



Morphological triggers of syntactic changes: Treebank-based Information Theoretic approach

Alexandra Simonenko, Benoît Crabbé, Sophie Prévost

► To cite this version:

Alexandra Simonenko, Benoît Crabbé, Sophie Prévost. Morphological triggers of syntactic changes: Treebank-based Information Theoretic approach. *Treebanks and Linguistic Theories (TLT)*, Dec 2015, Varsovie, Poland. hal-01240245

HAL Id: hal-01240245

<https://inria.hal.science/hal-01240245>

Submitted on 8 Dec 2015

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

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

Morphological triggers of syntactic changes: Treebank-based Information Theoretic approach*

Alexandra Simonenko, Benoit Crabbé, Sophie Prévost

ALPAGE, LaTTiCe

E-mail: alexandra.simonenko@mail.mcgill.ca
benoit.crabbe@linguist.univ-paris-diderot.fr
sophie.prevost@ens.fr

Morphological triggers of syntactic changes: Treebank-based Information Theoretic approach

Abstract

This paper addresses the classic problem of the triggers of the passage from a relatively free word order to a strict SVO in the history of French ([26], [9], [8], [28], [12], [7]). We present a corpus-based modelling of two, likewise classic, lines of analysis. First, we explore the link between the loss of word order freedom and the disappearance of morphological case marking ([22], [23], [6], [20], [14]). Second, we evaluate the syncretisation of verbal agreement and massive appearance of overt preverbal subject pronouns ([1], [19], [21]) as a potential analogical trigger of a generalized SVO (e.g. [3] for an analogy-based explanation of the change in nominal syntax in Old English). Although the analytical intuitions themselves have a long history, only recently has it become possible to perform their quantitative evaluations due to the availability of large (for historical data) annotated treebanks of Medieval French ([15], [16], and [24]).

1 Introduction

This paper presents a quantitative corpus-based investigation of the possible causes of the fixation of the word order in the history of French using Information Theoretic measures. Medieval French (MF) went from a (loose) V2 system, permitting for all six permutations of S, O, and V, to a relatively strict SVO (e.g. [8], [12]).

*We gratefully acknowledge funding support from Labex EFL, grant ANR-10-LABX-0083. We are also thankful to three anonymous reviewers of TLT 2015 for very helpful feedback.

- (1) [L' altre meitet]_{obj} avrat_v Rollant_{sbj}
the other half will.have Roland
‘Roland will have the other half’ (1100-ROLAND-V,36.446) V2 in Old French
- (2) Roland_{sbj} aura_v l' autre moitié_{obj}
Roland will.have the other half
‘Roland will have the other half’ SVO in Modern French
- (3) *L' autre moitié_{obj} aura_v Roland
the other half will.have Roland
*V2 in Modern French

The passage to SVO has been frequently attributed to the disappearance of morphological case marking, on the assumption that linear position and case both can mark syntactic roles and therefore the former can substitute for the latter (e.g. [27, 289], [6]). This is at the least a plausible analysis for French since by the X century, the distinction between nominative and accusative in MF was mostly retained only for masculine nouns (e.g. *reis_{nom,sg}*, *rei_{acc,sg}*), and even that was becoming unstable ([12]), as illustrated by the unmarked subject in (5).

- (4) **Reis Chielperics** tam bien en fist...
king Chilpéric so well of.it made
‘King Chilperic dealt with it so well...’ (0980-LEGER-V,XII.80)
- (5) É li nostre **rei** nus jugerá...
and the our king us will.judge
‘And our king will judge us.’ (1150-QUATRELIVRE-P,17.529)

In the typological perspective, the existence of some sort of an inverse dependency between the fixedness of the word order (i.e. arguments having strict positions with respect to the predicate: either SVO or OVS) and the availability of morphological case marking has been claimed to hold in the literature ranging from [22] to [2]. However, the position-for-case substitution in MF has remained in the hypothetical realm since until recently it had been virtually impossible to quantify the relevant changes. Another difficulty consisted in the absence of comparable measures of the contributions of the two markers for the syntactic role identification. Below we propose a way to circumvent both problems by using Information Theoretic notions and distributions drawn from MCVF.

