8.2267	0.4258	−19.32	$<2\times10^{-16}$
coefficient	0.0076	0.0003	22.98	$<2\times10^{-16}$

Table 6: Logistic regression estimates for overt pronominal subjects in prose (direct discourse)

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	0.8758	0.1556	5.626	1.85×10^{-08}
coefficient	0.0006	0.0001	5.716	1.09×10^{-08}

Table 7: Logistic regression estimates for overt pronominal subjects in prose (narrative)

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	−8.1465	0.2187	−37.24	$<2\times10^{-16}$
coefficient	0.0057	0.0002	35.29	$<2\times10^{-16}$

Table 8: Logistic regression estimates for overt pronominal subjects in verse (direct discourse)

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	−5.9545	0.6172	−9.64	$<2\times10^{-16}$
coefficient	0.0047	0.0005	9.01	$<2\times10^{-16}$

Table 9: Logistic regression estimates for overt pronominal subjects in verse (narrative)

Focusing only on direct discourse only, the change appears to proceed in a parallel way in verse and in prose, prose expectedly being more advanced than verse. We also see that in verse there is virtually no difference between direct speech and narrative. A result which is more difficult to interpret is the virtual stability of the rate of pronominal subject expression in prose narrative, as opposed to prose direct discourse, where the change progresses along an expected curve. As a consequence, it looks as though until approximately the end of the XIII c., prose narrative is more advanced than direct discourse, and then the situation reverses. This contrasts with the results of Glikman & Mazziotta (2013, 77) who report more overt subjects in direct discourse (in a sample of clauses from one text). This difference in results, however, may be due to a methodological difference: we excluded subjects omitted under coordination, while Glikman & Mazziotta (2013) included them. This explanation is supported by the fact that in the sample examined by Glikman & Mazziotta (2013, 79) we find more connectives such as *et* "and" in narrative (and therefore more contexts for subject ellipsis) than in direct discourse. This methodological point aside, our result runs counter to the commonly accepted idea that direct speech is more advanced than narrative with respect to the progress of grammatical changes. It has been largely acknowledged that direct speech (whatever the state of a language is) displays linguistic features closer to spoken language than narrative does, although it cannot be strictly equated with the latter. Because linguistic changes are expected to be more advanced in

spoken language than in written language, it is expected that innovating features appear first in direct speech.

Recall that we ran into a similar issue with the rate of pronominal subject expression in prose in general (expletive and personal subjects and direct discourse and narrative combined) in section 2. One feature which potentially sets apart prose narrative from both prose direct discourse and verse (narrative and direct discourse) is the frequency of subordinate clauses, which are known to favour subject expression significantly more than matrix ones (Adams 1987; Franzén 1939; Foulet 1928; Hirschbühler 1992; Prévost 2018; Roberts 2014; Vance 1997; Zimmermann 2014, among others). If so, the apparently stable high rate of pronominal subject expression in prose narrative may be due to a larger share of subordinate clauses in prose narrative than in any other text form we have examined, and to the fact that the change comes to completion earlier in subordinate clauses. This hypothesis can be tested if we check for the relative frequency of subordinate clauses in different text forms.

	MATRIX	MATRIX QUESTIONS	SUBORDINATE
PROSE NARRATIVE	0.53 (56964)	0.00	0.47 (50831)
PROSE DIRECT DISCOUSE	0.82 (13466)	0.02 (319)	0.16 (2647)
VERSE NARRATIVE	0.61 (33615)	0.00	0.39 (21159)
VERSE DIRECT DISCOURSE	0.89 (11638)	0.01 (116)	0.11 (1385)

Table 10: Frequency of clause types across text forms

In order to further test for the influence of discourse type and text form on the rate of subordinate clauses, we run a logistic regression model on a dependent variable CLAUSE TYPE with the values *matrix* and *subordinate* (ignoring the very infrequent matrix questions) with the predictor variables FORM (*prose*, *verse*) and DISCOURSE TYPE (*narrative*, *direct*). As the summary of the model's parameters in table 11 shows, both predictors are highly significant, with narrative affecting the probability of a subordinate clause positively and verse negatively. That is, prose narrative comes out as the environment favouring subordinate clauses the most, which can explain the high rate of pronominal subject expression in this environment.

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	−0.9503	0.0218	−43.50	$<2\times10^{-16}$
Discourse type (narrative)	2.2958	0.0229	100.32	$<2\times10^{-16}$
Form (verse)	−0.5466	0.0175	−31.15	$<2\times10^{-16}$

Table 11: Logistic regression estimates for clause type

In figure 4 we plot models fitting the distribution of the variable CLAUSE (*matrix*, *subordinate*) in prose and verse. The rate of subordinate clauses appears to be increasing in verse.[7]

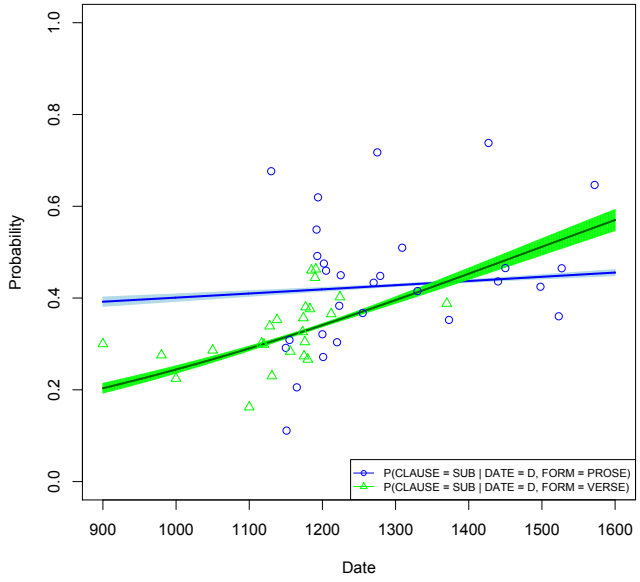

Figure 4: Overt subject emergence in prose and verse & direct and narrative discourse (matrix clauses)

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	-7.739×10^{-1}	6.542×10^{-2}	−11.83	$<2\times10^{-16}$
coefficient	3.722×10^{-4}	4.864×10^{-5}	7.65	1.97×10^{-14}

Table 12: Logistic regression estimates for subordinate clauses in prose

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	−3.481	0.136	−25.50	$<2\times10^{-16}$
coefficient	0.0023	0.0001	20.46	$<2\times10^{-16}$

Table 13: Logistic regression estimates for subordinate clauses in verse

In view of these results, les us focus our attention on matrix clauses alone. As figure 5 shows, if limited to this environment, the picture conforms to the traditional

[7]We cannot test for the significance of the difference between the model's coefficients in verse and prose due to insufficient data for verse in the later periods.

expectation of a faster change in environments approximating oral speech, that is, in direct discourse. Model estimates are given in tables 14–17.

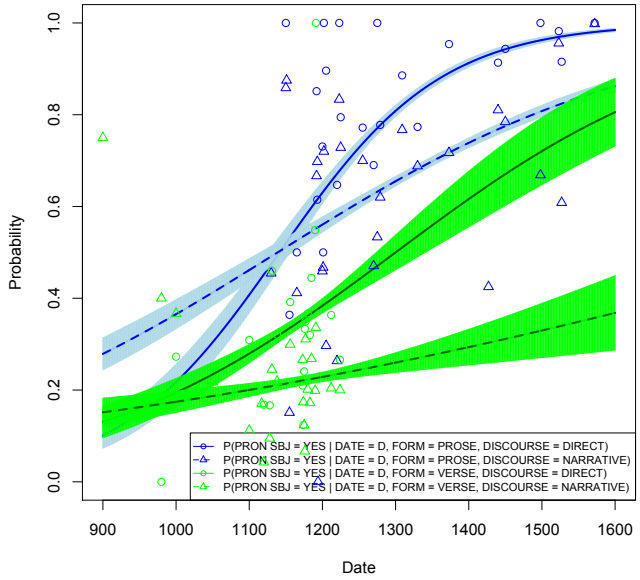

Figure 5: Overt subject emergence in prose and verse and direct and narrative discourse (matrix clauses)

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	-1.035×10	5.017×10^{-1}	-20.63	$<2\times10^{-16}$
coefficient	9.070×10^{-3}	3.879×10^{-4}	23.38	$<2\times10^{-16}$

Table 14: Logistic regression estimates for overt pronominal subjects in prose (direct discourse, matrix)

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	-4.5319	0.2704	-16.76	$<2\times10^{-16}$
coefficient	0.0039	0.0002	19.58	$<2\times10^{-16}$

Table 15: Logistic regression estimates for overt pronominal subjects in prose (narrative, matrix)

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	-6.1587	0.6605	-9.32	$<2\times10^{-16}$
coefficient	0.0047	0.0006	8.42	$<2\times10^{-16}$

Table 16: Logistic regression estimates for overt pronominal subjects in verse (direct discourse, matrix)

