

Investigating the history of English *do*-support using automatically annotated corpora

Aaron Ecay

March 9, 2015

For a color version of this handout, visit <http://aaronecay.com/papers/brazil-poster-2015/>.

1 *Do*-support

Do-support is the phenomenon whereby English requires the use of the auxiliary verb *do* in sentences where 1. there is not another auxiliary and 2. the adjacency of the V and T heads is interrupted. This interruption can be caused by a head which intervenes between T and the verb, as in (1). In cases of emphatic assertion, as in (2), the head that intervenes can have no segmental content (though it contributes a pitch accent to the phonology). The non-adjacency can also be caused by head movement of T, as in (3) or XP movement of VP (including V), as in (4).

Underlining indicates *do*-support

- (1) Terry does not eat meat.
- (2) Yes, Terry DOES eat meat.
(Capitals = emphatic pitch accent)
- (3) What does Terry like to eat?
- (4) Try though we did to be accommodating, Terry's dietary restrictions came as a surprise to us.

2 The diachrony of *do*-support

Do-support originated in a Middle English causative construction.

Note 1

Though this conclusion is disputed by some. The most common alternative theory is an origin by borrowing from Celtic languages. This proposed explanation has some shortcomings though: it cannot explain the almost 1,000 year time gap between the arrival of Germanic-speaking settlers in Great Britain and the emergence of *do*-support, nor the common features shared by English *do*-support and similar constructions in other Germanic languages. Finally, the Celtic proposal cannot explain the argument structure effects discussed below.

It spread through the language during the Early Modern English (EME) period, from 1500 on. It became mandatory some time in the 19th century.