MCVF is a treebank of tagged, parsed and functionally annotated French texts from X to XVIII cc. with Penn treebank style annotation scheme (approx. 1 mln words). We used CorpusSearch, a tool for matching tree patterns in corpora, which can search for the relations of precedence and dominance, for specific morphological forms as well as code utterances for parameters such as word order and presence of an overt subject.¹

In addition, we explore a second, and compatible, explanation of the passage to SVO. It has at its core the syncretisation of verbal subject agreement suffixes,

¹<http://corpussearch.sourceforge.net/>

and the massive emergence of pronominal subjects. The argument runs as follows: syncretisation of verbal agreement led to the replacement of pro-drop by overt pronominal subjects. The latter, being prevailingly preverbal, triggered reanalysis of the position of all subjects, including nominal, as preverbal. In order to build a quantitative model, we propose to treat verbal subject agreement as a signal of subject’s person feature and to quantify it using, once again, entropy measures. We then compare temporal profiles of agreement syncretisation and pronominal subject expression. Finally, we compare the rate of subject expression and the rate of preverbal nominal subjects to see if the two are correlated.

2 Loss of morphological case and word order flexibility

2.1 Morphology-syntax tradeoff hypothesis

Since both word order and morphological changes manifest themselves as gradual replacements of one alternative by another over the centuries rather than “overnight” categorical shifts, establishing a temporal relation between the two has been virtually impossible until very recently due to the absence of tools for quantifying the relevant changes. Establishing the temporal profiles of the changes is, in turn, indispensable for modelling grammatical relations (if any) between the corresponding phenomena. These points seem to have been overlooked in the debate about the relationship between case and word order, which led to claims such as the following one from [10, 22]: “a ... complication with the theory that phonetic attrition of the classical Latin case system necessitated a fixed Romance SVO order is that it is simply not true. ... [L]ate Latin and early Romance retained at least a binary case system (nominative vs. oblique) and were characterized by Verb Second constraint, such that SVO was just one of many possible word orders. From this we can only conclude that there is no necessary causal relation between phonetic attrition, in this case acting upon the case system, and the emergence of analytic structural changes.” As we show below, such conclusions are unwarranted by the corpus data, given that the *robustness* of nominative marking, estimated based on the proportion of nominative marked subjects among all subjects, was different at different points in time (overall decreasing), and so was the robustness of linear position marking (overall increasing).² The mere fact that in a given text we find both nominative marked subjects and SVO orders does not necessarily speak for or against a particular relation between case and order. In the following section we propose a way to track diachronic changes in the distribution of case markers and linear orders and to measure their contribution to the identification of syntactic

²Note that our approach is very different from approaches evaluating the role of case based on considering all factors, lexical and grammatical (e.g. verbal semantics and discourse context), which could potentially be used as keys for recovering grammatical functions ([23], [17]). While those studies evaluate how often morphological case was crucial for recovering grammatical functions (e.g. [17, 62] estimates that it was the case only in 5-10% of utterances in Late Latin), we are estimating its unambiguity as a signal (see below).

functions.

2.2 Methodology

Building on the classic insight of [5] and others that morphological case and linear position can be used to signal syntactic roles, we propose a way to quantify their efficiency using Shannon’s entropy in order to give them a common quantificational expression. We start with a working “tradeoff” hypothesis: the expectation that as one signal weakens, an alternative signal gets stronger. Informally, the strength of a signal, its efficiency, is a measure of a marker’s unambiguity. To illustrate this, imagine that in one text among arguments with accusative marking there are 80% of direct objects and 20% of subjects, while in another text the proportions are 50% and 50% respectively. Informally, accusative marker is a less ambiguous in the first text than in the second, where it is maximally ambiguous.

This can be formalized using conditional entropy measures. Let X and Y be two discrete random variables, the conditional entropy is the quantity:

$$H[Y|X] = - \sum_{x \in X} P(X = x) \sum_{y \in Y} P(Y = y|X = x) \log_2 P(Y = y|X = x) \quad (1)$$

where Y is the dependent variable, conditioned on some context X . In our example, Y is a grammatical function, subject or object and X represents the context of the dependent in terms of its position with respects to the head or its case properties³.

In the next section, we describe a method for estimating the conditional entropies $H[\text{FUNCTION}|\text{CASE}]$ and $H[\text{FUNCTION}|\text{POSITION}]$ using distributions from [15] and [16].

