# Facing the facts of *fake*: a distributional semantics and corpus annotation approach

Bert Cappelle, Pascal Denis and Mikaela Keller

Université de Lille, Inria Lille Nord Europe

*Fake* is often considered the textbook example of a so-called 'privative' adjective, one which, in other words, allows the proposition that '(a) fake x is not (an) x'. This study tests the hypothesis 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. We here use 'embeddings', that is, dense vector representations based on word co-occurrences in a large corpus, which in our study is the entire English Wikipedia as it was in 2013. 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. However, we contest that *fake* is an across-the-board privative adjective, since a *fake article*, for instance, is most definitely still an article. We extend a recent proposal involving the noun's qualia roles (how an entity is made, what it consists of, what it is used for, etc.) 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 privative 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.

Keywords: *fake*; privative adjective; modification; distributional semantics; bigram; word and phrase embedding; context-sensitivity

## 1. Introduction

On January 5th 2018, the venerable 129-year-old American Dialect Society chose *fake news* as its Word of the Year for 2017.[1] This phrase – not a word in the strict sense but nonetheless a "vocabulary item" – was not new but had in the past year gained new prominence and notoriety because of its repeated use, or misuse, by the newly elected president of the US. The American Dialect Society's press release stated that the phrase now in fact had two definitions: no longer just "disinformation or falsehoods presented as real news", but also

[1] *Fake news* had previously been named Word of the Year 2017 by the British dictionary publisher Collins and been nominated, again by the American Dialect Society, as Word of the Year for 2016, along with, among others, *post-truth* and *facticide*. In Germany, *fake news* was voted 2016 Anglicism of the Year (Stefanowitsch 2017), following the nomination of the affixoid *fake-* in 2013 (Flach 2014).

"actual news that is claimed to be untrue".[2] As Ben Zimmer, chair of the society's New Words committee, explains, "[w]hen President Trump latched on to *fake news* early in 2017, he often used it as a rhetorical bludgeon to disparage any news report that he happened to disagree with." While this may be the case, it does not warrant positing the new definition above, as this would mean, nonsensically, that Donald Trump frequently attempts to put down uncomfortable divulgements about his person or policy by calling them "actual news that is claimed [by him] to be untrue". In fact, when Trump dismisses a damaging fact as *fake news*, for his strategy to work, he can only be taken to use that phrase in its first (and only) sense. Of course, one could then accuse him of using that phrase dishonestly, but this is hardly a matter with which lexicographers should concern themselves.

The example of *fake news* does hint at two intriguing linguistic issues concerning the adjective *fake*. First, by saying of something that it is *fake news*, the speaker wants to distinguish it from what is considered to be *real news*. This adjective is therefore known as a 'privative' one, like *artificial*, *fabricated*, *false*, *fictional*, *fictitious*, *imaginary*, *phony*, and *mythical* (Nayak et al. 2014), all of which have a similar kind of 'negative' meaning.[3] (*Privative* comes from the Latin verb *privare*, which means 'rob', 'remove', 'strip' and which also lies at the origin of the English verb *deprive*.) Second, even though in the scholarly literature, *fake* is highlighted as the prime example of privative adjectives, in fact it is not so clear that *fake news* falls outside the set of entities – or aggregation of stuff – that we can legitimately refer to as *news*. Isn't *fake news* also *news*, but then news that contains falsehoods? If we briefly examine a set of corpus examples of *fake*-noun combinations, we soon find that in some cases, a *fake x* is clearly not an *x*, while in other cases, a fake *x* actually *is* an *x*, and in other cases still, a *fake x* is a controversial instance of *x*. For instance, most will agree that a *fake kidnapping* is not a real kidnapping and that a *fake video* (one that was doctored) is still an actual video. For a *fake passport*, opinions about class membership to the set of passports may remain divided – as they are among the authors of this paper. In any case, the addition of *fake* to a noun does not uniformly have the effect of denying that the referent of the noun phrase belongs to the noun's extension (i.e., the range of things that fall under the noun's label).

The goals of this paper are three-fold, the first two goals tying in with each of the above two issues separately and the third one treating the two issues together. First, we aim to test the assumption that adjectives (Adjs) with privative uses modify nouns (Ns) more drastically than do adjectives without such uses. The rationale behind this assumption lies in a well-known classification of adjectives between subsective and non-subsective ones (Kamp 1975; Kamp and Partee 1995; Partee 1995). As we will explain in some more detail below, privative adjectives are grouped in the category of non-subsective adjectives, which means that (when they are indeed used privatively) they do not allow one to make an inference of the type 'If *x* is an Adj N, then *x* is a N'. A reasonable expectation, then, is that the contexts in which we encounter an Adj+N sequence differ more greatly from the contexts in which we find the noun (modified or not) when the adjective is privative than when it is subsective. In a previous study, Boleda et al. (2013) failed to find such a distributional difference but their set of non-subsective adjectives included more adjectives than just privative ones, which are an extreme case.

[2] https://www.americandialect.org/fake-news-is-2017-american-dialect-society-word-of-the-year, accessed 30 May 2018.

[3] Even *alternative* can nowadays be used as a privative adjective when it is combined with the noun *fact*.