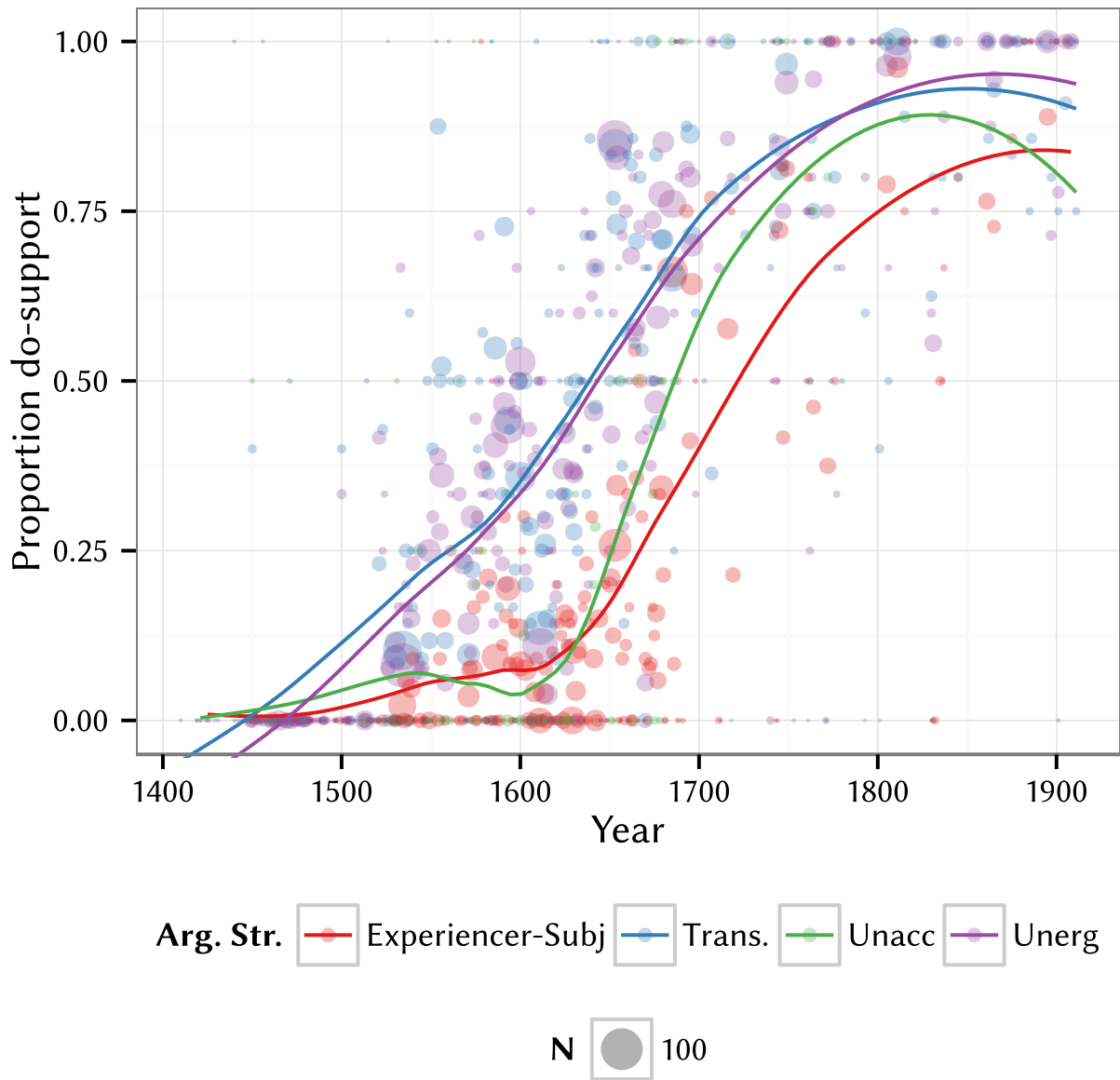


Figure 1: Incidence of *do*-support in negative declaratives of various argument structure types in the PPCHE

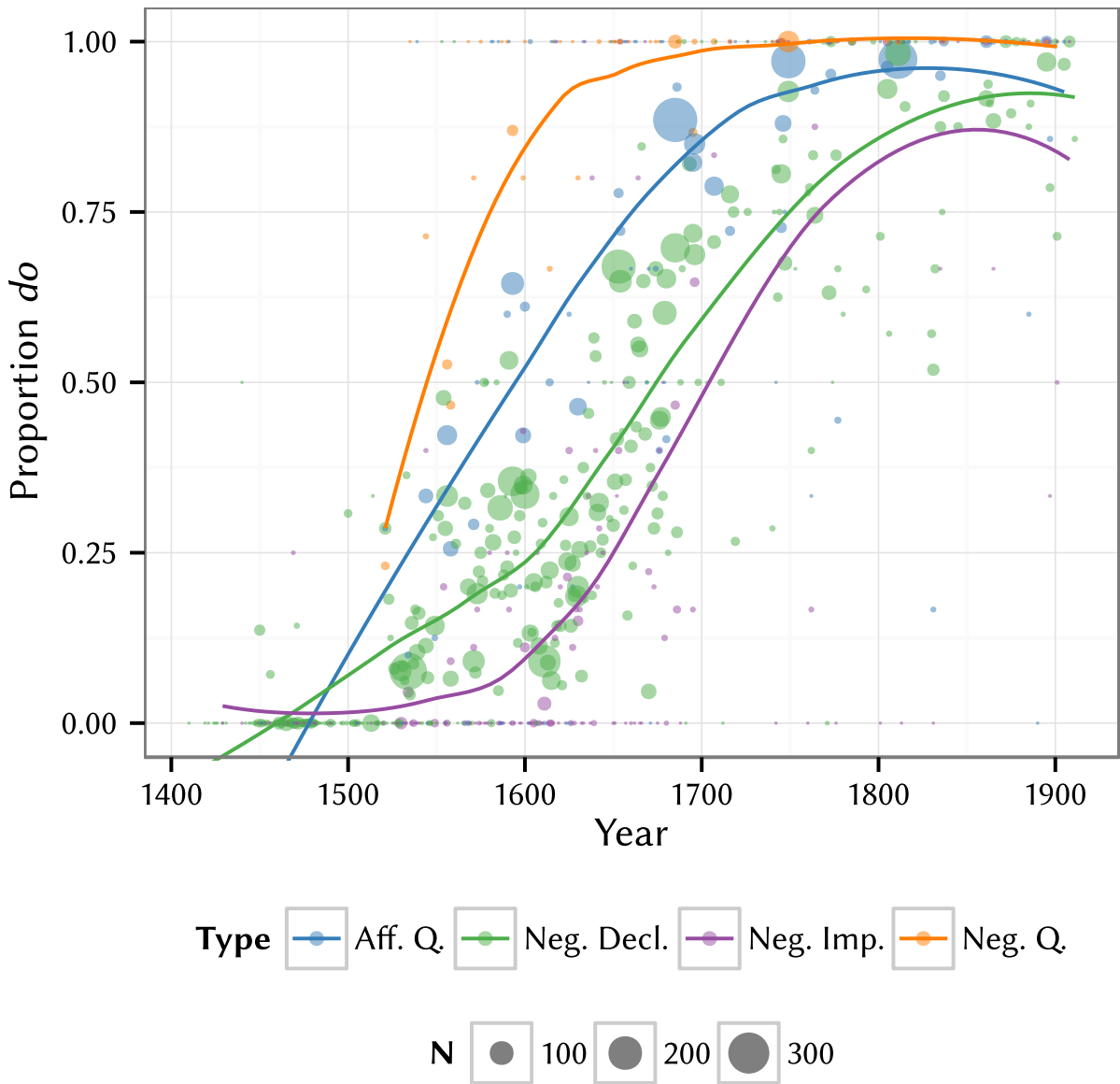


Figure 2: *Do*-support in various contexts, as measured in the Penn Parsed corpora of Historical English (PPCHE)