2.3 Data extraction

The corpora are morphologically and syntactically annotated using Penn Treebank kind of annotations. It consist of 35 texts from 980 to 1740, which gives about 1 mln words. We extracted all clauses with a finite verb form and a dependent, being either an overt nominal subject or a nominal object. We included only the nouns belonging to the traditional first declension class (e.g. *reis* “king”). As a preliminary step, we manually defined the declension class of each noun form in the corpus and listed them separately. This step was necessary since morphological case marking was not operative in the second declension class (*femme* “woman”) during the attested periods and we had to exclude it from our study of the case marking evolution. We also excluded nouns featuring suppletive case marking (e.g. *ber_{nom}* vs. *baron_{nom}* ‘baron’), as well as nouns whose stems end in *s/z/x*, since for those case marking is neutralised. There is a total of 15,768 examples for subjects and 10,033 examples for objects. Each example is coded with the following variables:

³Although our models may look similar to those of [2], one should observe that their goal is opposite: [2] tries to measure to which extent the dependency structure is a good predictor of word order, whereas in our case we try to predict the dependency type given word order and case.

1. DATE. Each clause was coded for the date of the text from which it was taken (e.g. 980, 1155 etc.): our query matched the identifier node appended to every finite clause with the date attributed to a given text by a scholarly consensus.⁴
2. Syntactic FUNCTION. Every clause was coded as containing an overt nominal subject – *sbj* – or a nominal object – *obj*.⁵ Clauses containing subjects are those clauses with a constituent NP-SBJ dominating one of the following four tags: NCS, NCPL, NPRS, NPRPL, which correspond to common singular noun, common plural noun, proper singular noun, and proper plural noun respectively (see Fig. 1). Clauses containing objects are those with a constituent NP-ACC dominating a nominal tag, (Fig. 2).
3. POSITION of the dependant with respect to the finite verb.⁶
 - The code *pre* was assigned if the dependent NP constituent precedes linearly the finite verb tag (AJ, EJ, LJ, MDJ or VJ in [15]).
 - The code *post* was assigned if the dependent NP constituent follows linearly the finite verb tag (AJ, EJ, LJ, MDJ or VJ in [15]).
4. Morphological CASE.
 - The code *nom* was given to forms ending in *s/z/x* in singular and zero in plural (nominative marking)
 - The code *acc* was given to forms that have no ending in singular and *s/z/x* in plural (accusative marking);

Figure 1 is an example of a coded clause with a nominal subject. The clause is taken from *La Chanson de Roland*, a poem dated from around 1100 and containing a preverbal nominal subject in singular and ending with *s* (nominative pattern).

Figure 2 is another example from *La Chanson de Roland*. It illustrates a coded clause with a preverbal nominal object in singular with a zero ending (accusative pattern).

Finally, we use an additional PERIOD factor partitioning our extracted observations by century intervals. For each such PERIOD, we estimated the conditional entropies $H[\text{FUNCTION}|\text{POSITION}]$ and $H[\text{FUNCTION}|\text{CASE}]$ from the data set by

⁴Since some datings are approximate (e.g. a manuscript can be dated by the first quarter of a century), in some cases we had to choose an arbitrary date within the attributed period.

⁵We ran the query twice: on clauses with a finite verb and a subject (whether or not they contained a direct object) and on clauses with a finite verb and a direct object (whether or not they contained a subject). We then merged the two sets of coding strings where each line ended up corresponding to a subject or a direct object token.

⁶In our sample there were no cases of discontinuous subject constituents headed by a noun whereby one part of the constituent would precede the verb and the other one follow, thus creating ambiguity for determining the precedence relation. Thanks to an anonymous reviewer for bringing up this potentially problematic issue.

