A New Look at Word Order Change: Constant Rate Hypothesis, Logistic Regression and Bayesian Statistics

Olga Scrivner
Indiana University

While the probabilistic approach is certainly a norm in psycholinguistics, natural language processing and sociolinguistics, the categorical approach is still dominant in historical linguistics, describing only "the extremes" of the phenomenon and excluding the gradient continuum. Given that the period of language change may span over several hundreds of years, finding empirical evidence for language change with limited access to data presents a great challenge.One such challenging linguistic phenomenon is the shift from object-verb (OV) to verb-object word order (VO) in the history of Latin and Old French. From the starting point, we inevitably embrace the core issue of decoding innumerable word order patterns. How to determine the rate of change when usage frequencies are "probabilistically influenced by features of context and situation" (Bock and Kroch, 1989)? In this study, I offer a novel methodology to examine historical data, by combining methodologies from different fields, including probabilistic and corpus linguistic methods.

First, this study focuses on infinitival transitive clauses, as it has been shown that the complexity of word order patterns and the structural ambiguity that obscure our analysis can be overcome through the examination of non-finite clauses (Zaring, 2010; Kroch, 2001). By applying corpus linguistic methods, the data are extracted from annotated corpora spanning over several centuries of Latin and Old French (PROEIL, Perseus, Opera Latina, MCVF and NCA). The infinitival clauses are further split into mono-clausal ("restructuring") and bi-clausal categories (Cinque, 2004). The data are then fitted into logistic regression (logit transform) (Kroch, 1989), where the slope represents the rate of change. The results show that infinitival constructions form three distinct groups: 1) independent infinitives, 2) bi-clausal (prepositional and bridge verbs) and 3) mono-clausal (modal and aspectual). Furthermore, the novel VO form seems to be initially spread through mono-clausal reduced infinitives.

Second, the corpus is codified for various pragmatic, semantic and syntactic factors. To analyze the relationship between VO order and independent variables I implement a Bayesian model, which has recently become a common statistical tool in various fields, such as biology and psychology, among others. While other statistical tools, such as VarbRul, or logistic regressions with SAS, SPSS or STATA, treat each parameter as a fixed entity with *p*-value, in the Bayesian analysis, each parameter is expressed in terms of a posterior distribution, combining information from the likelihood and prior distribution of the data (Kruschke, 2012). The results indicate that VO is a basic word order in Latin, contrary to the traditional view, while OV is a result of stylistic transformations. Furthermore, our study suggests that the origin of VO order seems to be in the reduction of the left periphery of clauses in Late Latin - a change that became evident only through analyzing reduced infinitival clauses.

This study shows that the combination of corpus linguistic, variationist and probabilistic methods yields a better understanding of language change. Furthermore, it shows that the findings obtained from statistical analyses can be fitted into linguistic formal theory, unifying the categorical and probabilistic approaches.

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

Kroch, Anthony. "Syntactic Change." The Handbook of Contemporary Syntactic Theory. Blackwell Publishers. 2001. 699–729.

Kruschke, John. "Bayesian Estimation Superses the t Test."*Journal of Experimental Psychology: General* 142/2 (2013): 573–603.

Zaring, Laurie. "Changing from OV to VO: More evidence from Old French."Ianua. *Revista romanicaPhilologica* 10 (2010): 1–18.