Our second aim is to get an idea of how often *fake* is used privatively and non-privatively, and whether this depends on the kind of noun *fake* modifies. Our intuition is that all uses of *fake* involve deceptive appearances and that privativity is a side effect of the semantic properties of the noun with which *fake* combines. Commenting on the question "whether a fake gun is or is not a gun", Taylor (2003: 96) writes:

> "In some respects it is, in other respects it isn't. (Recall the Aristotelian law of contradiction!) A fake gun is a gun to the extent that it possesses many of the attributes of a real gun; it looks like one, and performs at least some of the functions of a gun, e.g. it can be used to intimidate people. Yet a fake gun fails to perform a presumably essential function of a gun, i.e. it doesn't shoot. However, a real gun which fails to shoot, e.g. because it is rusty, or simply because it is unloaded, does not for this reason become a fake gun. Neither is a toy gun, which also fails to shoot, a fake gun. A fake gun has to have been constructed so that its appearance is such that it can be used with the intention of deceiving." (Taylor 2003: 96; cf. similar ideas previously expressed by Lakoff and Johnson 1980: 121-122)

We, too, consider the notion of deception central to the meaning of *fake*. In combination with the noun *gun*, which is defined explicitly as a weapon that shoots bullets, we obtain a privative combination – here, we are less vacillating than Taylor. By contrast, in combination with the noun *article*, which is simply defined as "a piece of writing included with others in a newspaper, magazine, or other publication",[4] the deception pertains to the verisimilitude of what is written and the way the author creates an aura of trustworthiness; since *fake* in this case does not affect any essential (i.e. definitional) attribute of what is denoted by the noun, we can interpret its use here as non-privative: a fake article is still an article. The question, in other words, is whether privative and non-privative uses of *fake* can be related more generally to semantically-based classes of nouns, such as those with and those without an explicit function as part of their conceptual meaning, among other potentially relevant semantic factors. In addressing this research question, we will take Del Pinal's (2015) analysis of *fake* as our point of departure. According to Del Pinal, *fake* operates in a predictable, invariant way on the noun's 'qualia' (i.e., essential attributes of what the noun denotes: how it comes into being, what it looks like, what it consists of, etc). Our approach in essence puts this analysis to the test, by checking whether it is really the case that *fake* always has the same effect with different kinds of nouns. After all, the example of *fake gun* versus *fake article* above suggests that different semantically-based types of noun may require, at the very least, different adjustments to a single basic interpretation that we assign to a *fake*-noun combination. To find out whether this assumption is correct, we will annotate a sample of different corpus-extracted *fake*-noun combinations.

A third and last aim of this paper is to find out whether, distributionally speaking, the modifier *fake* removes the adjective-noun combination further from the noun head when its use is privative than when its use is non-privative. Here, we are no longer looking for differences in distributional effects between adjectives that can be used privatively and adjectives that cannot (cf. our first aim) but between privative and non-privative uses of a single adjective, namely *fake* – uses which may be linked with noun semantics (cf. our second aim). What we aim to test is whether a noun such as *gun* occurs in contexts that are *more*

[4] https://en.oxforddictionaries.com/definition/us/article, last accessed 29 June 2018.

different from the contexts in which people use *fake gun* than the contexts of a noun such as *article* are different from the contexts of *fake article*, given that a fake gun is not actually a gun but a fake article is still an article.

In this study, we mix a distributional-semantic approach, which rests on the now generally agreed assumption that linguistic items occurring in similar contexts in a sufficiently large corpus of naturally occurring data have similar meanings (Harris 1954, Firth 1957, McDonald and Ramscar 2001, Sahlgren 2008), and manual annotation of some frequent *fake*-noun combinations extracted from Wikipedia. Let us summarize our three research aims and the methodology we use to tackle them.

- Do we find a larger distributional difference between Adj+N and N with 'privative' adjectives (e.g. *artificial*, *fake*) than with 'non-privative' adjectives (e.g. *nasty*, *stable*)?[5]
  Methodology: distributional semantics
- How many of the most common *fake*-N combinations are typically privative and does privativity or lack thereof depend on the semantics of the N?
  Methodology: manual annotation of corpus examples + distributional semantics
- Do we find a larger distributional difference between *fake*-N and N with privative uses of *fake* (e.g. as in *fake gun*) than with non-privative uses of this adjective (e.g. as in *fake article*)?
  Methodology: manual annotation of corpus examples (re-used) + distributional semantics

This paper is organized as follows. By way of background, Section 2 first provides a rapid overview of adjective classes. We then demonstrate that there are problems with the standard view according to which *fake* is (without further qualification) a 'privative adjective'. Next, we discuss Del Pinal's (2015) proposal, on which our paper builds, which relates the meaning of *fake* to the noun's qualia structure. In section 3, we describe in more detail our methodology to address our above-mentioned research objectives. Section 4 presents our results and section 5 discusses them. In section 6, we offer a conclusion and suggest some issues for further research.

## 2. Background: classification, problems, suggested solution

### 2.1 Standard classification of adjectives in general and of *fake* in particular

Adjectives are naïvely thought of as words that 'describe' or 'say something more' about an object. If something is referred to as a *red ball*, for instance, then we know that it's a red thing in addition to being a ball. Not so with *fake*. When we hear of an object that it is a *fake gun*, we typically cannot conclude that it is both a fake thing and a gun – rather, precisely because it is fake, we are led to conclude that it is *not* a gun. Previous work in linguistics (Kamp, 1975; Kamp & Partee, 1995; Partee, 1995) has resulted in a four-way classification of

[5] By 'privative adjectives', we mean adjectives that have a privative effect in combination with certain nouns, in certain contexts. So, while we recognize that adjectives such as *fake*, *false*, *artificial*, etc. are not privative in all their uses, we still consider 'privative adjective' a useful shorthand.

adjectives, based on such inferential properties. We here give a brief survey of each of the classes in this standard taxonomy.