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	-3.2436	0.5023	-6.45	1.07×10^{-10}
coefficient	0.0017	0.0004	3.98	6.69×10^{-05}

Table 17: Logistic regression estimates for overt pronominal subjects in verse (narrative, matrix)

To make an interim summary, in this study of pro-drop across text forms we have first established that if surface forms are counted indiscriminately, that is, all kinds of

null subjects together and without distinguishing discourse types, prose appears to have a very different change profile, with pronominal subject expression rates being very high right from the earliest texts on. If not predicted, this is at least not surprising in the two-grammar competition model where the old grammar allows for overt subjects under some pragmatically defined conditions. This pragmatic flexibility can arguably be exploited differently in different text forms. Once we look at the data in which the output of the two grammars is assumed to be categorically distributed, namely, clauses with expletive subjects (i.e. always null for the old grammar and always overt for the new one), the prose/verse distinction virtually disappears, as predicted by our grammar competition model. Once pragmatic factors are excluded, we expect a grammatical parallelism between the two text forms. It turns out that another way to uncover this parallelism is to look at the environment which is assumed to approximate oral speech the most, direct discourse. We found out that the rates of overt pronominal subjects are similar for verse and prose in this environment. We concluded that the major source of non-parallelism in other contexts is the uneven distribution of subordinate clauses, known to favour subject expression. If limited to matrix clauses, the change develops in parallel ways across prose and verse in pragmatically similar environments (either narrative or direct discourse). We thus conclude that the influence of pragmatic factors on the change is stable across text forms (as manifested by the absence of dramatic differences between rates of change in matrix clauses) if we properly control for the grammatical environments with which these factors interact, such as the distinction between matrix and subordinate clauses.

3 OV_{fin} decline in prose vs. verse

Early Medieval French is known to have a greater word order flexibility than Modern French, in particular, in allowing for both $V_{fin}O$ and OV_{fin}, the latter option disappearing with time (Marchello-Nizia 1995; Vance 1997; Labelle & Hirschbühler 2005; Labelle 2007; Zaring 2011; Marchello-Nizia & Rouquier 2012; Kroch & Santorini 2014). Examples below illustrate the OV_{fin} option unavailable in Modern French.

(2) $[\text{lei}]_{obj}$ consentit_{v} et observat_{v}
law agreed and observed

'he respected and observed the law' (0980-LEGER-V,XII.82)

(3) [Ja mais]$_{adv}$ [ledece]$_{obj}$ n'avrai$_{v}$
never joy won't.have
'I will never have joy' (10XX-ALEXIS-V,99.892)

(4) [Li quens Rollant]$_{sbj}$ [Gualter de l' Hum]$_{obj}$ apelet$_{v}$
the king Roland Walter of the Hum called
'The king Roland called Walter of Hum' (1100-ROLAND-V,65.779)

In what follows we will examine the effects of the verse/prose distinction on how this change proceeded.

3.1 From OV to VO: simple estimates

We first model this change by estimating the distribution of the variable $V_{fin}O$ (with the values *yes* and *no*) in a set of finite clauses with non-clitic direct objects excluding imperatives and wh-clauses targeting subject or object (total of 40,120). Some studies focus on tracking specifically base-generated OV orders. For instance, Kroch & Santorini (2014) in their study of the OV decline take into account only some non-finite clauses and exclude cases where the VO order could have been generated from OV by V-to-T or V-to-C movement. In contrast, we are examining the loss of object movement to the left-periphery, that is, to the left of the finite verb, assuming that a finite verb is at least as high as T.[8] That is, disregarding the question about the headedness of the VP, we suggest that the "old" grammar, inherited from Late Latin, allowed movement of direct objects to the clausal left periphery, while the new grammar that eventually took over did not allow for this sort of movement and generated only VO sequences.[9] We also assume that the old grammar could generate VO, or V1 ("verb-first"), orders in those cases where the verb moved higher than any of the arguments. This order is illustrated in (5). We assume, for now, that the old grammar generated such orders at some constant rate associated with a particular set of pragmatic conditions.[10] This assumption will be important in the discussion since a VO string is ambiguous as to

[8]Interestingly, though, the progression of $V_{non-fin}O$ reported in Kroch & Santorini (2014) and $V_{fin}O$ presented here turns out to be quite similar if we consider prose and verse combined.

[9]However, there seems to be no reasons to assume that OV was a predominant order even in Late Latin (e.g. Passarotti et al. (2015) and references therein).

[10]An underlying assumption here is that the product of the interaction between a given grammar and universal pragmatic laws is constant in the absence of external perturbing factors.

which grammar generated it.

(5) Baisset sun chef,
lowered his head
"He lowered his head." (1100-ROLAND-V,9.112)

For now let us abstract away from the exact structural positions of the arguments and simply look at the distribution of OV/VO sequences over time.

Figure 6 visualises logistic regression models of the $V_{fin}O$ variable for prose, verse, and the two forms combined. The slope of the model corresponds to the rate of replacement of the old grammar by the new one, assuming that the new grammar generated only VO, while the old one generated OV plus (a constant rate of) VO.[11] Since the rate of "old" VO, by assumption, is constant, it should not matter for the slope comparison.

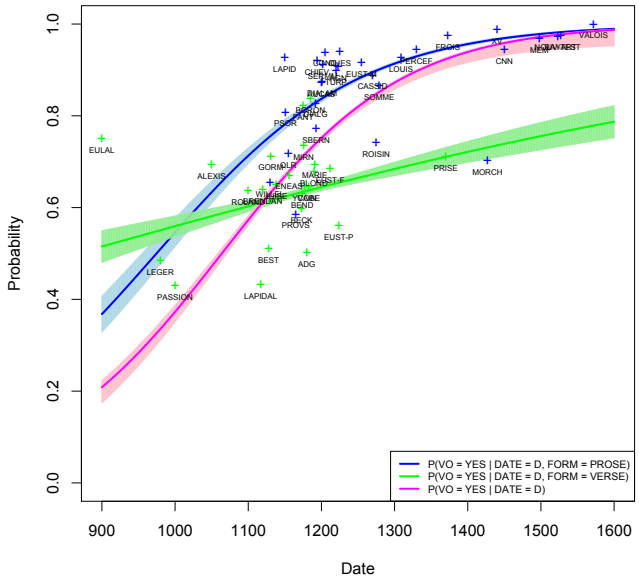

Figure 6: $OV_{fin} \rightarrow V_{fin}O$ in prose and verse

The parameter estimates for our model $P(V_{fin}O = \text{yes} \mid \text{Date} = d) = \frac{e^{\alpha+\beta\, d}}{1+e^{\alpha+\beta\, d}}$ are given in tables 18-20.

[11] Modern French makes use of the OV order under very restricted conditions discussed in Abeillé et al. (2008).

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	−7.0739	0.2963	−23.87	$<2\times10^{-16}$
coefficient	0.0072	0.0002	30.92	$<2\times10^{-16}$

Table 18: Logistic regression estimates for $V_{fin}O$ in prose

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	−1.5419	0.2952	−5.22	1.76×10^{-7}
coefficient	0.0017	0.0003	7.10	1.22×10^{-12}

Table 19: Logistic regression estimates for $V_{fin}O$ in verse

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	−8.6574	0.1796	−48.19	$<2\times10^{-16}$
coefficient	0.0081	0.0001	54.83	$<2\times10^{-16}$

Table 20: Logistic regression estimates for $V_{fin}O$ overall

While the rate of OV_{fin} for verse is almost constant over time (coefficient close to zero), prose appears to be more advanced than verse in the transition to $V_{fin}O$, at least during the XII c. for which we have data points for both prose and verse (although verse is still better represented). One way to interpret the logistic regression parameters we obtained is to say that the temporal "window" available for verse is such that we cannot really observe the decline of OV_{fin} in verse. This would be due to a problem of the corpus text sample, since we know for sure that OV_{fin} ends up disappearing almost completely even from verse.

3.2 Abstract grammar-based analysis

Before we concede, however, that there is an unsurmountable data sampling problem responsible for the difference or that there is actually a significant difference between the rates of change in prose and in verse during the available time window, let us consider another analytical possibility. Recall that our calculations of the rate of change from the "old" OV to the "new" VO order involved an assumption that even though both grammars can generate VO, we count all VO as "new" assuming that those that are generated by the old grammar (as V1 configurations) constitute a fixed proportion in the overall output of the old grammar at any given point. Thus miscounting them as produced by the new grammar does not affect the rate of the spread of the innovative

grammar. That is, even though, because of the added VO counts, the new grammar's probabilities would be "bumped" up at any given time point, this bumping up would be a constant over the whole Medieval period and independent of the prose/verse distinction, and thus could be neglected for the purposes of comparing the overall rates of change in prose and verse. However, if this assumption is wrong, that is, if for some reason the bumping-up effect varies depending on the text form and/or time, this could be a source of non-parallelism between prose and verse. In what follows we show that the original assumption is indeed problematic and that we do need to sort out VOs. The main culprit will turns out to be the VO orders with "true" pro-drop (that is, not cases of ellipsis under coordination), because (non-expletive) pro-drop rates vary depending on the prose/verse distinction (as we showed in section 2).

