

Learning how to count -- a treebank analysis of V2 word order in two Medieval Romance languages through time

As a central issue in syntactic analysis (Greenberg 1963), the analysis of word order has seen a renewed interest with the development of syntactically annotated treebanks (e.g. Liu 2010). This is equally the case in diachronic research, where both PENN and UD annotation systems have led to significant corpora production. Automatic extraction of word order in annotated datasets supports massive comparison across languages, and through time. Caution is however advisable in that results from automatic extraction can provide misleading results. The research identifies, through the conduct of a novel quantitative analysis on the gradual loss of V2 through time, particular configurations that require separate assessments for the results to be reflective of actual V2 word-order. The take-home message is that quantitative data are most valuable only when their investigations is informed by a qualitative analysis of the phenomenon at hand.

We present the result of a comparative analysis of V2 word order in two Medieval Romance varieties, French and Venetian (e.g. Wolfe 2018 and references therein). The protocol relies on a calibrated corpus to enhance comparability of results. The corpus is calibrated for each language with one text per century over the 14th, 15th and 16th century, at temporal intervals of about a hundred years. They are prose texts belonging to a non-literary genre of legal texts that contain traces of dialogal exchanges, and have been found through preliminary investigations to yield less conservative rates of use of changing variables (Larrivée 2022). The annotated versions of the texts are analysed for position of the finite verb in main and subordinates, using parallel extraction queries from the fine-grained PENN annotation set which is sensitive to phrase-structure. The extraction process however raises two types of methodological questions:

- Some configurations relating to a given word order need a separate assessment;
- Some configurations relating to a given word order should be set aside entirely.

On the first point, early Venetian displays an unexpected pattern by which there seems to be more V2 in subordinates than in main clauses. This makes sense once one realizes that this is due to a nearly categorical use of pre-verbal subjects in embedded clauses, irrespective of V-type (transitives, unergatives, unaccusatives). On the contrary, in main clauses we find that subject can be both pre and post-verbal, in between the auxiliary and the past participle, thereby attesting the expected asymmetry between main and embedded clauses typical of a V2 language.

The second point is illustrated by the surprising frequency of V1 word order. As both Venetian and French are expected to go from a V2 system to a SVO word order, the high proportions of V1 is troubling. Again, a qualitative examination of the data shows that the surprising proportions are due to two configurations that should be set aside: (i) coordinated subjectless clauses inside a sentence; (ii) relative clauses.

We conclude that: (a) despite the apparent prevalence of V2 in embedded clauses and V1 in main clauses, Old Venetian is still a totally regular V2 language in the early XIV c., with a clear asymmetry in the subject/verb position between main and embedded clauses; (b) The assessment of default assertive word order requires methodological and analytical decisions about what to count, and what not to.

Time permitting, we will also focus on the diachronic pattern of loss of the V2 property in the two languages. The refined data allow us to better quantify the rate of V2 and its diachronic demise.

References

Greenberg, J. 1966. *Language Universals, with Special Reference to Feature Hierarchies*. The Hague: Mouton.

Larrivé, Pierre. 2022. « Is Medieval French diglossic? New evidence on remnant V2 and register ». Ora Matushansky, Laurent Roussarie, Michela Russo, Elena Soare, Sophie Wauquier (Eds), *Romance Languages and Linguistic Theory (RLLT series: 17)*, *Isogloss* 8,2, 1-16.

Liu, Haitao. 2010. Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua* 120,6, 1567-1578.

Wolfe, Sam. 2018. *Verb Second in Medieval Romance*. Oxford: Oxford University Press.