

do-support in the parsed EME corpora: beyond Ellegård (1953)

Aaron Ecay

The University of Pennsylvania

Do-support is the phenomenon whereby, in modern English, the auxiliary verb *do* is inserted in negative sentences, emphatic affirmatives and questions when the sentence otherwise lacks an auxiliary verb. The emergence of this construction in the Early Modern English (EME) period (1300–1700) ranks among the most extensively studied syntactic changes in linguistics.

Ellegård (1953), the first rigorous quantitative study of the change, constructed a large diachronic corpus of potential *do*-support sentences, which enabled patterns to be noticed in the evolution of the change. Some of these patterns were noticed by Ellegård himself – such as the fact that a specific lexical class of verbs (with *know* as the most prominent member) resists *do*-support.

The grammatical analysis of *do*-support worked out in (Roberts 1985) was operationalized by Kroch (1989), who posited the Constant Rate Effect (CRE). This hypothesis states that superficially unrelated patterns which are reflexes of change in a single grammatical parameter will evolve in parallel (in the geometric sense). Kroch thus tied the rise of *do*-support to another change which takes place during the EME period: the loss of word orders where an adverb intervenes between a main verb and its object. These are both diagnostics of the value of a single Verb-Raising parameter (Emonds 1978; Pollock 1989).

Other investigations of *do*-support have focused on sociolinguistic aspects of its evolution. Warner (2005) finds that age-grading and register effects manifest differently at different stages of the change. This fact may explain one of the puzzling aspects of the evolution of *do*-support: that it does not have a smooth upwards trajectory over time, but rather a sharp “kink” around the year 1575, when its rate of use falls abruptly then recovers more slowly.

The work discussed here relies on the corpus collected by Ellegård. With the availability of parsed corpora of the relevant historical stages of the language (Kroch, Santorini, and Delfs 2005; Taylor et al. 2006), opportunities for new research emerge. Fortunately, only 6 % of the parsed corpora consists of material which is found in Ellegård’s corpus, which makes it possible to perform a replication of the earlier results. Broadly speaking, these are confirmed, which lends them empirical strength.

Secondly, the nature of information available in a parsed corpus allows additional hypotheses to be easily formulated and evaluated. Recent work, such as Ecay (2012), has integrated information available in the parsed corpora but not in Ellegård’s corpus. This work has uncovered a richer understanding of fine details, such as the effects of argument structure on *do*-support (noticed by Ellegård in the form of the *know* class). It also has postulated an intermediate grammar where *do* is an agentivity marker, which bridges the gap between Middle English use of *do* as a causative and modern usage which gives *do* no semantic import whatsoever.

The success of parsed corpus data at replicating and extending results gleaned from a special-purpose corpus indicates that general-purpose corpora are capable of generating interesting research results. Additionally, the possibility to quickly replicate and expand on previous results means that corpus-based research can proceed in the incremental fashion characteristic of other sciences.

This presentation will cover the replication of the studies which use Ellegård’s data and their extension using data from the parsed corpora. It is appropriate for advanced students and practitioners of historical syntax. Both theoretical and (abstract) methodological aspects of the investigation will be discussed. R language source and data files comprising to the analysis will be available to participants interested in concrete implementational details.

REFERENCES

- Ecay, Aaron (2012). *Three-way competition and the emergence of do-support in English*. Presented at DiGS14. July 2012. <http://www.ling.upenn.edu/~ecay/digs14/Ecay-DiGS14-slides.pdf>.
- Ellegård, Alvar (1953). *The auxiliary do: The establishment and regulation of its use in English*. Engelska språket.
- Emonds, Joe (1978). The Complex V-V' in French. *Linguistic Inquiry* 9:2, 151–175.
- Kroch, Anthony (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:3, 199–244. ISSN: 0954-3945.
- Kroch, Anthony, Beatrice Santorini, and Lauren Delfs (2005). *Penn-Helsinki parsed corpus of Early Modern English*. University of Pennsylvania. <http://www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-1/>.
- Pollock, Jean-Yves (1989). Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry* 20:3, 365–424.

- Roberts, Ian (1985). Agreement parameters and the development of English modal auxiliaries. *Natural Language and Linguistic Theory* 3:1, 21–58.
- Taylor, Ann, A. Nurmi, Anthony Warner, Susan Pintzuk, and T. Nevalainen (2006). *Parsed Corpus of Early English Correspondence, parsed version*. Compiled by the CEEC Project Team. York: University of York and Helsinki: University of Helsinki. Distributed through the Oxford Text Archive. <http://www-users.york.ac.uk/~lang22/PCEEC-manual/index.htm>.
- Warner, Anthony (2005). Why *do* *dove*: Evidence for register variation in Early Modern English negatives. *Language Variation and Change* 17:3, 257–280.