

Using Semantic Vector Space Models to investigate lexical replacement – a corpus based study of ongoing changes in intensifier systems

Martin Schweinberger

Universitat Hamburg

This study takes a corpus-based approach to examining co-occurrence patterns of amplifying intensifiers and adjectives (cf. 1) based on two components of the International Corpus of English (ICE). From a language variation and change perspective intensifiers are particularly interesting as they play a crucial part in the “social and emotional expression of speakers” (Ito & Tagliamonte 2003: 258) and because intensifier systems are prone to change (Ito & Tagliamonte 2003:257; Quirk et al. 1985:590). In addition, intensifiers allow for testing conditions of successful lexical replacement which is the topic of this paper.

- (1)
 - a. It’s a *very elegant* technique (ICE NZ: S2A-038)
 - b. oh wow that’s *really cool* (ICE NZ: S1A-096)
 - c. she looks *bloody old* in that picture any rate (ICE NZ: S1A-096)

The paper investigates whether vector space models (cf. e.g. Levshina 2015) can be used to predict successful lexical replacement in observable ongoing changes in intensifier systems. To elaborate, D’Arcy (2015) showed that the New Zealand English (NZE) intensifier system is currently undergoing a change in which the intensifier *really* replaces the intensifier *very*. In contrast, Schweinberger (2016) has shown that in Irish English (IrE), *very* is also waning but *really* is not replacing this traditional intensifier. The hypothesis of the present paper is that in case of lexical replacement (as in NZE) the innovative and the waning intensifier types should be highly similar in terms of their semantic space while these intensifier types should be semantically dissimilar in IrE which then would suggest that semantic dissimilarity blocks the potential innovative type (*really*) from replacing the traditional variant (*very*).

To extract all adjectives, the corpus data was POS-tagged by implementing a maximum entropy part-of-speech tagger. For each adjective, it was determined whether or not it was intensified and which type of intensifier occurred. The statistical analysis applies semantic vector space models to investigate the above stated hypothesis by determining similarities between collocational profiles of intensifiers and adjectives.

The results of the statistical analyses show that the co-occurrence patterns, i.e. the semantic space, of *really* and *very* in NZE are indeed highly similar thus which is taken to allow successful lexical replacement while the co-occurrence patterns of *really* and *very* in IrE are distinct which prohibits lexical replacement. The results are interpreted to suggest that in order for innovative variants to successfully replace traditional variants, it is not enough for innovative variants to attach to certain social groups (association with younger (female) middle

class speakers) as shown by Labov (2002) but that in addition innovative variants must show a certain intra-linguistic profile, i.e. a high degree of semantic similarity with the replaced waning variant, for lexical replacement to be successful.

The results of the semantic vector space models of co-occurrence pattern of *very* and *really* in NZE and IrE thus pose the question of whether there is a more general trend at work in lexical replacement: namely that lexical replacement requires an association with extra-linguistic social factors and intra-linguistic similarity.