To discuss the possible effect of grammar-ambiguous VOs, we need to be more specific about what kind of competing grammars we assume and what orders they can generate, including the position of the subject.

3.2.1 Grammar A ("old")

Table 21 gives an overview of the evolution of word order in transitive finite clauses with non-clitic objects.[12] The general obvious trend is the steady increase in SVO at the expense of all other permutations. Another immediate observation is the rarity of OSV and VOS orders, which we will therefore exclude from detailed examination.

	OSV	OV	OVS	SOV	SVO	VO	VOS	VSO
1100	0.00 (6)	0.26 (411)	0.08 (127)	0.04 (65)	0.19 (306)	0.40 (649)	0.00 (6)	0.02 (39)
1200	0.00 (60)	0.22 (3756)	0.05 (923)	0.05 (860)	0.32 (5487)	0.29 (4852)	0.01 (175)	0.05 (879)
1300	0.00 (25)	0.06 (343)	0.04 (225)	0.02 (128)	0.50 (2837)	0.28 (1598)	0.01 (36)	0.09 (515)
1400	0.01 (60)	0.05 (390)	0.02 (153)	0.01 (80)	0.56 (4749)	0.28 (2382)	0.01 (62)	0.07 (553)
1500	0.01 (28)	0.02 (92)	0.02 (100)	0.01 (32)	0.66 (3225)	0.25 (1208)	0.00 (19)	0.04 (193)
1600	0.00 (6)	0.01 (18)	0.01 (26)	0.00 (1)	0.74 (1829)	0.21 (516)	0.00 (9)	0.03 (81)

Table 21: Word order evolution in transitive clauses with non-clitic objects

A note is in order considering the scope of this investigation. The (evolution of)

[12]The reason we excluded pronominal clitic objects is that their syntax even in the earliest texts is already that of verbal clitics, meaning that they are much more syntactically constrained compared to nominal arguments, whereas pronominal subjects do not entirely cliticize until later. Specifically, the position of non-emphatic object pronominals is strictly dependent on the position of the verb: they immediately precede the verb if the verb is not clause-final, and they immediately follow it when the verb is clause-initial, a generalization known as the Tobler-Mussafia law. For a detailed corpus-based study of the syntax of object clitics see Simonenko & Hirschbühler (2012). We also excluded clauses with subject or object wh-dependency because of their idiosyncratic argument syntax.

clausal structure in Medieval French has been the subject of much attention in the literature (Vennemann 1974, Harris 1978, Fleischman 1992, Roberts 1993, Marchello-Nizia 1995, Vance 1997, Lafond 2003, Labelle & Hirschbühler 2005, Rouveret 2004, Mathieu 2006, Labelle 2007, Zaring 2011, Simonenko & Hirschbühler 2012, Kroch & Santorini 2014 to name just a few). Our focus here is limited to the disappearance of pre-verbal non-clitic objects in transitive finite clauses, and we are concerned only with the position of the main arguments. Most importantly, we are interested in how this change manifested itself depending on the text form, a topic which has not been yet explored at all in a systematic fashion, as far as we know. That is, such issues as the (un)availability of V3 in Old French, the syntax of different subordinate clauses and matrix and embedded questions and many others puzzles of the MF syntax are left out of the present picture.

We assume that the old grammar is characterised by an articulated left-periphery which involves an agreement projection, Agreement Phrase (AgrP), as well as (at least) two information structure-related projections, Focus Phrase (FocP) and Topic Phrase (TopP). In the following we briefly discuss our assumptions concerning the structures underlying each surface order.[13] Our eventual goal is to be able to classify as many surface strings as possible as generated by the old or by the new grammar, in order to track the disappearance of the OV-generating old grammar across text forms.

OVS. We begin our inventory of the configurations made available by the old grammar with OVS, (6). We assume that OVS corresponds to the object and subject placement in the Specifier of a discourse-oriented functional projection Topic Phrase (SpecTopP) and the Specifier of the Tense Phrase (SpecTP) respectively. This is accompanied by the movement of the verbal (complex) head to the functional head Agr(eement), which hosts subject person and number features, as in figure 7.

(6) [Messe e matines]$_{obj}$ ad [li reis]$_{sbj}$ escultet.
mass and matines has the king attended
"The king has attended mass and matines." (1100-ROLAND-V,54.647)

[13] We abstract away from fine details of the structure below the TP level, such as the presence of modal, aspectual, and agent-introducing projections. Triangular brackets indicate movement traces and regular brackets – the possibility of argument omission.

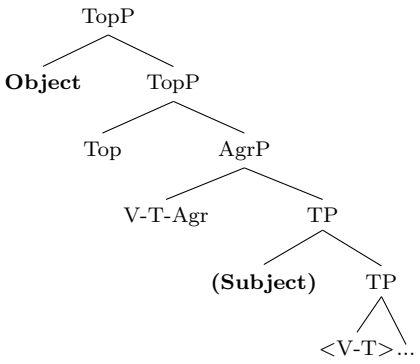

Figure 7: Grammar A generating OVS

Obviously, in order to test for the adequacy of this representation, we cannot probe directly into the information structure of MF by gathering speakers' judgements. However, as a proxy, we can look at the formal properties of the noun phrases involved, such as the presence/absence and semantics of determiners and modifiers, on the assumption that determiner types correlate with the information-structure statuses of arguments. Specifically, a number of determiners, such as definite and possessive ones, are commonly assumed to trigger presuppositions, that is, contraints on what kind of information a context should entail in order for the utterance in question to be felicitous in that context.

Table 26 in the Appendix gives the distribution of head types in direct object noun phrases in OVS configurations, and table 27 presents the distribution of the determiner types with nominal objects. We put the adjective *tel* 'such' in a separate category because of its frequency and special semantics. Noun phrases with such modifiers normally have an antecedent, and therefore can be assumed to be demonstrative-like.

Below we will compare these results with the determiner distribution in other syntactic configurations and show that this is a remarkably high incidence of demonstratives, both as heads and as pre-nominal determiners. Simonenko (2017) provides a semantic argument as well as arguments from synchronic studies that demonstratives are very likely to be shifted topics and that the position in question was likely associated with prosodic prominence (see also Rainsford (2011, 216) for Medieval French). This corresponds to the Top label of the relevant head in Figure 7.

SOV and OV. Another eventually disappearing configuration is SOV, (7), for which we assume the structure illustrated in Figure 8 where the subject and the object occupy the SpecTopP and the SpecFocP respectively.

(7) [Li reis Marsilie]$_{sbj}$ [le poign destre]$_{obj}$ i perdiet
the king Marsile the fist right there lost
"The king Marsile lost there his right fist.' (1100-ROLAND-V,200.2782)

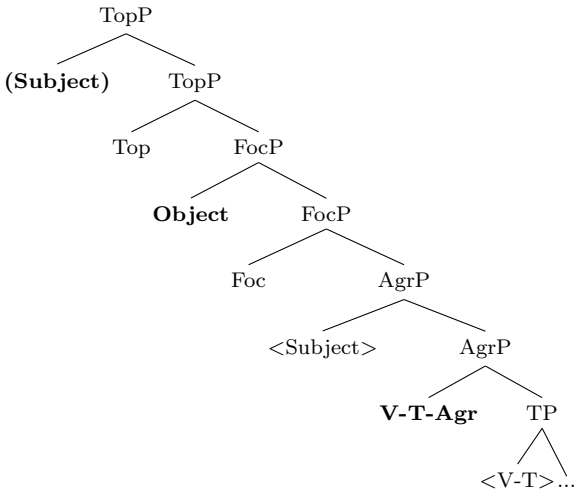

Figure 8: Grammar A generating SOV

Head types and determiner types with nominal objects in this configuration are distributed as in Tables 28 and 29 (Appendix) respectively, where we find a much lower rate of demonstratives than in OVS configurations.[14]

Under this configuration we subsume OV orders with null subjects. Specifically, we assume that if they are not contrastive, subjects are null in the old grammar. The distribution of head types and determiners in object phrases is remarkably similar in SOV and OV configurations, as a comparison between tables 30–31 (Appendix) on the one hand and tables 28–29 on the other shows. In fact, if we exclude full object pronouns, the difference in the distribution of the other heads types between OV and SOV is not statistically significant at the 0.05 significance threshold ($\chi^2 = 3.57$, df = 3, p = 0.31).[15] Another observation which suggests that OV and SOV should be grouped together in terms of clause structure is a similar rate of possessive determiners, which in both cases is much higher than in OVS orders. This can be viewed as a consequence of the requirement that a possessive pronoun co-indexed with the subject be c-commanded by the latter. Finally, the rate of object pronominalization is significantly higher in SOV than in OV, or, in other words, when the subject is overt, the object is more likely to be

[14]The difference is highly statistically significant at the 0.05 significance level ($\chi^2 = 84.6$, df = 1, p = 3.53×10^{-20}).

