
Danielle Barth
University of Oregon
dmathieu@uoregon.edu

Modeling reduction of *is*, *am* and *are* in grammaticalized constructions

Introduction

Grammaticalization is a type of language change characterized by the creation of a grammatical element from a lexical element or another grammatical element. A classic example is the development of the English future auxiliary *will* from the lexical verb meaning ‘want’. In the grammaticalization literature, phonetic reduction is often listed as a potential, although certainly not necessary, consequence to grammaticalization. For example, the auxiliary *will* can reduce to *'ll* as in (1), but the lexical verb cannot as in (2).

- (1) *I'll see you later,*
- (2) **I'll it to be so.*

However, phonetic reduction can be due to many factors, including frequency of use (Bybee 2007), low semantic weight (Heine 1993, Lehmann 1995), lexical class (van Bergem 1995) and context of use, etc. While changes in all of these parameters are associated with grammaticalization, they do not always go together. This paper examines a case where the less grammaticalized source construction is more frequent than the more grammaticalized target construction.

The Copular, Progressive and Passive Constructions

This paper examines the particular case of *be*. The forms *is*, *am* and *are* can reduce to *'s*, *'m* and *'re* in both the lexical source construction of the copula as the main verb (3) and two grammaticalized constructions of progressive (4) and passive (5) as the auxiliary.

- (3) *She's a welder.*

- (4) *She's working.*
- (5) *She's seen.*

The copular source construction is more frequent than the grammaticalized progressive and passive constructions as seen in a search in the Corpus of Contemporary American English (COCA) (Davies 2008-). All three variants could be characterized as having very little semantic weight because they are all grammatical elements. So this leads to the research question: Does one of these three construction types exhibit reduction more often, and if so which? And to follow up: what factors could influence the reduction of *is*, *am* and *are*? Following grammaticalization theory would predict that the passive and progressive constructions should display more reduction than the source copula construction. If frequency influences reduction then the copula construction should show the most reduction and the passive construction should show the least.

Modeling Reduction

The spoken section of the COCA, which contains transcripts of American TV and Radio programs with spontaneous speech, was used to build a sample of 3080 entries coded for presence/absence of reduction based on the orthographic transcription. A range of independent variables were coded by hand. Reduction was modeled using mixed-effects logistic regression. Random factors included the speakers and programs in which the target of interest occurred. Fixed factors included construction type, transitional frequency of collocates, log frequency of collocates and various phonetic and syntactic variables.

A series of models were built exploring the effects of these factors. Model comparison was used to select the simplest model that was as predictive as the most complex model. Models were compared with log likelihood tests and bootstrapping was done on logistic regression models with the random factors removed.

This paper presents the final four models: one for the combined results of *is*, *am* and *are* data and one for each individual verb-auxiliary set. The final models indicated that construction type played a role in influencing reduction, but in a more complicated way than expected. The progressive construction had the most reduction and was significantly different than either of the other two construction types. The copular and passive constructions were not significantly different from each other in any model except for the model of *am* only data.

Other significant factors included the type of subject NP, the preceding and following phonemes, the length of the preceding NP, preceding utterance and preceding word length, the occurrence of a reduced variant of *be* in the preceding utterance, log frequency of collocates and transitional probability of following collocates. The full model has a C index of concordance of .91, above the .8 threshold that Baayen (2008) argues is needed for adequate predictive capability. Models for each verb-auxiliary set also have C's above .8.

These results indicate that neither frequency nor grammaticalization alone provides adequate explanation in this case. The tendencies of grammaticalization favor the progressive construction in comparison to the copular construction and the tendencies of frequency favor the progressive construction in comparison to the passive construction.

Further analysis was done, coding separately from the corpus instances of the words *going* or *gone* following the target variant of *be*. This filtered out most cases of

the future construction, which appeared to be a factor influencing the results and was also the most frequent word following a variant of *be* in the progressive construction. In this subset of the data, the progressive still had significantly more reduction than the passive construction. The progressive also still showed more reduction than the copular in the full and *is* models.

Together these results show that, at the very least, reduction of *is*, *am*, and *are* varies by construction type and this pattern is robust. Frequency plays some role, because the least frequent construction type displays the least reduction. However, frequency is not enough to account for the highly frequent copular construction having less reduction than the progressive construction, even when instances of the future construction are coded separately. Grammaticalization also plays a role here because the grammaticalized progressive displays more reduction than its source copular construction. In the case of the future construction, it displays the most reduction and is more frequent than its source spatial construction, which is a classic example of grammaticalization. Neither grammaticalization nor high frequency guarantees that reduction will take place. Reduction is simply a possible outcome. What this study shows is that both are needed to account for the reduction of *is*, *am*, and *are*.

References

- Baayen, R. H. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- van Bergem, D. 1995. *Acoustic and lexical vowel reduction*. Amsterdam: IFOTT.
- Bybee, J. L. 2007. *Frequency of use and the organization of language*. Oxford & New York: Oxford University Press.
- Davies, Mark. (2008-) The Corpus of Contemporary American English (COCA): 410+ million words, 1990-present. Available online at <http://www.americancorpus.org>.
- Heine, B. 1993. *Auxiliaries*. Oxford: Oxford University Press.
- Lehmann, C. 1995. *Thoughts on Grammaticalization*. München, Newcastle: Lincom Europa.
-