

## Effects of literary form on grammatical changes: A treebank study

This paper investigates the effects of the text form (prose vs. verse) on diachronic grammatical changes in Medieval French using parsed treebanks [5] and [6] ( $\approx 1$  mln words annotated with Penn scheme).

Despite the common intuition that the prose is somehow more “advanced” than the verse contemporary to it with respect to grammatical changes, in the absence of quantificational evaluations, the magnitude of the difference has remained unknown. At the same time, the prevalence of verse in the earliest periods of documented French (i.e. X–XII c.) results in a strong and unavoidable correlation between time and form, which potentially undermines the results of the studies attempting to formally model Medieval French evolution. This study contributes to solving this problem by comparing two historical changes, the loss of pro-drop and that of  $OV_{finite}$  order, across text forms.

We model the loss of pro-drop using logistic regression on a sample of finite clauses with either an overt pronominal or null subject (total of 95,574), excluding relatives, imperatives, and wh-questions because of their idiosyncratic subject syntax. Our model  $P(\text{Subject} = \text{overt} | \text{Date}) = \frac{e^{\alpha + \beta * \text{Date}}}{1 + e^{\alpha + \beta * \text{Date}}}$  has  $\alpha = -1.889$  and  $\beta = 0.0027$  for prose and  $\alpha = -2.083$  and  $\beta = 0.0021$  for verse (Fig. 1). That is, the rates of change ( $\beta$ ) are nearly identical, whereas the difference in intercepts means that the verse is more “conservative”, which is in line with the traditional intuition. This also points towards a possible extension of Kroch’s (1989) Constant Rate Hypothesis according to which a given grammatical change has the same rate across (grammatical) contexts unto different text forms. This, however, does not hold for the loss of OV (Fig. 2,  $P(\text{Position} = OV | \text{Date}) = \frac{e^{\alpha + \beta * \text{Date}}}{1 + e^{\alpha + \beta * \text{Date}}}$  with  $\alpha = 7.682$  and  $\beta = -0.008$  for the prose and  $\alpha = 0.276$  and  $\beta = -0.0007$  for the verse), which appears to be almost constant for the verse. This suggests that the verse is to be treated separately at least for the evolution of OV, which has not been the practice yet (cf. [7], [3], [2], [8], [4], [1]). The analysis of the effect (e.g. verse being an “artificial” syntactic register etc.) remains for future research.

Figure 1. Predicted probability of overt subject given Date

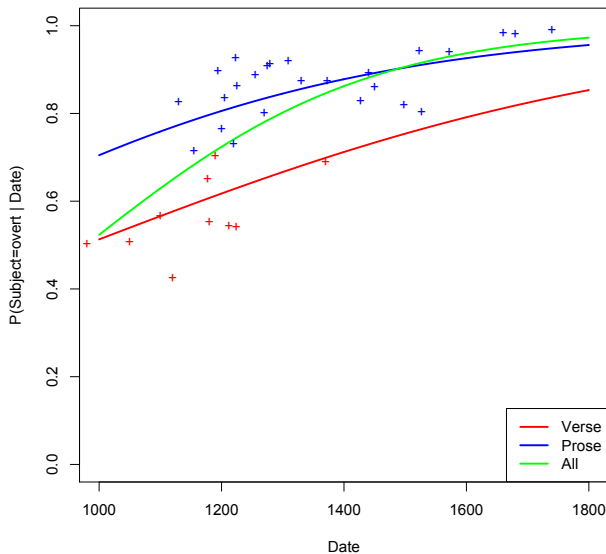
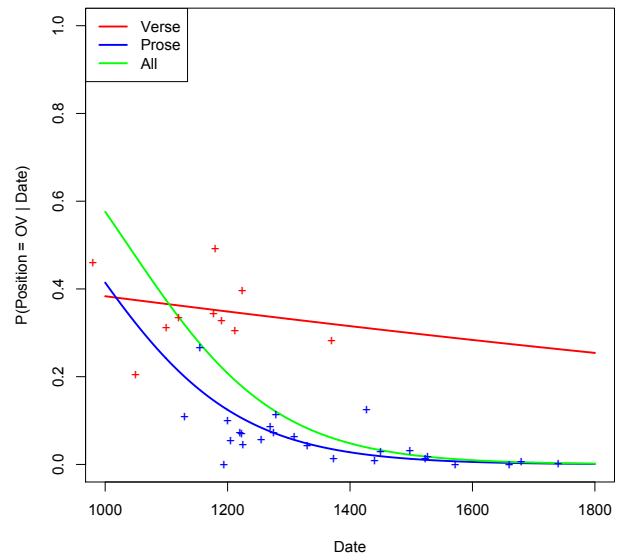


Figure 2. Predicted probability of preverbal direct object given Date



## References

- [1] Anthony Kroch and Beatrice Santorini. On the word order of Early Old French, 2014. *SinFonIJA* 7, Graz, Austria, September 2014.
- [2] Marie Labelle. Clausal architecture in Early Old French. *Lingua*, 117(1):289–316, 2007.
- [3] 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.

- [4] 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.
- [5] Corpus MCVF annoté syntaxiquement, sous la direction de France Martineau, avec Paul Hirschbühler, Anthony Kroch et Yves Charles Morin, 2010.
- [6] Penn Supplement to the MCVF Corpus by Anthony Kroch and Beatrice Santorini, 2010.
- [7] Barbara Vance. *Syntactic change in medieval French*. Studies in Natural Language and Linguistic Theory. Kluwer, Dordrecht/Boston/London, 1997.
- [8] Laurie Zaring. On the nature of OV and VO order in early Old French. *Lingua*, 121:1831–1852, 2011.