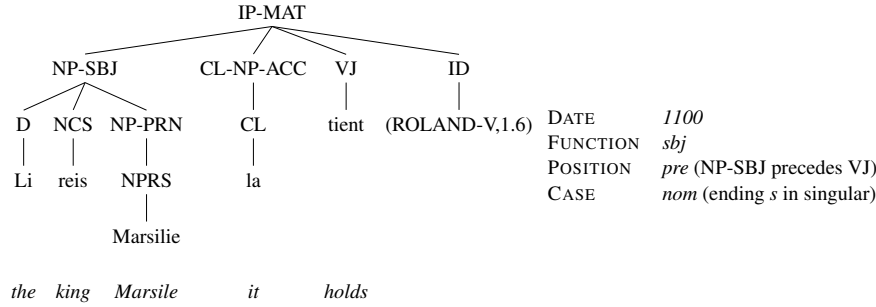


Figure 1: Coding for subject “The king Marsile holds it”

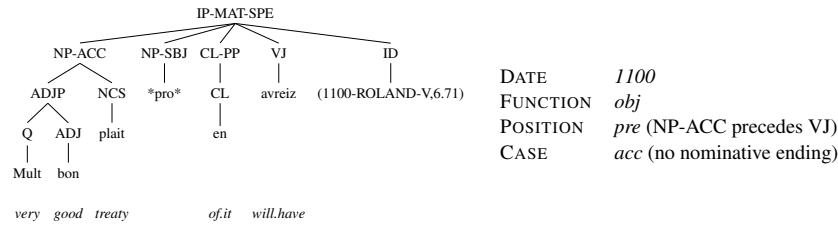


Figure 2: Coding for object “You will have a very good treaty out of this.”

maximum likelihood estimation. Note that this partition has been defined with the goal of avoiding data sparsity issues and ensuring that the actual counts in the data set are sufficiently high.⁷

2.4 Results

Entropy measures for the 1st declension are illustrated in Fig. 3, where high entropy corresponds to “weak” and low entropy to “strong” signals. For instance, high conditional entropy of FUNCTION given POSITION means that the probability for an argument in the preverbal position of being a subject was similar to that of being an object, while low entropy indicates a substantial difference. Overall, we can see that the entropy of FUNCTION given POSITION goes down, whereas the entropy of FUNCTION given CASE goes up.

A note is in order concerning an apparent zig-zag of the case signal measure, which, as it were, descends at the XIII c. and then goes back up at the XIV c. Upon closer examination, it turns out that the higher (compared to the following period) entropy in the XII c. is due to the lexical properties of one text, namely, *Li Quatre Livre des Reis*. Here among accusative marked arguments there are 496 objects and 398 subjects. However, among the latter, there are 215 tokens of the name *David*. In the corpus this name appears in the nominative form, *Davids*, only

⁷In other words, we do not face the same kind of estimation problems that are reported for instance by [2]. We also illustrate this in the next few sections by reporting error bars, on the plots, computed by statistical bootstrapping.

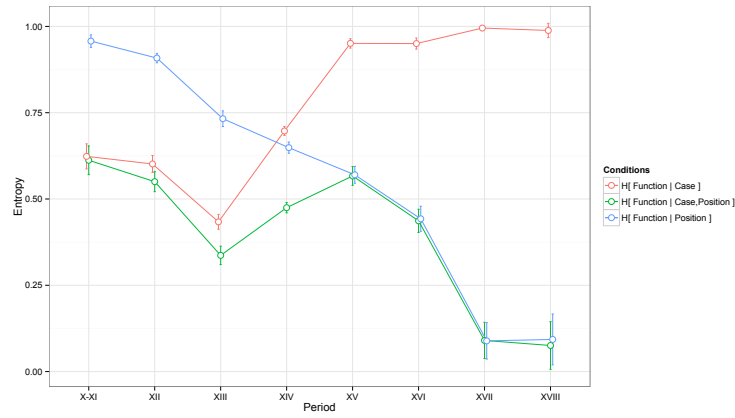


Figure 3: Morphological case and position as signals of grammatical function

twice, in the chronicles of Jean Froissart dated from approximately 1370. Given that proper nouns may have different morphological behaviour than common noun, we also did entropy estimations on the set of common noun only, Fig. 4.