First, *red* belongs to the class of 'intersective' adjectives, along with such other examples as *round* and *carnivorous*. An intersective adjective expresses a property that is not semantically tied up with the meaning of the noun it happens to be found with. Consequently, we can allow this adjective to 'skip' from one noun to another noun that can also be used to characterize an entity, without thereby changing the truth value of the sentence. For example, if we know that Tasmanian devils are *carnivorous animals* and if we then learn that they are marsupials, we are allowed to infer that they are *carnivorous marsupials*. Second, adjectives such as *good*, *enthusiastic* and *skillful* belong to the class of (non-intersective) 'subsective' adjectives. In contrast to intersective ones, these closely interact with the meaning of the noun they modify. For instance, given the information that Tasmanian devils are *good swimmers*, the additional information that they are also hunters does not allow us to update our knowledge with the proposition that they are *good hunters*. Just because Tasmanian devils are good swimmers does not mean that they are generally 'good' at everything, including hunting – indeed, they may be poor hunters and mainly obtain their meat from scavenging. The adjective *good* expresses a property that always has to be relativized to something specific; nothing is 'good' in an absolute sense (Siegel 1976a, b). Crucially, both intersective and subsective adjectives allow a simple inference pattern of the type 'If *x* is an Adj N then *x* is an N'. For instance, once we know that Tasmanian devils are *carnivorous animals*, we can also conclude that they are *animals*, and given that Tasmanian devils are *good swimmers*, we can infer that they are *swimmers* (taken here in the general sense of organisms that can swim). Note that intersective adjectives, which allow a conjoined entailment (*x* is an N and *x* is Adj), are in fact a special case (i.e., a subtype) of subsective adjectives, which merely require that the entailment '*x* is an N' holds. In other words, all intersective adjectives are subsective, but not vice versa. A third class of adjectives, the so-called 'modal' adjectives, also known as 'plain non-subsective' adjectives, such as *alleged* or *possible*, block this basic kind of inferential pattern that we find with (intersective and non-intersective) subsective ones. On knowing that John is an *alleged murderer*, for instance, we can be certain neither of John being a murderer nor of John not being a murderer; we can only infer that he *may* be a murderer. Finally, there is a fourth class of adjectives, the 'privative' adjectives, such as *apparent*, *artificial* and *imaginary* – and, indeed, *fake*, which in the linguistic literature is considered as the prototypical example of this class (Kamp 1975; Kamp and Partee 1995; Partee 1995; Coulson and Fauconnier 1999; Nayak et al. 2014; Del Pinal 2015; Morzycki 2015; Chatzikyriakidis and Zhaohui 2017). Privative adjectives are defined as blocking, just like the modal adjectives, the inference pattern 'If *x* is an Adj N then *x* is an N', but, what is more, as also permitting the inference that '*x* is *not* an N'.[6]

Figure 1 provides a diagrammatic representation of this standard classification of adjectives, using Venn diagrams. Examples of each type of adjective are given, as well as inferential patterns that distinguish between the four types. Sets represent the respective extensions of nouns and adjective-noun combinations and dots in these sets schematically represent entities making up the extension. Note that in the diagram of plain non-subsective adjectives, depicted bottom left, elements in the adjective-noun set are connected by means of

[6] As a somewhat (but not entirely) off-topic remark, that Tasmanian devils are not real devils does not mean that *Tasmanian* is also a privative adjective. It only goes to show how important multi-word units are in natural language understanding.

dashed lines with elements, labelled with a question mark, in the noun set. This represents the idea that there may or may not exist elements in the latter set that are coreferential with elements in the former. For privative adjectives, observe that there is no overlap of the set of entities in the extension of adjective-noun combinations and the set of entities in the extension of the unmodified noun.