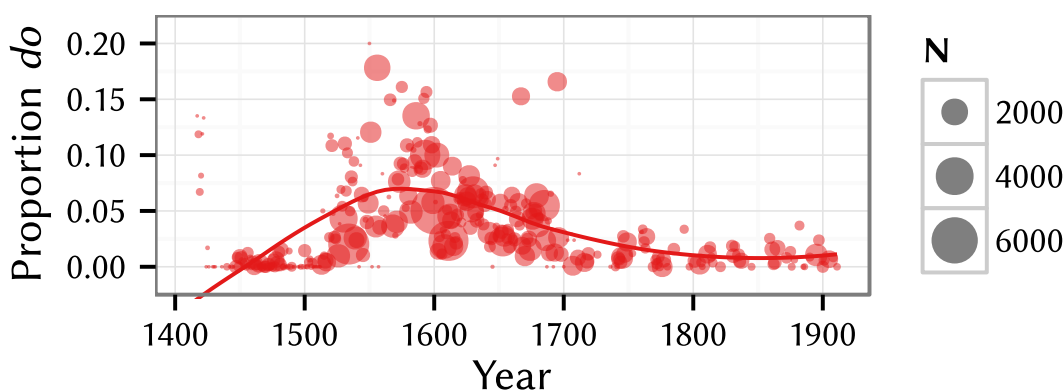


Figure 3: Do in affirmative declarative sentences from the PPCHE.

Note 2

Though some unusually conservative dialects resist *do*-support, at least in questions (thanks to Elyse Jamieson for bringing this fact to my attention). The construction has also been the locus of further innovation, such as the extension of *do*-support with possessive *have*, which began in North America in the early 1800's and has since spread to other dialects.

The chart shown in Figure 2 shows the progress of this construction. The phenomenon was first studied quantitatively by Ellegård (1953). The corpus created by Ellegård was reanalyzed in a variety of later studies. This work represents the first application of parsed corpus data (specifically from the PPCHE) to the study of this construction.

3 Affirmative declarative *do*-support

In addition to its use in the Modern English *do*-support contexts in (1–4), in EME *do*-support was also used in non-emphatic affirmative declarative sentences, a usage which is not allowed in present-day English. This usage peaked around roughly 10% of affirmative declarative sentences, before beginning to decline in the late 16th century before being lost from the language. Its evolution is shown in Figure 3.

4 Argument structure effects

In the process of investigating the diachrony of *do*-support in the PPCHE, I discovered that there is a difference in the usage of *do*-support across different argument structure contexts.

An argument structure effect

Before 1575, *do*-support is not widely used with unaccusative verbs in any context. Furthermore, affirmative declarative *do*-support is never used with unaccusatives.

This fact is illustrated in Figures 4 and 1. It leads to an account of an intermediate grammar of *do* where *do* has been bleached of its causative semantics, but still has agentive semantics. This intermediate grammar is responsible for generating the attested affirmative declarative *do* sentences. By extension, these sentences should not be analyzed as tokens of *do*-support in the modern sense, but rather tokens of the usage of *do* as an agentivity marker.

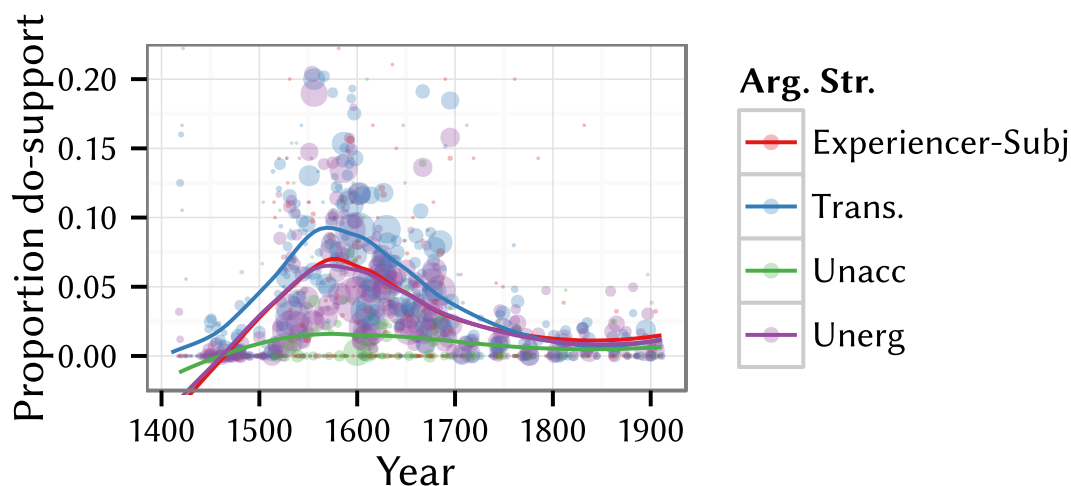


Figure 4: Incidence of *do*-support in affirmative declaratives of various argument structure types in the PPCHE

5 Possible lexical effects

The presentation of the data on argument structure from the PPCHE obscures a fact about the data: it is sparse enough that the so-called argument structure classes are determined by just a few words. Specifically, the experiencer-subject class is dominated by *know*, and the unaccusative class by *come* and (to a lesser extent) *go*. We would like to know whether the properties our analysis imputes to lexical classes are in fact generalizable, or whether they are peculiar to only these lexical items. However, the PPCHE do not contain enough information to investigate the question.