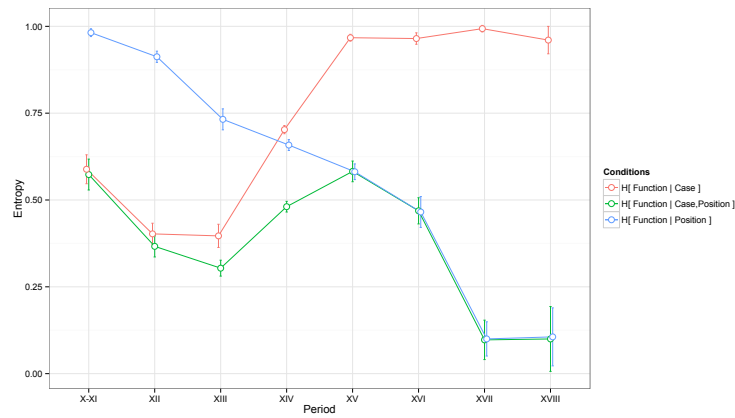


Figure 4: Morphological case as a signal of grammatical function for common nouns

Coming back to the general picture, if we assume that the purported “trade-off” in signal strength is immediate, we then expect that when the position-signal was still very weak (entropy around 0.9), the case-signal should have been strong in order to efficiently mark subject/object distinction. Instead, what we find is a weak case signal in the earliest periods of MF (entropy around 0.6). That is, it appears that if the weakening of case signal was indeed the trigger of the word order changes, the effect was not immediate. That there can be a temporal lag be-

tween morphological changes and their possible syntactic consequences has been suggested in the studies of the relation between the impoverishment of verbal inflection and the disappearance of verbal movement in Germanic languages (e.g. [25]).

Another question, however, is why it should be the position and not a set of new morphological markers which replaces the lost case, and, specifically, why subjects occupy preverbal and not postverbal position. Below we investigate the hypothesis that nominal subjects became strictly preverbal by analogy with pronominal subjects, whose rate soared in MF, ([4], [18]), following the syncretisation of verbal subject agreement.

3 Verbal inflection and loss of pro-drop

Part of the morphological impoverishment of MF was the spread of the subject agreement ending *e* from the 3rd to the 1st person singular in verb forms of the traditional 1st conjugation class (with *-er* infinitives) in indicative and subjunctive moods of the present tense ([11, 200,207]) (*aim* ‘(I) love’ becomes *aime*, as in *il aime* ‘he loves’), as well as the spread from the ending *s* from the 2nd to the 1st person singular in verb forms of the traditional 2nd conjugation class (with *-ir*, *-oir* and *re* infinitives). A non-syncretised paradigm identifies the subject’s person right at the position of V, which is impossible with an ambiguous *e*, given the possibility of pro-drop and a flexible word order. However, if the pronominal subject is always overt (i.e. there is no pro-drop), identification of the subject’s person is more efficient: pronouns in MF are most often preverbal and unambiguous as to their grammatical role.

The disappearance of pro-drop in MF has been linked to the impoverishment of the verbal inflection ([21], [13]), but there has been no quantificational studies of the data bearing on the possible connection. We examine the two phenomena, again, in terms of entropy measures. In order to estimate the efficiency of verbal inflection for identification of subject’s person we define a binary variable PERSON with sample space {1st, 2nd, 3rd} and estimate its entropy given endings *e* and *s*. Most likely syncretisation extended beyond these endings in oral language affecting all final stops and fricatives and making all endings phonologically indistinguishable except for 1st and 2nd person plural. However, due to the unavailability of oral data, we have to approximate this process by focusing on the fate of *e* and *s*, which can be quantified.

3.1 Data extraction

We extracted all clauses with 1st conjugation verb forms ending in *e* or with 2nd conjugation verb forms ending in *s* and with an overt nominal or pronominal subject (total of 3,202). This allowed us to estimate how good the two endings were to predict subject’s person. Below we explicate the coding procedure. The variables

we coded for are as follows:

1. DATE is extracted as in section 2.3.
2. CONJUGATION of the verbal form:⁸
 - *first* if the form belonged to the first conjugation.
 - *first* if the form belonged to the second conjugation.
3. ENDING of the verbal form. The codes were assigned corresponding to endings of verbal forms, such as:⁹
 - The code *e* was assigned if the verbal form ended in *e*, *ë*, *é* or *è*.
 - The code *s* was assigned if the verbal form ended in *s*, *z* or *x*.¹⁰
4. PERSON of the subject: *first*, *second*, or *third*.¹¹

We estimated the conditional entropy $H[\text{PERSON}|\text{ENDING}]$ from the data set by maximum likelihood estimation. In order to track the evolution of pro-drop, we estimated the entropy $H[\text{SUBJECT}]$ of the variable SUBJECT, which coded all clauses with a finite verb and either a null or a pronominal subject for the presence/absence of an overt pronominal subject (*yes*, *no*).¹²

3.2 Results

The results are illustrated in Fig. 5. Entropy of subject's person given ending predictably increases and eventually goes up to 1, meaning that *e* and *s* progressively become indiscriminate with respect to the person of the subject, reaching maximum ambiguity by the end of the MF period.