[15]We had to remove free relatives from consideration because the number of observations is too small.

pronominal.[16] Recall, however, that these are non-clitic objects, which means that they were most likely contrastively focused (otherwise a clitic variant would have been chosen), which is reflected in their position in figure 8.

VSO. We assume that VSO orders, as in (8), have an in-situ object inside of VP and a subject in the canonical subject position in the Specifier of TP, as in Figure 9. The Specifier of the TopP in such configuration is occupied by an indirect object or a non-argument constituent.

(8) De Guenelun atent [li reis]$_{sbj}$ [nuveles]$_{obj}$...
from Guenelun awaits the king news
"The king awaits news from Guenelun..." (1100-ROLAND-V,53.642)

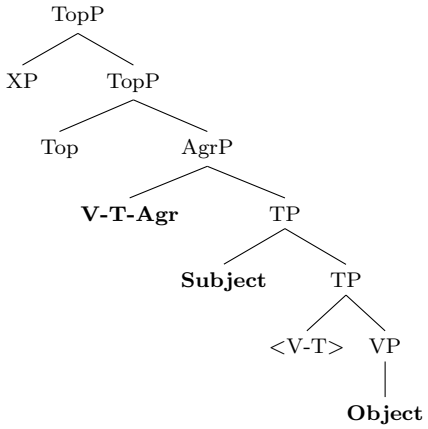

Figure 9: Grammar A generating VSO

The distribution of determiners with the objects is given in tables 32 and 33 (Appendix).

SVO and VO. Finally, for the old Grammar A, let us consider the pair SVO and VO. As far as Grammar A goes, we assume that these orders resulted from a structure as in Figure 10. An overt subject occupes the Specifier of the Topic projection. The SVO string, however, is ambiguous, as it could also be the output of the new grammar, as will be illustrated in section 3.2.2. In our estimates of the disappearance of OV we will not try to disambiguate SVO and will count them all as the output of the new grammar.

[16]This can be related (at least in cases where the subject is nominal) to the first Preferred Argument Structure constraint identified by Du Bois (2003, 34): "Avoid more than one lexical core argument".

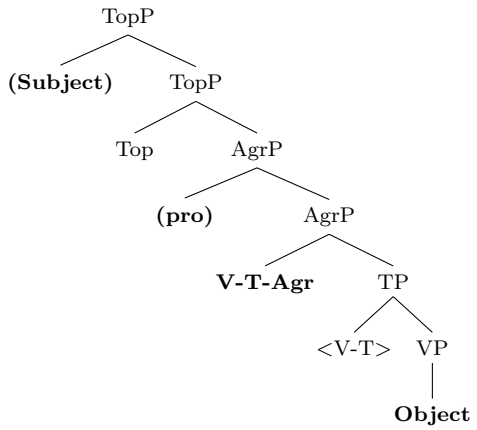

Figure 10: Grammar A generating SVO & VO

3.2.2 Grammar B ("new")

We assume that in contrast to Grammar A, Grammar B lacks an articulated left-periphery and an agreement head. It is also characterised by obligatory subject expression, the subject by default occupying SpecTP. It is well established that Medieval French underwent verbal agreement syncretisation (Bettens 2015; Buridant 2000; Dees et al. 1980; Foulet 1935; De Jong 2006; Marchello-Nizia 1992; Morin 2001; Simonenko et al. 2018). As a result, Modern French finite verbs do not distinguish between 1st, 2nd, and 3rd person singular in the present indicative. The only subject-less (non-imperative) finite clauses Grammar B generates are those where the subject is elided under coordination with the preceding clause, just like in Modern French. A simple declarative clause with a transitive predicate could thus be schematized as in Figure 11.

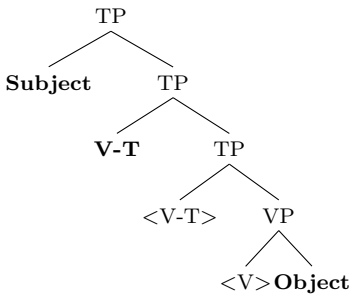

Figure 11: Grammar B generating SVO

3.3 Transition from Grammar A to B

We now classify all strings as generated by Grammar A or B. Seeing the loss of OV as resulting from the loss of a grammar with an extended left-periphery is in line with the

tradition analysing word order changes in Medieval French as reflecting a transition from Topic-initial to Subject-initial utterance organization (Vennemann 1974, Harris 1978, Marchello-Nizia 1995, 100). Similarly, Labelle & Hirschbühler (2005) suggested that during the medieval period French lost an information structure-related projection in the clausal left-periphery.

String type	Generating grammar
OVS	Grammar A
SOV & OV	Grammar A
VOS	Grammar A
VSO	Grammar A
"true" VO (i.e. subject omitted not under coordination)	Grammar A
"false" VO (i.e. subject omitted under coordination)	Grammar A or B
SVO	Grammar A or B

We are now in a position to model the transition from a grammar with a rich left periphery to a grammar without one as the distribution of a binary variable Grammar with values A and B, where all OVS, SOV, OV, VOS, VSO, and true VO are classified as Type A and all SVO as Type B, false VO being excluded from consideration.[17] We fit the following logistic regression model to our data: P(Grammar = B | Date = d, Form = f) = $\frac{e^{\alpha+\beta\, d}}{1+e^{\alpha+\beta\, d}}$ and the result is visualised in figure 12.

[17]We count all SVO as generated by Grammar B on the assumption that Grammar A generated such strings at a rate which was stable both across time periods and across text forms.

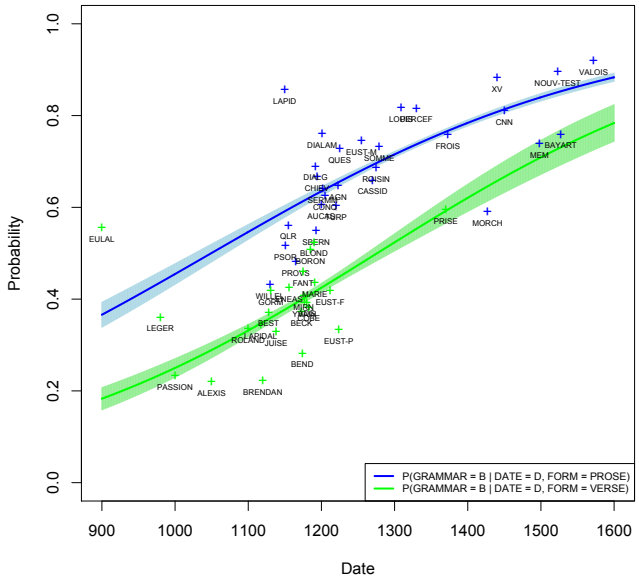

Figure 12: Passage from Type A to Type B Grammar

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	−3.860	0.1870	−20.64	$<2\times10^{-16}$
coefficient	0.0036	0.0001	25.87	$<2\times10^{-16}$

Table 22: Logistic regression estimates for the spread of Grammar B in prose

	Estimate	Std. Error	z value	Pr(>|z|)
intercept	−5.0760	0.3393	−14.96	$<2\times10^{-16}$
coefficient	0.0039	0.0003	13.89	$<2\times10^{-16}$

Table 23: Logistic regression estimates for the spread of Grammar B in verse

Estimated this way, the passage to the SVO grammar proceeds in parallel in prose and verse. An explicit testing for the difference in the slope parameter by means of a comparison based on an analysis of deviance of a model where the coefficient parameter can vary depending on the prose/verse distinction with one where the coefficient is not sensitive to these contexts reveals that the prose/verse parameter does not significantly contribute to better predict the data ($\chi^2 = 0.93$, df $= 2$, p $= 0.62$). In other words, the distinction between the rate of change in verse and in prose is not statistically significant.

The main change in our estimates compared to just tracking the distribution of OV/VO orders, as in section 3.1, is counting true VO as belonging to the same grammar as OV. Before doing a more fine-grained investigation of word-order changes, we speculated that if the rate of true VO were different in verse and in prose because of the

difference in the null subject rate, this would have affected our simplistic estimates of the passage from OV to VO, since the rate of the latter would have been disproportionately bumped up in verse. In contrast, classifying true VO as generated by the old Type A grammar results in the rates of change being now very similar across text forms. In other words, given that null subjects are more frequent in verse (see Figure 3), counting all VO including the true subjectless ones as generated by Grammar B led to an overestimation of the probability of the latter in verse in the first periods, where null subjects were still very frequent. Comparison between tables 24 and 25 (the only centuries for which there is enough data in both text forms are considered) makes it obvious that the main difference between prose and verse is the relative frequency of true OV and VO orders: the frequency of these orders is higher in verse, but it drops in the 14 c. This is consistent with what we know about the decline of null subjects (and thus true OV and VO), and this is, we suggest, the source of the non-parallelism between prose and verse we initially observed in Figure 6.

