The first decade of *because*-NP: 2007–2016

We examine the *because*-NP construction (WOTY, 2013) using a large Internet corpus and an online survey. *Because*-NP is notable as a modern, primarily Internet-based syntactic development. However, much is still unknown about its origins, structure, and sociolinguistic correlates (cf. Bailey 2014; Blamire 2017; references therein).

We expand upon Bland, Raess & Baclawski’s (2016) study by including a larger sample of data, consisting of the entire Reddit Corpus from 2007–2015 and the entire Twitter Stream Grab Corpus from 2011–2016, analyzed in monthly chunks. Together, these account for over 350,000 tokens of *because*-NP. The results show a strong S-curve in both Reddit and Twitter: *because*-NP arose in 2011 and experienced a dramatic rise in late 2011–2012 before leveling off (Fig. 1–2). Intriguingly, even though *because*-NP was a shibboleth of Internet speech around 2012, its usage has remained steady since then.

In order to examine attitudes toward *because*-NP, we also report the results of an Amazon Mechanical Turk survey of 150 native American English speakers. Participants were randomly assigned tokens of *because*-NP and non-elided tokens (e.g. *because of* NP) adapted from natural corpus examples. They were then asked to rate how likely they would be to say or hear/read the tokens, as well as the perceived age, gender, nationality, and medium (online vs. in-person) of the token’s speaker. Participants also completed a short sociolinguistic questionnaire on their background and Internet usage. Results align with the construction’s novelty and Internet basis. *Because*-NP is quite stigmatized—participants reported they were far less likely to utter the tokens than to hear them from others. It is also significantly associated with young and online speakers. Participants who reported a high use of Wikipedia were significantly unwilling to accept *because*-NP, suggesting a formality effect. Finally, tokens of *because*-NP that alternated with a version with *of* were rated significantly higher than those with other elided elements (e.g. *I’m, there are*). This lends credence to Blamire’s (2017) hypothesis that *because*-NP is a case-licensing phenomenon. However, several tokens that were rated highly in our survey (e.g. *because my NP*) do not align with previous hypothesis.