Making sense of fake's fickleness: The role of context

Bert Cappelle\(^1\), Pascal Denis\(^1,2\) and Mikaela Keller\(^1,2\)
\(^1\)Université de Lille, \(^2\)Inria Lille – Nord Europe
{bert.cappelle/pascal.denis/mikaela.keller}@univ-lille.fr

Keywords: corpus linguistics, privative adjective, modification, word and phrase embedding, context-sensitivity

Few words have enjoyed more recent attention than fake, especially (but not exclusively) in combination with news. Not only has fake featured prominently in word-of-the-year contests around the world, it has also been treated extensively in the linguistic literature. Formal semanticists have classified fake as a so-called ‘privative’ adjective, for which sentences of the form No Adj N is (an) N are always true. We here contest that fake is an across-the-board privative adjective. For instance, while a fake beard is not, actually, a beard, a fake article is most definitely still an article. We suggest that privativity may be a side effect of the core meaning of fake, which involves an act whereby someone intends to deceive (Lakoff & Johnson 1980: 121-122; Coulson & Fauconnier 1999, Taylor 2003: 96). This still raises the question of how humans effortlessly manage to interpret a fake+noun combination correctly.

Using a distributional-semantic approach (Turney & Pantel 2010), we first demonstrate that the overall semantic contribution of ‘privative’ adjectives (artificial, fake, false, fictitious) cannot be detected straightforwardly as a general distancing effect. We here use ‘embeddings’, that is, dense vector representations based on word co-occurrences in a large corpus, which in our study is a dump of the entire English Wikipedia. A plausible hypothesis would be that the contexts of an adjective-noun combination are more different from the contexts of the noun when the adjective is such a ‘privative’ one than when it is an ordinary (subsective) one, like red. Comparing the cosine distance between the adjective-noun bigram and single noun embeddings across two sets of adjectives, privative and ordinary ones, we fail to find a noticeable difference. This suggests that a classification of adjectives into these two sets (privatives and non-privatives) is too coarse. We therefore extend a recent proposal (Del Pinal 2015) involving the noun’s qualia roles (how an entity is made, what it consists of, what it is used for, etc.; Pustejovsky 1995) and propose several interpretational types of fake-noun combinations, some but not all of which are privative. These interpretations, which we assign manually to the 100 most frequent fake-noun combinations in the Wikipedia corpus, depend to a large extent on the meaning of the noun, as combinations with similar interpretations tend to involve nouns that are linked in a distributions-based network. When we restrict our focus to the private uses of fake only, we do detect a slightly enlarged difference between fake + noun bigram and noun distributions compared to the previously obtained average difference between adjective + noun bigram and noun distributions. This result contrasts with negative or even opposite findings reported in the literature (Boleda et al. 2012, 2013).

Our study shows that the precise interpretation of fake is highly sensitive to the kind of noun it combines with. We argue that a cognitive-linguistic, usage-based approach is well equipped to face this fact about fake. Further discussion of authentic examples (e.g., of fake gun referring, in a crime film, to a real gun that was not used as a murder weapon but was planted on someone) demonstrates that speakers also have to bring to bear highly detailed frame-semantic knowledge of the world and knowledge of the wider discourse context.

References