	OSV	**OV**	OVS	SOV	SVO	**VO**	VOS	VSO
1200	0.00	0.04	0.09	0.08	0.58	0.08	0.02	0.11
1300	0.01	0.02	0.05	0.02	0.67	0.09	0.01	0.13
1400	0.01	0.01	0.02	0.00	0.79	0.06	0.01	0.10

Table 24: Word order in transitive clauses with non-clitic objects in prose

	OSV	**OV**	OVS	SOV	SVO	**VO**	VOS	VSO
1200	0.01	0.18	0.07	0.07	0.40	0.20	0.01	0.06
1300	0.01	0.23	0.05	0.08	0.37	0.22	0.01	0.04
1400	0.00	0.15	0.02	0.07	0.60	0.13	0.00	0.03

Table 25: Word orders in transitive clauses with non-clitic objects in verse

4 Conclusions

We have examined two changes affecting different components of the Medieval French grammar across two text forms, prose and verse. First, we quantified the changes as variation in two surface forms, an "old" and a "new" variant. For the change in subject expression that meant quantifying occurrences of null vs. overt subjects, and for the

word order change – quantifying instances of OV_{fin} vs. $V_{fin}O$ orders. In both cases this approach revealed a puzzling non-parallelism between verse and prose, namely, either prose or verse would appear almost to stagnate across the Medieval period. This does not go well with the obvious fact that in both text forms all changes came to completion much earlier than today's French.

We then switched from estimating surface form competition to a more abstract modelling of variation as a competition between two grammars for which we assumed a certain mapping between abstract representations and surface forms. For the null subjects case we assumed an old grammar which could generate both null and overt personal pronominal subjects and a new one generating only overt ones. On these assumptions only expletive subjects unambiguously signalled which grammar was used. Estimated as the variation in null/overt expletives, the change progresses in a parallel fashion in prose and verse. These results suggest that in the grammar allowing for null subjects (the old grammar), the expression of *personal* pronominal subjects depends on the text form and, therefore, is not subject to strict grammatical constraints. This is a welcome result given that in modern null subject languages the conditions on the use of overt personal subjects are commonly defined in information-structural or pragmatic terms (e.g. aboutness-shift in Italian and Spanish, Frascarelli (2007) and Jiménez-Fernández (2016) respectively), and that the structuring of discourse depends largely on how the speaker chooses to relate a semantic representation to the utterance context. We further discovered that a major grammatical factor influencing such pragmatic choices is the clause type, matrix vs. subordinate: once we control for it, we see a prose/verse parallelism in the emergence of the overt personal pronominal subject. This suggests that pragmatic factors interact with grammatical choices in a stable way across time, which may be interpreted as an indication of the universality of pragmatic reasoning.

We also discovered that the difference in personal pronominal subject expression between prose and verse had repercussions for the estimation of the loss of the OV order as a simple competition between OV and VO. A higher rate of null subjects in verse resulted in what seemed like a very early dominance of VO. Once recast in terms of abstract grammars whereby the old grammar could generate subjectless VO sentences (and other argument permutations) and the new one only SVO, we once again see

parallel changes in prose and verse.

This project demonstrates that suitably large treebanks make it possible to engage tools of statistical analysis to test some of the traditionally accepted impressionistic and/or intuitive claims in the literature, thus strengthening the empirical basis of the field.

References

Abeillé, Anne, Danièle Godard & Frédéric Sabio. 2008. Deux constructions à SN antéposé en français. In Jacques Durand, Benoît Habert & Bernard Laks (eds.), *Congrès mondial de linguistique française (CMLF08)*, Paris: Institut de Linguistique Française.

Adams, Marianne. 1987. From Old French to the theory of pro-drop. *Natural Language & Linguistic Theory* 5(1). 1–32.

Alexiadou, Artemis & Elena Anagnostopoulou. 1998. Parametrizing AGR: Word order, V-movement and EPP-checking. *Natural Language & Linguistic Theory* 16(3). 491–539.

Barbosa, Pilar. 1995. *Null subjects*: University of Massachusetts, Amherst dissertation.

Bettens, Olivier. 2015. Chantez-vous français ? Remarques curieuses sur le français chanté de Moyen Âge à la période baroque. `http://virga.org/cvf/`.

Buridant, Claude. 2000. *Nouvelle grammaire de l'ancien français*. Paris: Sedes.

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, Steintje Meilink, Karin van Reenen-Stein & 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.

Du Bois, John W. 2003. Argument structure: Grammar in use. In J. D. Du Bois (ed.), *Preferred argument structure. grammar as architecture for function*, 11–60. Amsterdam: John Benjamins.

Dufter, Andreas. 2010. Subordination et expression du sujet en ancien français. In *Actes du XXVe Congrès International de Linguistique et de Philologie Romanes (Innsbruck, 3-8 septembre 2007)*, vol. 2, 443–458.

Fleischman, Suzanne. 1992. Discourse and diachrony: The rise and fall of old french *si*. In Marinel Gerritsen & Dieter Stein (eds.), *Internal and external factors in syntactic change*, 433–73. Berlin: Walter de Gruyter.

Fontaine, Carmen. 1985. *Application de méthodes quantitatives en diachronie: L'inversion du sujet en français*. Université du Québec à Montéal MA thesis.

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*: Uppsala dissertation.

Frascarelli, Mara. 2007. Subjects, topics and the interpretation of referential pro. *Natural Language & Linguistic Theory* 25(4). 691–734.

Glikman, Julie & Nicolas Mazziotta. 2013. Répresentation de l'oral et syntaxe dans la prose du *Queste del saint Graal (1225-1230)*. In Dominique Lagorgette & Pierre Larrivée (eds.), *Représentations du sens linguistique 5*, 69–90. Université de Savoie.

Guillot-Barbance, Céline, Bénédicte Pincemin & Alexei Lavrentiev. 2017. Représentation de l'oral en français médiéval et genres textuels. In *Autour du clivage langue parlée/langue écrite du latin au français*, Clermont-Ferrand, France.

Harris, Martin. 1978. *The evolution of French syntax: A comparative approach*. Longman Publishing Group.

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.

Ingason, Anton Karl. 2016. PaCQL: A new type of treebank search for the digital humanities. *Italian Journal of Computational Linguistics* 2(2). 51–66.

Jelinek, Eloise. 1984. Empty categories, case, and configurationality. *Natural Language & Linguistic Theory* 2(1). 39–76.

Jiménez-Fernández, Ángel. 2016. When discourse met null subjects. *Borealis-An International Journal of Hispanic Linguistics* 5(2). 173–189.

Kaiser, Georg A. 2009. Losing the null subject. A contrastive study of (Brazilian) Portuguese and (Medieval) French. In *Null-subjects, expletives, and locatives in Romance*, 131–156.

Kauhanen, Henri & George Walkden. 2017. Deriving the Constant Rate Effect. *Natural Language and Linguistic Theory* 1–39.

Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1(03). 199–244.

Kroch, Anthony & Beatrice Santorini. 2014. On the word order of Early Old French. Handout for a talk at SinFonIJA 7, Graz, Austria.

Labelle, Marie. 2007. Clausal architecture in Early Old French. *Lingua* 117(1). 289–316.

Labelle, Marie & Paul Hirschbühler. 2005. Changes in clausal organization and the position of clitics in Old French. In Montserrat Batllori, Maria-Lluïsa Hernanz, Carme Picallo & Francesc Roca (eds.), *Grammaticalization and parametric variation*, 60–71. Oxford: Oxford University Press.

Lafond, Larry. 2003. Historical changes in verb-second and null subjects from Old to Modern French. In D. Eric Holt (ed.), *Optimality Theory and language change*, 387–412. Dordrecht: Kluwer Academic Publishers.

Lagorgette, Dominique & Pierre Larrivée. 2013. *Représentations du sens linguistique 5*. Université de Savoie.

Marchello-Nizia, Christiane. 1992. *Histoire de la langue française aux XIVe et XVe siècles*. Paris: Dunod.

Marchello-Nizia, Christiane. 1995. *L'évolution du français. Ordre des mots, démonstratifs, accent tonique*. Paris: Armand Colin.

Marchello-Nizia, Christiane. 2012. L'oral représenté: un accès construit à une face cachée des langues 'mortes'. In Bernard Combettes, Céline Guillot, Evelyne Oppermann-Marsaux, Sophie Prévost & Amalia Rodrígez Samolinos (eds.), *Le changement en français. Etudes de linguistique diachronique*, 247–264. Bern: Peter Lang.