At the same time, entropy of SUBJECT goes down, that is, the probability of having an overt pronominal subject becomes progressively greater than not having one. We also see that entropy of subject's person given ending had already been well above 0 when entropy of *Subject* was still 1, meaning, in our model, that syncretisation precedes the decline of pro-drop, which corroborates (but does not prove, of course) the hypothesis that the former triggered the latter.

⁸Similarly to our treatment of nominal declensions, we extracted verbal forms from the clauses with an overt nominal or pronominal subject and listed them separately according to their conjugation type.

⁹This is not an exhaustive list of endings we used in our coding, but in this paper we are interested only in *e* and *s*.

¹⁰Our query made sure to avoid confusion between other endings with final *s*, *z*, *x* (such as 2nd and 1st person plural endings *ez*, *ons* etc.) and the relevant endings.

¹¹We extracted all pronominal forms from the corpus and classified them by person.

¹²We excluded from our counts coordination structures with subject ellipsis, since this phenomenon persists in Modern French as well and is therefore irrelevant for the question of the evolution of overt pronominal subjects.

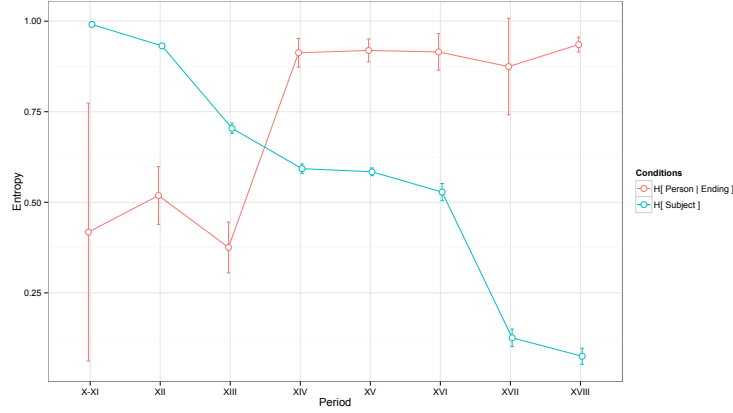


Figure 5: Pro-drop and verbal endings *s* and *e* as signals of subject's person

4 Pronominal and nominal subjects: analogy

We now evaluate the hypothesis that the massive appearance of overt pronominal subjects, almost always preverbal, triggered an analogical change in the syntax of nominal subjects which progressively became preverbal. First, we need to establish the fact that the growing rate of utterances with overt subjects is due to the emergence of overt pronominal subjects, whereas the rate of nominal subjects was declining. To that end, we coded the corpus for the following variables.¹³

1. DATE is extracted as in section 2.3.
2. PRONOUN received value *yes* if an utterance contained an overt pronominal subject and *no* otherwise.
3. NOUN received value *yes* if an utterance contained a non-pronominal subject and *no* otherwise.
4. POSITION of the subject with respect to the finite verb (*pre* vs. *post*).

Fig. 6 shows that the probability of having a non-pronominal subject was slowly going down from 25% to 0.05%, whereas the probability of an overt pronominal subject raised from 41% to 92%.

Pronominal subjects in MF are overwhelmingly preverbal. For instance, in X–XI cc. there was about 56% of preverbal nominal subjects (465 out of 832) whereas among pronominal subject the rate was 68% (1039 out of 1534). On the hypothesis about an analogical change in the syntax of nominal subjects, we compare the profile of the emergence of overt pronominal subjects and the fixation

¹³We excluded relatives clauses, imperatives, and wh-questions because of their idiosyncratic subject syntax, as well as coordination structures with subject ellipsis.

