The rise of the 19th century English progressive: variation between individual verbs

Alina Ladygina and Igor Yanovich

DFG Center for Advanced Studies: Words, Bones, Genes and Tools, University of Tübingen

Introduction. We examine the development of the English progressive for the 42 most frequent lexical verbs in the speech of the British Parliament throughout the 19th century. Using exploratory data analysis and mixed-effects regression modeling, we find that individual verbs show distinctive trajectories of change that cannot be obviously predicted by general linguistic features known to influence the construction. Previous studies would not detect this because our dataset is 2 orders of magnitude larger, allowing us to consider individual verbs in detail. We conclude that in the rise of the progressive, "each word has its own history".

Background. The progressive as it exists today is a relatively young feature of English. Although already established in Early Modern English, it only became obligatory in "semantically progressive" contexts in the 20th century [Kranich 2010]. The 19th century was crucial for the process of the progressive integration [Smitterberg 2005], with the overall frequency of the progressive greatly increasing throughout the period [Strang 1982], [Denison 1998: Sec. 3.3.3].

Motivation. Our aim is to see how the rise of the progressive plays out on the level of individual lexical verbs: which actual trajectories of change different verbs have. To be sure that we are dealing with true trajectories coming from a single speech community, as opposed to aggregated from heterogeneous usage by individuals who have never been in linguistic contact with each other, we use the debates of the British Parliament as our source. Though MPs come from different places and have different life histories, we can at least be sure that they were in regular linguistic contact with each other. The specificity of speech situations in the parliament helps, ensuring that different speakers in the corpus attempt to use roughly the same register in our texts. While parliamentary debates represent formal speech and thus could in principle show fossilized linguistic norms, we do observe clear changes in the usage throughout the 19th century, Fig. 1. Importantly for us, considerable variation by verb (gray) is clearly seen around the general trend (black).

Methods. The debates from the Hansard Archive from 1803 to 1889 were divided by calendar decade, and POS-annotated by TreeTagger [Schmidt 1994]. We identified 50 most frequent verbs over the whole period and extracted their finite progressives using regular expressions (N=89553 occurrences of the progressive in total, cf. N=53 for debates in [Smitterberg 2005]). The verbs for which a high proportion of non-progressive forms, e.g. predicative adjectival participles (*That correction is wanting*) was extracted, were discarded, leaving 42 lexical verbs for which the number of false positives was negligible. Instances of *BE+going* followed by *to* were also excluded, as that combination usually is a futurate. Our metric was the number of *BE+ing* hits for the verb divided by the overall frequency of the verb in the corresponding time period. While this is only a rough proxy to the variationist variable we are interested in, namely the proportion of progressives in contexts where one would be possible, a proper delineation of that sociolinguistic variable would involve analyzing by hand millions of contexts to arrive at a comparably sized dataset.

Analysis. Our 42 verbs can be divided by trajectories they exhibit into those where the progressive's share increases (Fig. 2), decreases (Fig. 3) or stays largely stable (Fig. 4) throughout the period. The verbs within these three classes do not show obvious semantic or aspectual similarities. Conversely, verbs with similar features do not behave alike. E.g., the verbs of movements or speech demonstrate high variability in the relative frequency of progressives (Fig. 5-6). Interestingly, lexical aspect does not play a clear role either: even statives show a heterogeneous pattern. In general, they rarely occur in the progressive, but one of them – the verb *suffer* – shows a dramatic increase in relative frequency.

The informal observations above are supported by regression modelling. We fit the following set of mixed-effects linear models[1], all with fixed effect of time: 1) varying intercept for individual verb (AIC=-2053.6, logLik=1030.8); 2) the same plus varying slope for individual verb (AIC=-2132.0, logLik=1071.0); 3) var. intercept and slope for "trajectory class", according to the division into verbs in Fig. 2-4, plus var. intercept for individual verb (AIC=-2104.4, logLik= 1058.2); 4) var. intercept and slope for movement vs. speech (claimed to be leaders of change) vs. other verbs, plus var. intercept for ind. verb (AIC=-2049.6, logLik= 1030.8). Not surprisingly, Model 2, which assigns a "random" slope to each verb, is the best. But importantly, allowing slopes to vary by "trajectory class", Model 3, is clearly helpful, while doing the same for semantic classes, Model 4, does not improve even the baseline model that didn't have any varying slopes. We do not suggest that "trajectory classes" are a real phenomenon. Rather their relative statistical success illustrates that differences between individual verbs cannot be obviously traced to any general linguistic features.

---

[1] Normally, logistic regression would be more appropriate for such counts data as ours. But in our case it is not, for two reasons: 1) we only observe the overall frequency of the lexical verb, not the # of contexts where the progressive is possible, hence our counts do not come from a binomial distribution; 2) the trajectories in Fig. 6 obviously do not uniformly show the S-curve shape.

Conclusion. The history of individual verbs in the rise of the progressive does not appear reducible to the influence of general linguistic factors. **Individual word histories are significant**. Future analyses should thus strive to distinguish true class effects from spurious ones driven by individual items (e.g. by *suffer* for statives.)

In our debates data, the membership of each verb in its "trajectory class" appears generally stable throughout the 19th century. But we do not know whether such membership is stable across different genres and speech communities. It remains for future research to show whether individual verb histories that we see in the parliamentary debates generalize across text types and speech communities, or belong specifically to some of those, but not others.

References

Hansard Parliamentary Debates (1st-3rd Series), February 1820 - August 1891. http://www.hansard-archive.parliament.uk/

[Denison 1998] David Denison (1998). Syntax. In: *The Cambridge History of the English Language, Vol IV: 1776-1997*, ed. Suzanne Romaine, CUP.

[Kranich 2010] Svenja Kranich (2010). The progressive in Modern English. A Corpus-Based Study of Grammaticalization and Related Changes. Issue 72 of Language and Computers - Studies in Practical Linguistics. Amsterdam: Rodopi.

[Schmidt 1994] Helmut Schmid (1994): Probabilistic Part-of-Speech Tagging Using Decision Trees. Proceedings of International Conference on New Methods in Language Processing, Manchester, UK.

[Smitterberg 2005] Erik Smitterberg (2005). The progressive in 19th-century English: A process of integration. Issue 54 of Language and computers: studies in practical linguistics. Amsterdam - New York, NY, Rodopi, 2005

[Strang 1982] Barbara M.H. Strang (1982). Some aspects of the history of the be + -ing construction. In: *Language Form and Linguistic Variation: Papers dedicated to Angus McIntosh*, ed. John Anderson, Amsterdam: John Benjamins.

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6