Marchello-Nizia, Christiane & Magali Rouquier. 2012. De (S)OV à SVO en français: où et quand? L'ordre des constituants propositionnels dans la Passion de Clermont et la Vie de saint Alexis. In Monique Dufresne (ed.), *Constructions en changement. Hommage à Paul Hirschbühler*, 111–155. Presses de l'Université de Laval.

Mathieu, Éric. 2006. Stylistic Fronting in Old French. *Probus* 18(2). 219–266.

Mathieu, Éric. 2009. From local blocking to cyclic agree. In Jila Ghomeshi, Ileana Paul & Martina Wiltschko (eds.), *Determiners: Universals and variations*, vol. 147, 123–158. John Benjamins Publishing.

MCVF. 2010. Corpus MCVF annoté syntaxiquement, dirigé par France Martineau, avec Paul Hirschbühler, Anthony Kroch et Yves Charles Morin. `http://www.voies.uottawa.ca/corpus_pg_en.html`.

Mitchell, Bruce. 1985. Old English syntax.

Morin, Yves Charles. 2001. La troncation des radicaux verbaux en français depuis le Moyen Age. Études diachroniques. *Recherches Linguistiques de Vincennes* (30). 63–85.

Passarotti, Marco, Barbara McGillivray & David Bamman. 2015. A treebank-based study on Latin Word Order. In Gerd Haverling (ed.), *Latin Linguistics in the Early 21st Century. Acts of the 16th International Colloquium on Latin Linguistics*, 338–350. Uppsala: Uppsala Universitet.

Penn Supplement to MCVF. 2010. Penn Supplement to the MCVF Corpus by Anthony Kroch and Beatrice Santorini.

Pintzuk, Susan. 2004. Variationist approaches to syntactic change. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 509–528. Wiley Online Library.

Prévost, Sophie. 2018. Increase of pronominal subjects in Old French: Evidence for a starting-point in Late Latin. In Anne Carlier & Céline Guillot (eds.), *Latin tardif - français ancien: continuités et ruptures* Coll. Beihefte zur Zeitschrift für romanische Philologie, 169–198. Berlin: De Gruyter.

Rainsford, Thomas Michael. 2011. *The emergence of group stress in medieval French*: University of Cambridge dissertation.

Roberts, Ian. 1993. *Verbs and diachronic syntax: A comparative history of English and French*. Dordrecht: Kluwer.

Roberts, Ian. 2014. Taraldsen's Generalization and language change: Two ways to lose null subjects. In Peter Svenonius (ed.), *Functional structure from top to toe: The cartography of syntactic structures*, vol. 9, 115–148. Oxford University Press.

Rouveret, Alain. 2004. Les clitiques pronominaux et la périphérie gauche en ancien français. *Bulletin de la Société de Linguistique de Paris* 99(1). 181–237.

Schøsler, Lene. 2002. La variation linguistique: Le cas de l'expression du sujet. In Rodney Sampson & Wendy Ayres-Bennett (eds.), *Interpreting the history of French, A festschrift for Peter Rickard on the occasion of his eightieth birthday*, 187–208. Amsterdam - New York, NY: Rodopi.

Simonenko, Alexandra. 2017. Determiners as a probe into diachronic information structure & the Definiteness cycle. Handout for a talk at Formal Diachronic Semantics 2, Saarland University, Saabrücken, November 20-21. `https://www.academia.edu/35249405/Determiners_as_a_probe_into_diachronic_information_structure_and_the_Definiteness_cycle`.

Simonenko, Alexandra & Anne Carlier. 2016b. The evolution of the French definite article: from strong to weak. Handout for a talk at *Going Romance*. Johann Wolfgang Goethe-Universität, Frankfurt. 08-10/12/2016. `https://www.academia.edu/28473104/The_evolution_of_the_French_definite_article_from_strong_to_weak`.

Simonenko, Alexandra, Benoit Crabbé & Sophie Prévost. 2018. Agreement syncretisation and the loss of null subjects: Quantificational models for Medieval

French. Accepted in *Language Variation and Change*. http://ling.auf.net/lingbuzz/003491.

Simonenko, Alexandra & Paul Hirschbühler. 2012. Placement de clitiques dans les propositions V1 et évolution de la structure de la proposition en ancien français. In Monique Dufresne (ed.), *Constructions en changement* Les voies du français, 11–53. Sainte-Foye: Presses de l'Université Laval.

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

Vennemann, Theo. 1974. Topics, subjects and word order. In J. M. Anderson & Charles Jones (eds.), *Historical linguistics: Proceedings of the first international conference on historical linguistics*, vol. 1, 339–376. Amsterdam: North Holland.

Walkden, George & Kristian Rusten. 2016. Null subjects in Middle English. *English Language and Linguistics* 1–35.

Zaring, Laurie. 2011. On the nature of OV and VO order in early Old French. *Lingua* 121. 1831–1852.

Zimmermann, Michael. 2014. *Expletive and referential subject pronouns in Medieval French*, vol. 556. Walter de Gruyter GmbH.

A Appendix

We used the following set of queries to search MCVF and Penn Supplement to MCVF for the relevant data. Queries are composed using the search language of the CorpusSearch software. An on-line manual for the software can be found at the following address http://corpussearch.sourceforge.net/. Ingason (2016) gives a detailed overview of the search language. Lists of tags used in the corpora can be found at the following link http://gtrc.voies.uottawa.ca/manuel/manuel-morpho-fr/index.htm, as well as a syntactic annotation manual http://gtrc.voies.uottawa.ca/manuel/syntax-manual-fr/index.htm and additional information about text editions: http://www.ling.upenn.edu/ beatrice/corpus-ling/frenchTexts.html.

A.1 Queries

node: IP*
add_to_ignore: ITJ*|NP-VOC*|NEG*|PON|*-LFD
define: GTRC.def
coding_query:

```
1: {
\0842: (*STRASB* inID)
\0900: (*EULALI* inID)
\0980: (*LEGER* inID)
\1000: (*PASSION* inID)
\1050: (*ALEXIS* inID)
\1100: (*ROLAND* inID)
\1117: (*LAPIDAL* inID)
\1120: (*BRENDAN* inID)
\1128: (*BESTIAIRE* inID)
\1130: (*WILLELME* inID)
\1131: (*GORMONT* inID)
\1138: (*JUISE* inID)
\1150: (*LAPIDFP* inID)
\1151: (*PSORNE* inID)
\1155: (*QUATRELIVRE* inID)
\1156: (*ENEAS* inID)
\1165: (*PROVSERLO* inID)
\1173: (*BECKET* inID)
\1174: (*BENDUC* inID)
\1175: (*FANTOSME* inID)
\1176: (*MIRNDORL* inID)
\1177: (*YVAIN* inID)
\1183: (*ADGAR* inID)
\1185: (*COBE* inID)
\1190: (*BORON* inID)
\1191: (*BLONDNESLE* inID)
\1192: (*DIALGREG* inID)
\1193: (*SBERNAN* inID)
\1180: (*MARIE* inID)
\1194: (*CHIEVRES* inID)
\1200: (*AUCASSIN* inID)
\1201: (*DIALAME* inID)
\1202: (*SERMMADN* inID)
\1205: (*CLARI* inID)
\1212: (*EUSTACE-FISHER* inID)
\1220: (*PSEUDOTURPIN* inID)
\1223: (*AGNES* inID)
\1224: (*EUSTACE-PETERSEN* inID)
\1225: (*QUESTE* inID)
\1255: (*EUSTACE-MURRAY* inID)
\1270: (*CASSIDORUS* inID)
\1275: (*Roisin* inID)
\1279: (*SOMME* inID)
\1309: (*JOINVILLE* inID)
\1330: (*PERCEFOREST* inID)
\1370: (*PRISE* inID)
\1373: (*FROISSART* inID)
\1427: (*MORCHESNE* inID)
\1440: (*JOIES* inID)
\1450: (*CNNA* inID)
\1498: (*COMMYNES* inID)
\1523: (*TESTAMENT* inID)
\1527: (*BAYART* inID)
\1572: (*VALOIS* inID)
\1572: (*ALL* inID)
\1660: (*GENDRON* inID)
\1680: (*MORIN* inID)
\1740: (*DUPLESSIS* inID)
}
```

2: {
imp: (IP-IMP* idoms finiteVerb) AND (finiteVerb idoms !**)
mat: (IP-MAT* idoms finiteVerb) AND (finiteVerb idoms !**)
whsbj: ((IP-SUB* idoms finiteVerb) AND (finiteVerb idoms !**)) AND (IP-SUB* idomsmod NP-SBJ* *T**)
whobj: ((IP-SUB* idoms finiteVerb) AND (finiteVerb idoms !**)) AND (IP-SUB* idomsmod NP-ACC* *T**)
sub: ((IP-SUB* idoms finiteVerb) AND (finiteVerb idoms !**))
na: ELSE
}