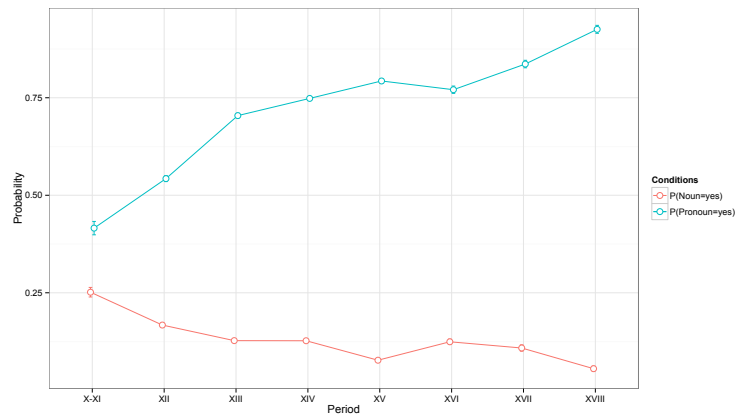


Figure 6: Pronominal and non-pronominal subjects

of nominal subjects in the preverbal position. Fig. 7 shows the probability of an overt pronominal subject calculated on the sample of clauses with a pronominal subject or without an overt subject, $P(\text{SUBJECT} = \text{yes})$, and the probability of *non-pronominal subjects* being preverbal, $P(\text{POSITION} = \text{pre}, \text{NOUN} = \text{yes})$. The two measures are significantly correlated (Pearson's $r = 0.82$, $p = 0.01$).

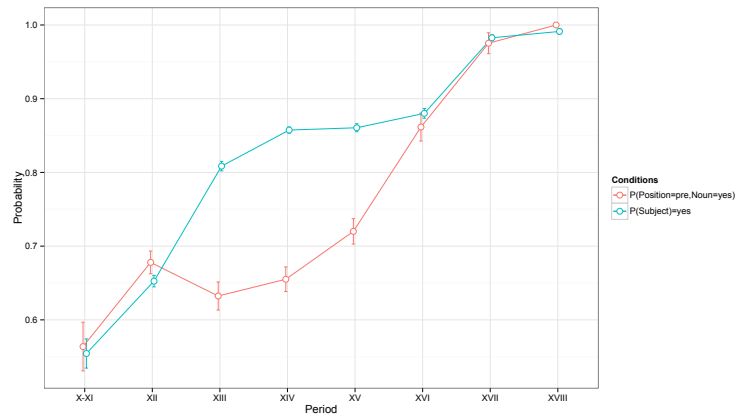


Figure 7: Preverbal non-pronominal subject and overt pronominal subject

5 Conclusions

In this paper we showed, first, that the loss of case and the fixation of argument positions, if taken as signals of grammatical functions, are in a tradeoff relation, assuming that a tradeoff does not have to be immediate. That is, the position signal is still very weak at the time when case signal is already imperfect. It must be noted, however, that due to the lack of data prior to X c., we cannot estimate whether morphological case was ever a perfect function signal (i.e. completely unambiguous). Second, our results suggest that a similar tradeoff relation was holding between the degree of unambiguity of verbal endings and the rate of expression of pronominal subjects. As a side note, one cannot help noticing the striking similarity between the temporal profiles of the two morphological phenomena, case and endings signals, which we will have to leave to future research. Third, we found a strong correlation between the replacement of pro-drop by overt pronominal subjects and the migration of non-pronominal subjects to the preverbal position. A correlation does not of course entail causality, but the results suggest that the two were related in a highly non-accidental manner. Parsed treebanks made it possible for us to develop with Information Theoretic expressions for morphological and syntactic phenomena thereby making them comparable on the diachronic plane, which is a novel contribution.

References

- [1] Lucien Foulet. *Petite syntaxe de l'ancien français*. Champion, troisième édition revue. Réédition 1982, Paris, 1928.
- [2] Richard Futrell, Kyle Mahowald, and Edward Gibson. Quantifying word order freedom in dependency corpora. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pages 91–100, 2015.
- [3] Aafke Hulk and Ans van Kemenade. Verb second, pro-drop, functional projections and language change. In *Clause structure and language change*, pages 227–256. Oxford University Press, 1995.
- [4] Georg A. Kaiser. 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”*, pages 131–156, 2009.
- [5] Edward Keenan. Toward a universal definition of ‘subject’. In C Li, editor, *Subject and Topic*. Academic Press, New York, 1976.
- [6] Paul Kiparsky. The rise of positional licensing. In Ans van Kemenade and Nigel Vincent, editors, *Parameters of morphosyntactic change*, pages 460–494. Citeseer, 1997.