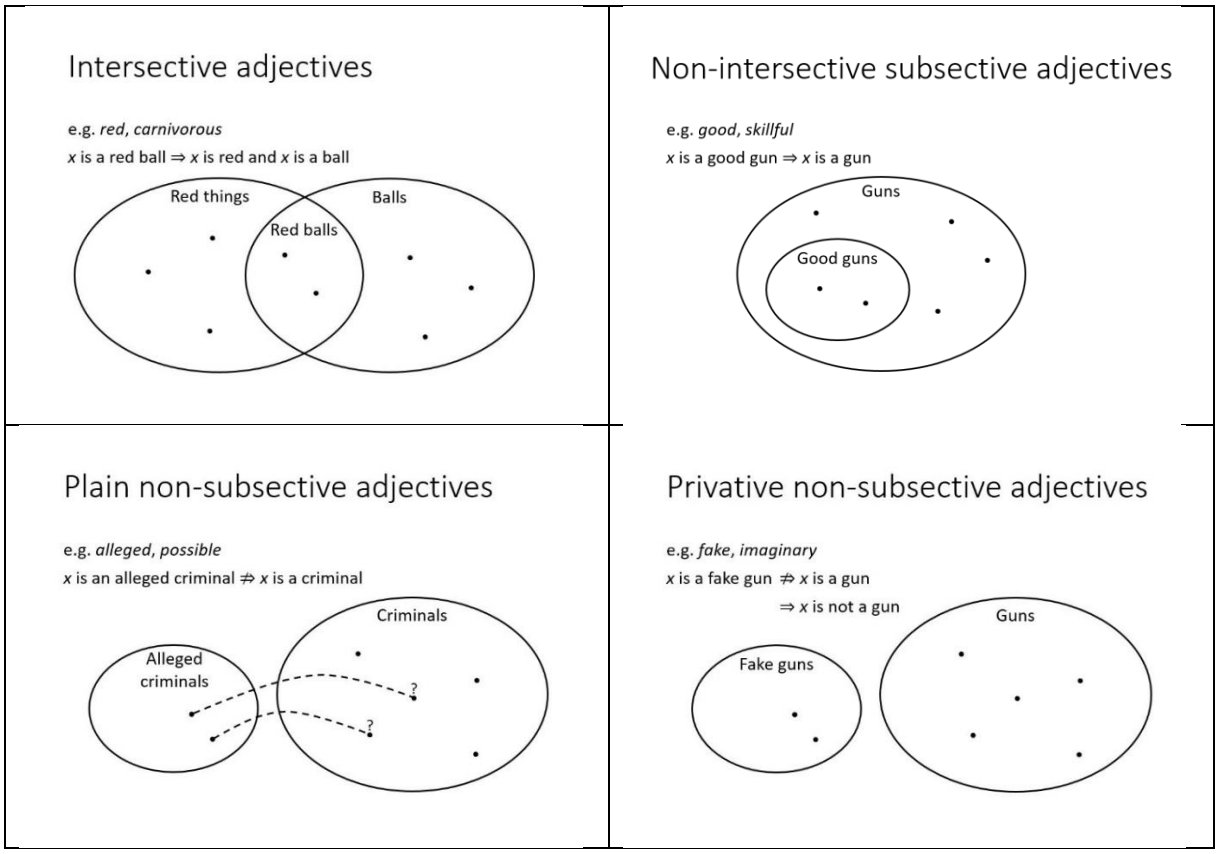



**Figure 1**. Common classification of adjectives based on entailment relations, following Kamp and Partee (1995).

As mentioned above, *fake* has come to be seen as the example *par excellence* of a privative adjective. As such, it excludes the referent of a noun phrase containing the modifier *fake* (e.g., the thing referred to as a '*that fake gun*') from the unmodified noun's extension (i.e., the set of things that fall under the label '*gun*'). In what can be credited as being the most extensive treatment of *fake*, Del Pinal (2015) considers this exclusion from the noun/nominal's extension a central, ever-present ingredient of the meaning of *fake*. He summarizes his position about the inherently privative nature of *fake* as follows: "So any adequate treatment of privative Adjs such as *fake* has to explain why, for instance, from knowing that x is a fake gun we can infer i. that **x is not a gun**, and ii. that x was created with the intention that it only seem like a gun, but not that it can be used to shoot" (Del Pinal 2015: 8, our emphasis).[7] The point that it only seem like a gun is of vital importance, as we will

[7] Del Pinal's (2015) analysis is couched in formal semantics. His claim that there is an invariant privative aspect to *fake* follows from these statements and formulas, taken from various places in his paper but, as far as we can see, not taken out of context: "Assume *e* stands for an arbitrary common N (e.g., an artifact or natural kind term)" (p. 3); "*e* has a complex semantic structure. Part of this structure determines *e*'s extension (call this its *E-structure*), and part of it does not determine *e*'s extension (call this its *C-structure*)" (p. 4); "we will include a function $Q_E$ which takes the meaning of terms and returns their E-structure denotations" (p. 13); "*Qualia functions:* partial functions from the meaning of terms into their respective C-structure denotations, namely,

show later on, but here we want to draw the attention to the first point. Apparently, Del Pinal analyzes *fake* as being always privative in the technical sense going back to Kamp (1975): it allows us to judge as true any sentence of the form "no fake noun/nominal is (a) noun/nominal". This is clearly wrong, however, as we will now show.

## 2.2 Caveats and problems

'Privative adjective' is a term that we should use with extreme caution. We can more safely speak of 'privative uses' of an adjective. In fact, Kamp (1975: 125), who introduced the term 'privative' for adjectives such as *fake*, *false* and *fictional*, already remarked, "I doubt that there is any adjective which is privative (in the precise sense here defined) in all of its possible uses." That so-called privative adjectives are not *always* used privatively is mentioned in passing more recently by Boleda et al. (2012: 1228) when they contrast *false floor*, "which is not a real floor" and *false rumor*, "which is a real rumor that is false". As pointed out also by Nayak et al. (2014), so-called 'privative adjectives' may occur in idiomatic and/or fixed combinations in which the adjective does not have a straightforwardly privative effect. For instance, a *false alarm*, similarly to *false rumor*, really *is* an alarm, in the dictionary sense of a warning or a feeling of fear or anxiety, albeit one that turns out to be needless, as the dangerous situation fails to materialize.

Moreover, even when a 'privative adjective' *is* really used privatively, its meaning cannot simply be reduced to its privative effect. That is why privative adjectives are sometimes far from substitutable. For example, a *fake hero* is someone pretending to be heroic (or someone unjustifiedly made to be heroic by others), hence a person who is *not* a hero, while a *mythical hero* is someone who only belongs to a mythical world of legends and folk tales, or at best an idealized past, but whose deeds in that fictional or romanticized world really *are* heroic. What this shows is that privative adjectives cannot be seen as simple operators that remove the referent of a noun phrase from the extension of the head (i.e., the set of real-world entities the noun denotes). We will come back to this point in section 2.3.