3: {
proncp: ((IP* idoms NP-SBJ*) AND (NP-SBJ* idoms PRO)) AND (NP-SBJ* idoms CP*)
proncp: ((IP* idoms NP-SBJ*) AND (NP-SBJ* idoms D)) AND (NP-SBJ* idoms CP*)
prondem: ((IP* idoms NP-SBJ*) AND (NP-SBJ* idomsmod PRO DemPronSg|DemPronPl))
pron: ((IP* idoms NP-SBJ*|CL-NP-SBJ*) AND (NP-SBJ*|CL-NP-SBJ* idoms PRO))
pron: ((IP* idoms NP-SBJ*|CL-NP-SBJ*) AND (NP-SBJ*|CL-NP-SBJ* idomsonly D))
imp: (IP* idoms NP-SBJ*|CL-NP-SBJ*) AND (NP-SBJ*|CL-NP-SBJ* idoms PROIMP)
prop: (IP* idoms NP-SBJ*) AND (NP-SBJ* idoms NPRS|NPRPL)
frel: (IP* idoms NP-SBJ*) AND (NP-SBJ* idoms CP-FRL*)
noun: (IP* idoms NP-SBJ*) AND (NP-SBJ* idoms NCS|NCPL|NP*|ADJ*|Q*|NUM*|QTP*)
coord: (IP* idoms NP-SBJ*) AND (NP-SBJ* idoms *con*)
impnull: (IP* idoms NP-SBJ*) AND (NP-SBJ* idoms *proimp*|*PROIMP*)
null: (IP* idoms NP-SBJ*) AND (NP-SBJ* idoms null)
other: ELSE
}

4: {
proncp: ((IP* idoms NP-ACC*) AND (NP-ACC* idoms PRO)) AND (NP-ACC* idoms CP*)
proncp: ((IP* idoms NP-ACC*) AND (NP-ACC* idoms D)) AND (NP-ACC* idoms CP*)
prondem: ((IP* idoms NP-ACC*) AND (NP-ACC* idomsmod PRO DemPronSg|DemPronPl))
cl: (IP* idoms CL-NP-ACC*)
pron: ((IP* idoms NP-ACC*) AND (NP-ACC* idomsonly PRO|D))
frel: (IP* idoms NP-ACC*) AND (NP-ACC* idoms CP-FRL*)
prop: (IP* idoms NP-ACC*) AND (NP-ACC* idoms NPRS|NPRPL)
noun: (IP* idoms NP-ACC*) AND (NP-ACC* idoms NCS—NCPL—NP*|ADJ*|Q*|NUM*|QTP*)
na: (IP* idoms !NP-ACC*)
other: ELSE
}

5: {
sbj-obj-v: (((((IP* idoms NP-ACC*) AND (IP* idoms NP-SBJ*|CL-NP-SBJ*)) AND (NP-SBJ*|CL-NP-SBJ* precedes NP-ACC*)) AND (IP* idoms finiteVerb)) AND (NP-ACC* precedes finiteVerb)) AND (NP-SBJ*|CL-NP-SBJ* idoms !**)
obj-sbj-v: (((((IP* idoms NP-ACC*) AND (IP* idoms NP-SBJ*|CL-NP-SBJ*)) AND (NP-ACC* precedes NP-SBJ*|CL-NP-SBJ*)) AND (IP* idoms finiteVerb)) AND (NP-SBJ*|CL-NP-SBJ* precedes finiteVerb)) AND (NP-SBJ*|CL-NP-SBJ* idoms !**)
sbj-v-obj: (((((IP* idoms NP-ACC*) AND (IP* idoms NP-SBJ*|CL-NP-SBJ*)) AND (IP* idoms finiteVerb)) AND (NP-SBJ*|CL-NP-SBJ* precedes finiteVerb)) AND (finiteVerb precedes NP-ACC*)) AND (NP-SBJ*|CL-NP-SBJ* idoms !**)
obj-v-sbj: (((((IP* idoms NP-ACC*) AND (IP* idoms NP-SBJ*|CL-NP-SBJ*)) AND (IP* idoms finiteVerb)) AND (NP-ACC* precedes finiteVerb)) AND (finiteVerb precedes NP-SBJ*|CL-NP-SBJ*)) AND (NP-SBJ*|CL-NP-SBJ* idoms !**)
v-sbj-obj: (((((IP* idoms NP-ACC*) AND (IP* idoms NP-SBJ*|CL-NP-SBJ*)) AND (NP-SBJ*|CL-NP-SBJ* precedes NP-ACC*)) AND (IP* idoms finiteVerb)) AND (finiteVerb precedes NP-SBJ*|CL-NP-SBJ*)) AND (NP-SBJ*|CL-NP-SBJ* idoms !**)
v-obj-sbj: (((((IP* idoms NP-ACC*) AND (IP* idoms NP-SBJ*|CL-NP-SBJ*)) AND (NP-ACC* precedes NP-SBJ*|CL-NP-SBJ*)) AND (IP* idoms finiteVerb)) AND (finiteVerb precedes NP-ACC*)) AND (NP-SBJ*|CL-NP-SBJ* idoms !**)
obj-v: ((IP* idoms NP-ACC*) AND (IP* idoms finiteVerb)) AND (NP-ACC* precedes finiteVerb)
v-obj: ((IP* idoms NP-ACC*) AND (IP* idoms finiteVerb)) AND (finiteVerb precedes NP-ACC*)
na: ELSE }

6: {
prose: (*PROVSERLO*|*DIALGREG*|*SBERNAN*|*CHIEVRES*|*AUCASSIN*|*DIALAME*|*SERMMADN*
|*CLARI*|*PSEUDOTURPIN*|*AGNES*|*QUESTE*|*EUSTACE-MURRAY*|*CASSIDORUS*|*Roisin*|*SOMME*
|*JOINVILLE*|*PERCEFOREST*|*FROISSART*|*MORCHESNE*|*JOIES*|*CNNA*|*COMMYNES*|*TESTAMENT*
|*BAYART*|*VALOIS*|*ALL*|*GENDRON*|*MORIN*|*DUPLESSIS*|*STRASB*|*WILLELME*|*LAPIDFP*|*PSORNE*
|*QUATRELIVRE* inID)
verse: (*BECKET*|*BENDUC*|*FANTOSME*|*MIRNDORL*|*YVAIN*|*ADGAR*|*COBE*|*BORON*|*BLONDNESLE*
|*MARIE*|*EUSTACE-FISHER*|*EUSTACE-PETERSEN*|*PRISE*|*ENEAS*|*GORMONT*|*JUISE*|
EULALI*|*LEGER*|*PASSION*|*ALEXIS*|*ROLAND*|*LAPIDAL*|*BRENDAN*|*BESTIAIRE* inID) }

7: { def: (IP* idoms NP-SBJ*) AND (NP-SBJ* idomsmod D detDef)
def: (IP* idomsmod NP-SBJ* NP) AND (NP idomsmod D detDef)
dem: (IP* idoms NP-SBJ*) AND (NP-SBJ* idomsmod D detDem)
dem: (IP* idomsmod NP-SBJ* NP) AND (NP idomsmod D detDem)
indef: (IP* idoms NP-SBJ*) AND (NP-SBJ* idomsmod D detIndef)
indef: (IP* idomsmod NP-SBJ* NP) AND (NP idomsmod D detIndef)
part: (IP* idoms NP-SBJ*) AND (NP-SBJ* idomsmod DF detPart)
part: (IP* idomsmod NP-SBJ* NP) AND (NP idomsmod DF detPart)
poss: (IP* idoms NP-SBJ*) AND (NP-SBJ* idoms DZ*)
poss: (IP* idomsmod NP-SBJ* NP) AND (NP idoms DZ*)
tel: (IP* idoms NP-SBJ*) AND (NP-SBJ* idomsmod ADJ tel)
tel: (IP* idomsmod NP-SBJ* NP) AND (NP idomsmod ADJ tel
q: (NP-SBJ* idoms Q|QP*) OR (NP-SBJ* idoms ADJNUM*|ADJR*|ADJS*)
q: (NP-SBJ* idoms ADJ*) AND (ADJ* idoms [Nn]ul|[Nn]uls)
q: (NP-SBJ* idomsmod NP Q|QP*)
q: (NP-SBJ* idomsmod ADJ* Q|QP*)
zero: ELSE
}