6 A new corpus

Thus, I have constructed a new corpus of EME text, which is much larger than the PPCHE.

PYCCLE

- The Penn-York Computer-annotated Corpus of a Large amount of English
- 1 billion (10^9) words
- Based on the EEBO and ECCO corpora digitized by the TCP
- Annotated with POS tags using a 100% automatic process; PPCEME and PPCMBE used as training data

This corpus is limited by the lack of syntactic information as well as errors (both random and systematic) in the annotation. However, it proves a useful source of information about *do*-support. As demonstrated in Figure 5, the lexical class hypotheses derived from the PPCHE data are not driven by peculiarities of individual lexical items – rather *come* and *go* pattern with other unaccusatives and against transitives, especially in the earliest data.

7 Bibliography

- Ellegård, Alvar (1953). *The auxiliary do: The establishment and regulation of its use in English*. Engelska språket.
- Kroch, Anthony, Beatrice Santorini, and Lauren Delfs (2005). *Penn-Helsinki Parsed Corpus of Early Modern English*. University of Pennsylvania. URL: <http://www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-1/>.
- Kroch, Anthony, Beatrice Santorini, and Ariel Dierani (2010). *Penn Parsed Corpus of Modern British English*. University of Pennsylvania. URL: <http://www.ling.upenn.edu/hist-corpora/PPCMBE-RELEASE-1/>.

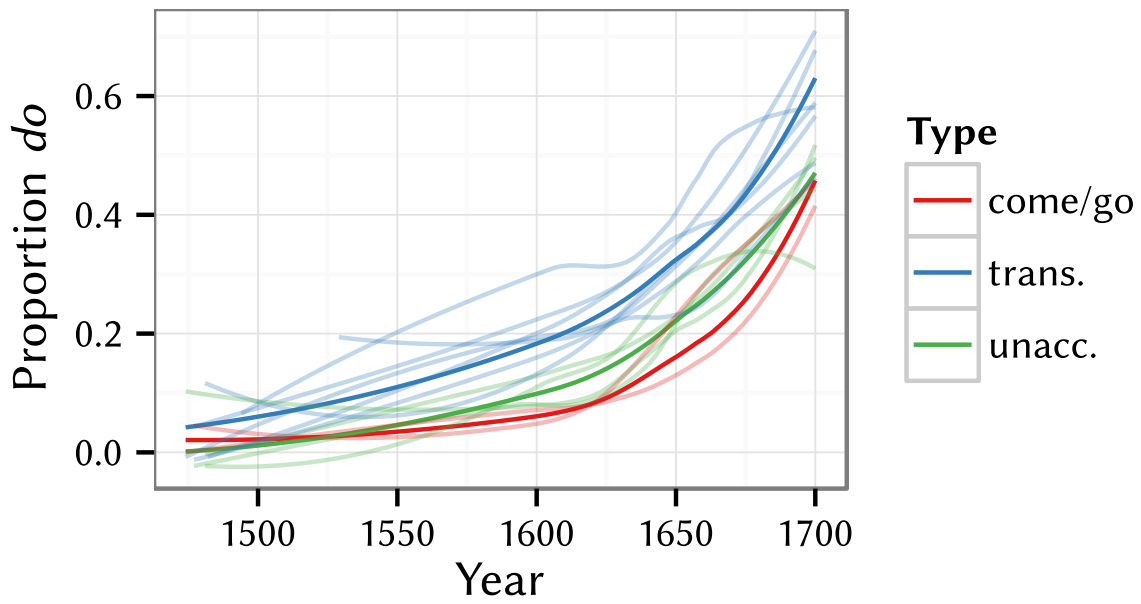


Figure 5: *Do*-support with various lexical items in the PYCCLE. Unaccusatives are *stand*, *fall*, *happen*, and *die*. Transitives are *make*, *give*, *read*, *take*, *find*, *love*, and *receive*. Each faint line represents a single verb, and the darker line is the class-wide average.

Taylor, Ann et al. (2006). *Parsed Corpus of Early English Correspondence*. Compiled by the CEEC Project Team. Distributed through the Oxford Text Archive. URL: <http://www-users.york.ac.uk/~lang22/PCEEC-manual/index.htm>.