In any case, note that there are diagnostics for privativity of *fake*-noun combinations. One is simply looking up the definition of the noun. The definition of *gun* is "a weapon incorporating a metal tube from which bullets, shells, or other missiles are propelled by explosive force, typically making a characteristic loud, sharp noise" (via Oxford Dictionaries[8]). As fake guns practically always lack this internal mechanism and therefore aren't meant to shoot, they do not classify as guns. This contrasts with *story*, which is defined as "an account of imaginary or real people and events told for entertainment" or "an account of past events in someone's life or in the development of something". It has as one of the subsenses "a report of an item of news in a newspaper, magazine, or broadcast" or "a particular person's representation of the facts of a matter" (via Oxford Dictionaries[9]). A fake story, whether one told in a conversation or one published or broadcast in the media, always remains a story in one of these (sub)senses, regardless of the fact that its contents are partially or wholly made up.

---

constitutive, formal, telic, and agentive. The qualia functions are $Q_C$, $Q_F$, $Q_T$, $Q_A$" (p. 13); "$D_C$ is a variable that ranges over ordered sets of the E-structure and C-structure of common Ns." (p. 14); "Consider then the full lexical entry for *fake*: […]: E-structure: $\lambda D_C.[\ \boldsymbol{\lambda x.\neg Q_E(D_C)(x)} \wedge \neg Q_A(D_C)(x) \wedge \exists e_2[\text{MAKING}(e_2) \wedge \text{goal}(e_2, Q_F(D_C)(x))]]$ […]" (p. 20; our emphasis).

[8] https://en.oxforddictionaries.com/definition/us/gun, last accessed 10 August 2018.

[9] https://en.oxforddictionaries.com/definition/us/story, last accessed 10 August 2018.

Another, related, diagnostic involves the use of a definitional pattern ('A fake N/N' is an N/N' that…').[10] This reveals that while sometimes *fake-noun*/nominal combinations clearly exhibit a privative use, as appears from the oddity of defining them by starting with the bare noun/nominal (cf. (1), where '!' indicates the sentence is semantically anomalous), there are many other combinations for which this definitional format does not sound semantically odd at all (cf. (2)), and where *fake* cannot be considered privative:

(1) a. !A fake gun is a gun that…
b. !Fake leather is leather that…
c. !A fake orgasm is an orgasm that…
d. !A fake painting by Van Gogh is a painting by Van Gogh that…

(2) a. A fake story is a story that…
b. A fake document is a document that…
c. Fake products are products that…
d. Fake versions of the new £5 note are versions of the new £5 note that…

The distinction may be a subtle one, but observe that we can, additionally, test whether using scare quotes (and/or ironic intonation) is appropriate. If the speaker/writer knows that a particular object is a fake gun and if this speaker/writer knows that the hearer/reader also knows this, the speaker can refer to this object as *a "gun"*, with intonational, gestured and/or typographical quotation marks. By contrast, such shared knowledge about the factual falsity of a story does not license the use of scare quotes and/or ironic intonation: talking about *!a "story"* (note again the quotation marks) to refer to what is known to be a fake story makes no sense. This test is useful in cases where there is no dictionary definition available, for instance when we are dealing with complex nominals, as in (1d) and (2d). A gang of counterfeiters may among themselves refer to a fake painting they want to pass off as a painting actually made by Van Gogh as *our "painting by Van Gogh"*, whilst counterfeiters who print fake fivers will not talk to each other about *!our "version of the new £5 note"*. A version of something, even if it is a fake one, is still a version.

For some combinations of *fake* and a noun/nominal, it is far from obvious to decide whether using the noun/nominal as head of the definition is semantically odd or not. An example is *passport*, which is basically a document one needs to have checked by the authorities each time one enters or leaves one's country. In that respect, even a fake passport is still a passport (just like a fake story is still a story; see also Pavlick and Callison-Burch (2016: 115) for a similar treatment of *fake ID*). But if you find it also crucial for a passport to be considered as such that it is delivered by the authorities in the first place, contains only fully accurate information, and is not illicitly modified once it was delivered, then you might not consider a fake passport truly a passport. Thus, the following part of the definition may or may not sound semantically odd, depending on what you consider a crucial part of the meaning of *passport*:

(3) (!)A fake passport is a passport that…

[10] N' (read: 'N bar') is a structure larger than a single noun but smaller than a noun phrase, notably not including a determiner (cf. Jackendoff 1977). It corresponds to what some grammars, e.g. Huddleston and Pullum *et al.* (2002), call a 'nominal', a term we also use in our text. Note, indeed, that *fake* may modify a noun (e.g. [$_{N}$ *gun*]), resulting in a nominal ([$_{N'}$ *fake* [$_{N}$ *gun*]]), or a structure which is already a nominal (e.g. [$_{N'}$ *nude photograph of the actress*]) resulting in another, higher-order, nominal ([$_{N'}$ *fake* [$_{N'}$ *nude photograph of the actress*]]).

## 2.3 *Fake* and its interaction with the noun's qualia structure

The above observations can be summed up as follows: with some nouns/nominals, *fake* has a clearly privative effect, with some others, it clearly does not, and with others still, whether or not the effect is privative remains unclear. A reasonable conclusion from this situation is that privativity is not the main semantic import of *fake*. That is, the modification of a noun/nominal by *fake* is *not* meant, in the first place, to predicate of a noun/nominal's referent that it is not an instance of what this noun/nominal generally denotes. More likely, what it does, as its general semantic contribution, is indicate that this referent is deceptive. A fake gun is an object that deceptively looks like a real shooting weapon, a fake story is a sequence of narrated events that have a deceptive aura of truth, a fake passport is an object that deceptively looks as though it was a completely legitimately obtained travel document, and so on. The presence or absence of a privative interpretation is then only a by-product of this semantic constant of *fake* and its interaction with other semantic contributors.