8: { def: (IP* idoms NP-ACC*) AND (NP-ACC* idomsmod D detDef)
def: (IP* idomsmod NP-ACC* NP) AND (NP idomsmod D detDef)
dem: (IP* idoms NP-ACC*) AND (NP-ACC* idomsmod D detDem)
dem: (IP* idomsmod NP-ACC* NP) AND (NP idomsmod D detDem)
indef: (IP* idoms NP-ACC*) AND (NP-ACC* idomsmod D detIndef)
indef: (IP* idomsmod NP-ACC* NP) AND (NP idomsmod D detIndef)
part: (IP* idoms NP-ACC*) AND (NP-ACC* idomsmod DF detPart)
part: (IP* idomsmod NP-ACC* NP) AND (NP idomsmod DF detPart)
poss: (IP* idoms NP-ACC*) AND (NP-ACC* idoms DZ*)
poss: (IP* idomsmod NP-ACC* NP) AND (NP idoms DZ*)
tel: (IP* idoms NP-ACC*) AND (NP-ACC* idomsmod ADJ tel)
tel: (IP* idomsmod NP-ACC* NP) AND (NP idomsmod ADJ tel)
q: (NP-ACC* idoms Q|QP*) OR (NP-ACC* idoms ADJNUM*|ADJR*|ADJS*)
q: (NP-ACC* idoms ADJ*) AND (ADJ* idoms [Nn]ul|[Nn]uls)
q: (NP-ACC* idomsmod NP Q|QP*)
q: (NP-ACC* idomsmod ADJ* Q|QP*)
zero: ELSE }

A.2 Definitions

Search queries contain the following abbreviations given in the GTRC.def file:

finite verb: AJ|EJ|LJ|MDJ|VJ

tel: [Tt]el|[Ii]tel|[Tt]ele|[Ii]tele|[Aa]ltretel|[Aa]utretel|[Ii]tels|[Tt]els|[Aa]utres|[Aa]utre|[Tt]elles |[Tt]elle|[Aa]ultre|[Tt]eil|[Tt]eu|[Tt]au

detPart: @au|@aus|@el|@es|@es@|@eus|@l|@ou|@u|d#|d'|d@|de|de@|@ES

detIndef: ·I·|UN|UNE|.i.|.j.|@un|@ una|@una@|@une|j.|u|u@|um|Un|un| un'|una|Une|une|unes|Ung|ung|ungs|unne|Uns|uns|unz|un[e]|Un[s]|un[s]|vn|vns|·i·|ún|úne|UNS

detDef: @u|L@|@ú|@U|@OU|s'|@ES|Lí|U|@0|L#|LES|@L|LA|LE|LI|L'|
@au|@au@|@aus|@el|@els|@es|@es#|@es@|@eu|@eus|@ez|@l
|@l#|@l@|@la|@la@|@li|@lo|@ls|@o|@ou|@ous|@s|@s@|@u|@u@|@us|
@us@|@ux|@ux@|@uz|@z|@[l]|@[s]|il|ils|l#|l'|La|la|La@|la@|las|Le|le|Le@|le@|Les
|les|Les@|les@|lez|Li|li|li@|lis|Lj|lj|lla|Lo|lo|lo@|los|lou|lu|ly|l[e]|l[i]|l'|l|lá|lé|lí|@S

detDem: se@|se|Ses|C@|cest[e]|CESTE|ICEST|se|ses|CELE|CIL|CEL|CEST|@cel|@cela
|@cest|@cesta|@ceste|@ci|@equesta|Ce|ce|Ce=sta|Ce@|ce@|ceals|ceau|Cel|cel|cel'|Cela
|cela|Cele|cele|celes|Celi|celi|Celle|celle|celles|Cellui|cellui|cels|Celui|celui
|Celuy|celuy|Cenes|Ces|ces|Cest|cest|cest'|Cesta|cesta|Ceste|ceste|Cestes|cestes|cesti
|cestre|Cestui|cestui|cestuj|Cestuy|cestuy|cest[e]|Cet|cet|Cete|cete|Cette|cette
|ceu|Ceus|ceus|Cez|cez|Che|che|chel|chele|chelui|Ches|ches|chest|cheste|chesti
|chestui|chiel|Chil|chil|Chils|chils|Chis|chis|chist|chu|Chus|chus|ci|ciel|ciel'
|ciels|ciest|Cil|cil|cilla|Cils|cils|Cilz|cilz|Cis|cis|Cist|cist|Ciz|ciz|cé
|cés|cést|el|equesta|I=quist|iceals|iceau|Icel|icel|Icela|icela|Icele|icele|Iceles
|iceles|iceli|icelle|icelles|icellui|Icelluy|icelluy|icels|icelui|iceluy|iceol|iceols|Ices|ices
|Icest|icest|icest'|Icesta|icesta|Iceste|iceste|Icestes|Iceulx|iceulx|iceulz|iceus|Icez|icez
|ichel|Ichele|Ichis|ichis|Ichés|Icil|icil|Icist|icé|icés|iqueaus|Iqueste|Iquist
|iquist|quist|sest|set|sete|y=celle|ycelle|ycelui|yceulx|yceux|yceuz|ícel|íces
|ícest|[Cc]estei|[Cc]eos|[Cc]eu

DemPronSg: @se|@se|C#|C'|C@|CELE|CEO|CIL|Ce|Cecy|Cela|Cele|Celle|Ceo|Ceste|
Chil|Chils|Chist|Ci|Cil|Cils|Cilz|Cio|Cist|Co|Icelle|Icil|Icist|Ico|Il|Il=l@|Iquist|S'|Se|Seci
|an|c#|c'|c@|ce|cecy|cel|cela|cele|celle|ceste|cestes|ch'|ch@|chil|chils|chis|chius|cho|chou
|cil|cils|cilz|cio|cis|cist|co|c*|có|ice|icelle|ichist|icil|ico|ico|s#|s'|se|secy|ycelle
|Ç'|Ço|Çou|Çó|ç'|ça|ço|çou|çó|íço

DemPronPl: CELES|CIL|Celes|Chil|Chils|Chist|Ci|Cil|Cils|Cilz|Cist|Icelle
|Icil|Icist|autres|celes|celles|cestes|chil|chils|chis|chius|cil|cils|cilz|cis|cist|ichist|icil|lesquelles

A.3 Head types and determiners with direct objects

Head type	
free relative	0.03 (5)
noun	0.47 (735)
personal pronoun	0.03 (53)
pronoun with a CP-complement	0.09 (147)
demonstrative	**0.37** (584)
proper noun	0.02 (30)

Table 26: Head types in object phrases in OVS

Determiner	
definite	0.19 (139)
demonstrative	**0.21** (155)
tel	0.09 (65)
possessive	0.07 (50)
quantifier	0.15 (111)
indefinite	0.02 (16)
partitive	0.01 (6)
zero	0.26 (193)

Table 27: Determiners with nominal objects in OVS

It has to be noted that a zero determiner in medieval French is not to be equated with indefiniteness. The spread of overt determiners was another change that progressed gradually over the medieval period (e.g. Simonenko & Carlier 2016b), and in the earlier texts bare nouns occurred frequently in the contexts which in Modern French require a definite determiner, a demonstrative or a possessive pronoun (Mathieu 2009).

Head type	
free relative	0.002 (2)
noun	0.78 (913)
personal pronoun	0.06 (70)
pronoun with a CP-complement	0.05 (60)
demonstrative	0.05 (59)
proper noun	0.05 (62)

Table 28: Head types in object phrases in SOV

Determiner	
definite	0.25 (234)
demonstrative	0.06 (54)
tel	0.02 (20)
possessive	0.19 (170)
quantifier	0.1 (92)
indefinite	0.03 (26)
partitive	0.001 (1)
zero	0.34 (316)

Table 29: Determiners with nominal objects in SOV

Head type	
free relative	0.003 (17)
noun	0.83 (4158)
personal pronoun	0.01 (54)
pronoun with a CP-complement	0.06 (281)
demonstrative	0.06 (284)
proper noun	0.04 (216)

Table 30: Head types in object noun phrases in OV

Determiner	
definite	0.26 (1086)
demonstrative	0.02 (91)
tel	0.02 (82)
possessive	0.2 (845)
quantifier	0.14 (572)
indefinite	0.03 (142)
partitive	0.005 (21)
zero	0.3 (1319)

Table 31: Determiners with nominal objects in OV

Head type	
free relative	0.01 (27)
noun	0.84 (1909)
personal pronoun	0.00 (1)
pronoun with a CP-complement	0.1 (215)
demonstrative	0.01 (22)
proper noun	0.04 (86)

Table 32: Head types in object phrases in VSO

Determiner	
definite	0.25 (478)
demonstrative	0.03 (64)
tel	0.02 (39)
possessive	0.17 (330)
quantifier	0.14 (278)
indefinite	0.04 (81)
partitive	0.01 (32)
zero	0.31 (607)

Table 33: Determiners with nominal objects in VSO

Head type	
free relative	0.01 (167)
noun	0.84 (15652)
personal pronoun	0.002 (30)
pronoun with a CP-complement	0.08 (1534)
demonstrative	0.02 (417)
proper noun	0.034 (633)

Table 34: Head types in object phrases in SVO

Determiner	
definite	0.27 (4293)
demonstrative	0.03 (578)
tel	0.02 (248)
possessive	0.16 (2611)
quantifier	0.11 (1738)
indefinite	0.04 (687)
partitive	0.02 (287)
zero	0.33 (5210)

Table 35: Determiners with nominal objects in SVO