- [7] Anthony Kroch and Beatrice Santorini. On the word order of Early Old French, 2014. SinFonIJ 7, Graz, Austria, September 2014.
- [8] Marie Labelle. Clausal architecture in Early Old French. *Lingua*, 117(1):289–316, 2007.
- [9] Marie Labelle and Paul Hirschbühler. Changes in clausal organization and the position of clitics in Old French. In M. Batllori, M.L. Hernanz, C. Picallo, and F. Roca, editors, *Grammaticalization and Parametric Variation*, pages 60–71. Oxford University Press, Oxford, 2005.
- [10] Adam Ledgeway. *From Latin to Romance: Morphosyntactic typology and change*. Oxford University Press, 2012.
- [11] Christiane Marchello-Nizia. *Histoire de la langue française aux XIVe et XVe siècles*. Dunod, Paris, 1992.
- [12] Christiane Marchello-Nizia and Magali Rouquier. 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, editor, *Constructions en changement. Hommage à Paul Hirschbühler*, pages 111–155. Presses de l’Université de Laval, 2012.
- [13] Eric Mathieu. Stylistic fronting in old french. *Probus*, 18(2):219–266, 2006.
- [14] Thomas McFadden. *The position of morphological case in the derivation: A study on the syntax-morphology interface*. PhD thesis, University of Pennsylvania, 2004.
- [15] Corpus MCVF annoté syntaxiquement, sous la direction de France Martineau, avec Paul Hirschbühler, Anthony Kroch et Yves Charles Morin, 2010.
- [16] Penn Supplement to the MCVF Corpus by Anthony Kroch and Beatrice Santorini, 2010.
- [17] Harm Pinkster. *Latin Syntax and Semantics*. Routledge, New York, 1990.
- [18] Sophie Prévost. Emergence and development of personal pronoun subjects: study of a bilingual Latin-old French corpus, To appear.
- [19] Ian Roberts. *Verbs and Diachronic Syntax: A Comparative History of English and French*. Kluwer, Dordrecht, 1993.
- [20] Ian Roberts. Directionality and word order change in the history of English. In Ans Van Kemenade and Nigel Vincent, editors, *Parameters of morphosyntactic change*, pages 397–426. Cambridge University Press, 1997.
- [21] Bernhard Wolfgang Rohrbacher. *The Germanic VO languages and the full paradigm: a theory of V to I raising*. PhD thesis, University of Massachusetts, Amherst, 1994.

- [22] Edward Sapir. *Language, an introduction to the study of speech*. Harcourt, Brace and Co., New York, 1921.
- [23] Lene Schøsler. La déclinaison bicasuelle de l'ancien français: son rôle dans la syntaxe de la phrase, les causes de sa disparition. In *Études romanes de l'Université d'Odense*, volume 19. Odense University Press, 1984.
- [24] Achim Stein and Sophie Prévost. Syntactic annotation of medieval texts: the Syntactic Reference Corpus of Medieval French (SRCMF). In P. Bennett, M. Durrell, S. Scheible, and R. Whitt, editors, *New Methods in Historical Corpus Linguistics. Corpus Linguistics and International Perspectives on Language*, volume 3, pages 75–82. CLIP, Tübingen, 2013.
- [25] John D. Sundquist. *Morphosyntactic change in the history of the Mainland Scandinavian languages*. PhD thesis, Indiana University, 2002.
- [26] Barbara Vance. *Syntactic change in medieval French*. Studies in Natural Language and Linguistic Theory. Kluwer, Dordrecht/Boston/London, 1997.
- [27] Theo Vennemann. An Explanation of Drift. In Charles Li, editor, *Word Order and Word Order Change*, pages 269–305. University of Texas Press, Austin, 1975.
- [28] Laurie Zaring. On the nature of OV and VO order in early Old French. *Lingua*, 121:1831–1852, 2011.