Though Del Pinal (2015), as we have seen, mistakenly assumes that *fake* always has a privative effect, his analysis is in fact much more sophisticated, as is indicated by the second part of the quotation from his work we gave in Section 2.1. Del Pinal sees *fake* as operating also on the noun/nominal's 'qualia structure', a notion introduced by Pustejovsky (1995) but ultimately going back to ideas of Aristotle. A noun/nominal's qualia structure specifies, simply put, what an entity expressed by a noun/nominal looks like, what it consists of, what it is used for and how it came into being. Del Pinal suggests that the modifying adjective *fake* has as its meaning not just that the referent of the noun phrase whose noun/nominal it modifies does not fall under that noun/nominal's extension – this is the too general privative part of its meaning – but also something in addition to this, relating to the noun/nominal's qualia structure (cf. also Lakoff and Johnson 1980: 120-121). *Fake* conveys the information that the object concerned did not come into being the way the corresponding non-fake entity does, that it was made to have the perceptual features we know the corresponding non-fake entity to have and that it lacks the goal specified for the corresponding entity. This part of Del Pinal's analysis is interesting, but again not appropriate for all imaginable *fake*-noun/nominal combinations.

Del Pinal suggests that for nouns that are not artefact terms, we may modify the noun's meaning by relying on general encyclopaedic knowledge to enrich its TELIC qualia role, which specifies what its goal is or what it is used for.[11] No enrichment is needed for the non-artefact noun *blood*, as this denotes something with an identifiable function ("delivers necessary substances such as nutrients and oxygen to the cells and transports metabolic waste products away from those same cells", from Wikipedia[12]). This function is not 'just' technical, encyclopaedic knowledge but is, perhaps in a more simplified form, a firm part of the meaning of *blood* that can be mentally accessed by speakers. Note also that it can be found in the definition of *blood* given by some dictionaries (e.g., "The red liquid that circulates in the arteries and veins of humans and other vertebrate animals, carrying oxygen to and carbon dioxide from the tissues of the body", in Oxford Dictionaries[13]). This aspect of the

[11] Del Pinal also suggests that, complementarily, we may contextually modulate the meaning of *fake*, but he then appears to argue against this (Del Pinal 2015: 19).
[12] https://en.wikipedia.org/wiki/Blood, last accessed 10 August 2018.
[13] https://en.oxforddictionaries.com/definition/us/blood, last accessed 10 August 2018.

meaning of the noun, belonging to the TELIC qualia role, is not preserved when *fake* is added as a modifier. However, many non-artefact nouns denote objects that have no obvious function. What, for instance, is the function of *beards*? There appears to be none. In some regions in the world, men have a reduced capacity to grow a full beard, and women generally do not have visible facial hair, so a full beard does not seem to be crucial for a man's reproductive success, let alone for our species' survival. A beard may be grown, by someone who can, to comply with societal or religious norms, to signal masculinity or wisdom, to avoid being recognizable to those who don't know the person has a beard, and so on. If we use this encyclopaedic information to fill the underspecified TELIC qualia role of beard, then we have to conclude that *fake* does *not* appear to affect this. On the contrary, a fake beard may be worn precisely to obtain one or more of these effects (fit in with the hipsters, look old and wise, seem devote, look like Santa, dress in disguise to remain undetected, etc.). We could conjecture that when a noun refers to a concept which has a clear goal or function, this semantic aspect will be affected by *fake*; when there is no clear goal or function, then obviously *fake* cannot affect what is not present. Perhaps a similar point can be made with respect to other qualia roles, for instance the so-called AGENTIVE qualia role (dealing with how an entity came into being): if this aspect is not featured prominently or at all in the conceptual content of a noun, then *fake* cannot affect this qualia role in the same way as it does with nouns that do make central reference to the entity's origins.

# 3. Method

## 3.1 Computation of semantic distances between adjective-noun combinations and nouns: 'privative' vs. 'non-privative' adjectives

Our first hypothesis, which is that privative adjectives 'shift' the usage contexts of a noun more markedly than non-privative adjectives do, can be tested with the help of distributional-semantic methods. Distributional semantics is an approach to measuring similarities and differences in meanings between words based on the hypothesis that not just the proverbial birds but also *words* of a feather flock together. That is, linguistic items that are found in similar contextual environments have similar meanings (Harris 1954, Firth 1957, McDonald and Ramscar 2001, Sahlgren 2008). Natural language processing (NLP) has had successes in capturing the contexts of words or combinations of words as so-called 'word embeddings', which are dense numerical vectors (i.e. sequences of, typically, 300 or 500 numbers) representing linguistic items' co-occurrences with words, collected in a continuous vector space (Turney and Pantel 2010). This vector space has undergone, by mathematical techniques, a reduction of dimensionality – from very many dimensions, where each and every word in the corpus is a potential context word for a linguistic item and hence a distinct dimension, to relatively few, namely 'only' a few hundreds.

For the present study, we used word embeddings based on a corpus containing the entire set of English-language texts on Wikipedia as it was in 2013. This corpus consists of approximately 75 million lines, where each line consists of a sentence, totaling almost 1.6 billion words (tokens). While this corpus is more restricted in terms of register and genre variety than most other large corpora currently available, we chose to use it because we could readily use the word embeddings based on it, as these are provided in the Mangoes environment, which was developed by the Magnet team at Inria.[14] Moreover, the corpus is

[14] https://gitlab.inria.fr/magnet/mangoes/wikis/resources, last accessed 1 November 2017.

formatted in the right way – it has been tokenized, for instance – to allow quite straightforward extraction of embeddings of word combinations (phrase embeddings).

Following part of the experimental protocol of Boleda et al. (2012), we calculated the semantic similarity between the embeddings of nouns and adjective + noun bigrams for a set of 'privative' adjectives and a set of matched 'non-privative' adjectives, on the English Wikipedia 2013 corpus. We chose four privative and four non-privative adjectives with similar type frequencies of adjective + noun combinations in the corpus (see Table 1). The non-privative adjectives were chosen from the list of non-intentional (i.e., intersective and non-intersective subsective) adjectives provided in Boleda et al. (2012).

**Table 1.** Selection of four privative and four non-privative adjectives, with numbers of distinct adjective + noun bigrams (types) with occurrences > 10 in the English Wikipedia 2013 corpus

| **Privatives** | Number of bigrams | **Non-privatives** | Number of bigrams |
|---|---|---|---|
| *false* | 318 | *stable* | 343 |
| *artificial* | 317 | *safe* | 270 |
| *fake* | 194 | *loose* | 196 |
| *fictitious* | 61 | *nasty* | 52 |
| totals | 890 | | 861 |

Although technical details and their implications are not our central concern here, let us nonetheless specify the decisions made in our procedure. We used cosine distance between vectors as a measure of semantic similarity between noun and bigram embeddings. As it was done in Boleda et al. (2012) we computed the embeddings of the part-of-speech-aware lemmas (distinguishing for example between *fake* the adjective and *fake* the verb). We computed the co-occurrence of these target lemmas with context POS-aware lemmas in a symmetric half-window of four words around each target lemma. The obtained co-occurrence matrix (target lemmas x context lemmas) was normalized using positive point-wise mutual information (PPMI). Finally, the embedding vectors were obtained by projecting the PPMI data into the 500 first singular vectors computed by Singular Value Decomposition (SVD). Unlike what is done in Boleda et al. (2012), the embedding vectors for the adjective + noun bigram were produced in a second step by computing the co-occurrence of the observed bigram (lemmatized and POS-informed) with the context lemma in a window of half-size four, by normalizing the counts using the PPMI weights obtained on the overall corpus and by projecting the resulting sparse vectors into the 500 first singular vectors obtained previously.

Boleda et al. (2013) control for the number of occurrences of the adjective by choosing several adjectives with a wide range of numbers of occurrences (some rare, others very frequent) in order to have a 'diverse' picture of the cosine similarity. Here, we controlled for the number of occurrences of the nouns, also taking into account Adj + N bigram frequency. Instead of looking directly at the cosine similarity, we computed an interpolated line representing the average cosine similarity value for each possible value of the log((N - bigram) occurrences) and looking at how selected bigram and N pairs diverge from the average behaviour.

## 3.2 Exploring context-sensitivity

### 3.2.1 Manual annotation of corpus-extracted examples

Our second hypothesis concerns the adjective *fake*, as this is the adjective that has gained the status of being *the* most clearly privative modifier. More specifically, we assume that its interpretation is to some extent context-sensitive, depending on the kind of noun it occurs with. Our intuition is that the combination of *fake* with a noun tends to be interpreted in a

particular way when the noun refers to something containing verbal or visual information (e.g., *story*, *news*, *video*) and a different way when the noun refers to a natural or manmade entity with a well-described purpose other than informing people (e.g., *blood*, *gun*, *medicine*). When the noun refers to an event (e.g., *burglary*, *death*, *wedding*), a *fake*-noun combination is again likely to be interpreted in yet a slightly different way, and so on. To test this assumption, we need to examine different *fake*-noun combinations and annotate them, using a sufficiently fine-grained set of possible semantic values.

From the English Wikipedia 2013 corpus, we extracted all sentences with the word *fake*, resulting in a collection of just over 14,000 tokens of *fake* with its sentence-length context. Using Stanford POS Tagger in Natural Language Toolkits (NLTK), we provided all words in these sentences with part-of-speech labels. We then extracted from our set of sentences all tokens of *fake* as an adjective followed by zero, one or more adjectives, followed by one or more singular or plural nouns. Our script allowed us to exclude occurrences of *fake* as a noun or a verb, as well as 'incomplete' sequences such as *fake shark* or *fake rewards* occurring as part of *fake shark attack* or *fake rewards card*, respectively, and to *include* occurrences of *fake* modifying such noun-noun compounds (but see further). The script collected a total number of 8,432 types, of which 4,163 are hapaxes (unique occurrences).

We selected the 100 most frequent types thus obtained, excluding however those with multiple nouns (e.g. *fake field goal attempt*), because it is not always easy to find a definition for such complex nouns in dictionaries (e.g. for *antivirus software* in *fake antivirus software*) and because we used definitions to decide whether the combination has a privative effect; moreover, the internal bracketing is not always obvious with these multi-noun sequences (e.g. [*fake* [*news stories*]] or [[*fake news*] *stories*]?). Also excluded were those types with a plural noun for which the singular noun occurred higher up in the list of 100 most frequent *fake*-noun combinations (although we could alternatively have decided to join singular and plural occurrences as manifestations of a single type), as well as some other kinds of doubles (e.g. *fake mustache* was retained, but not *fake moustache*). Removed from the list were also *fake one* and *fake ones*, as the head noun in these cases does not have a lexical meaning, but is instantiated by an anaphoric pronoun, and irrelevant cases such as *fake book*, which has a very specific meaning among jazz musicians.

The 100 most frequent *fake*-noun combinations (types) represent 1920 occurrences (tokens). The most frequent type in our set is *fake blood* (with 123 tokens in the Wikipedia corpus) and not the nowadays very commonly used type *fake news* (which with its 35 tokens comes in at place 19) – note, though, that we used a version of Wikipedia frozen in time, predating the role of fake news in the 2016 US presidential elections. Note also that the nature of Wikipedia (containing mostly objective, factual subject matter) is such that one or more otherwise frequent *fake*-noun combinations are not among the 100 most frequent ones in this corpus. For example, *fake smile* is in the top 10 most frequent *fake*-noun combinations in the Corpus of Contemporary American English, but not in the top 100 in the English Wikipedia corpus.

Taking Del Pinal's (2015) analysis as a starting point, we checked whether the interpretation he proposes applies to the 100 extracted *fake*-noun combinations. We developed a coding scheme in a conservative way, by adapting his proposal, informally rendered in (4) below, when it turned out to be inapplicable to a given combination, making as minimal adjustments as possible. While avoiding creating a new interpretational type for each different noun, we added a new type to our coding scheme whenever the interpretation of a *fake*-noun combination could not be adequately rendered by means of any of the types already

distinguished. In (4)-(10), we list, and define, all the interpretational types that we needed in order to annotate the 100 most frequent *fake*-noun combinations. (In these definitions, 'x' stands for the noun at hand, which may be countable or uncountable and may start with a vowel or silent *h* or not – which is why we use '(a(n)) x'. The *fake*-noun combination provided as an example for each interpretational type is the one with the highest number of tokens in the Wikipedia corpus for that type.)

(4) Type 1, based on Del Pinal (2015), e.g. *fake blood*
(A) fake x is not (a(n)) x, did not come into being the way (a(n)) x normally comes into being, and was not made with the goal of fulfilling the core function of (a(n)) x but was made to be perceived like (a(n)) x, so that it has one or more of the effects of (a(n) x that came into being the normal way.

(5) Type 2, e.g. *fake punt* (i.e. a form of trick play in American football where the ball, despite being played from the player formation typical of punts, ends up not being punted (i.e. kicked) but passed)
(A) fake x is not (a(n)) x, does not exist (or occur) as such, but elements one associates with (a(n)) x make it look like there is (a) really existing (or occurring) x, so that these elements have one or some of the effects of (a(n)) x that really exists (or occurs).

(6) Type 3, e.g. *fake beard*
(A) fake x is not (a(n)) x, did not come into being the way (a(n)) x normally comes into being but was made to be perceived like (a(n)) x, so that it has one or more of the effects of (a(n)) x that came into being the normal way.

(7) Type 4, e.g. *fake passport*
(A) fake x is (a(n)) x that did not come into being the way (a(n)) x normally comes into being but was made to be perceived like (a(n)) x that came into being the normal way so that it functions like, or at least has one or some of the effects of, (a(n)) x that came into being the normal way.

(8) Type 5, e.g. *fake article*
(A) fake x is (a(n)) x that did not come into being the way (a(n)) x normally comes into being but was made to be perceived like (a(n)) x that came into being the normal way so that, while what it conveys (in words or images) is not (completely) true (i.e., corresponding to reality), this is perceived as completely true, and so that it functions like, or at least has one or some of the effects of, (a(n)) x that came into being the normal way.

(9) Type 6, e.g. *fake version*
(A) fake x is (a(n)) x that did not come into being like the original x but was made to be perceived like the original x so that it functions like, or at least has one or some of the effects of, the original x.

(10) Type 7, e.g. *fake person*[15]

[15] This interpretation (type 7) is included here for the sake of completeness. In fact, though, the one sentence in our sample that appears to display this interpretation for a combination contains a typo (*parson* misspelt as *person*). An example with the plural noun *people* (not included in the sample) is given in (i):

(i) He realizes that the woods are now his home and he will never be happy in modern society, with its noise, pollution, and fake people.

This use of *fake people* has an extra connotation: people leading a life that fails to be authentic, in the sense of purposive, mindful, true to one's inner self. This shows that individual pairs may have specific interpretations that speakers probably store in their mental lexicon. The most typical interpretation of *fake person* is of type 2.

(A) fake x is (a(n)) x that is made to look like something or someone it is (or they are) not or that pretends to be something or someone it is (or they are) not, so that it has (or they have) one or more of the effects of that something or someone.

In some cases, the noun that is part of a *fake*-noun combination has multiple unrelated meanings. An example of this is *notes*, which in combination with *fake* is used either with the meaning of 'banknotes' or with that of 'short letters'. (The meaning of 'musical sound' is not easily eligible to modification by *fake*, and in any case, this meaning is not attested in any of the 13 extracted sentences with *fake notes*.) Similarly for *arm*: *fake arm* sometimes denotes 'artificial upper limb' and sometimes 'fake weapon'. In those cases, we considered the meaning that in the set of extracted sentences turned out to be the commonest. (For *notes*, it is that of 'bank notes', and for *arm* it is 'limb'). The only exception to this procedure is *fake person* (cf. footnote 11). For *title*, which when modified by *fake* can mean 'name of a book/film/…' or 'name indicating rank or position', the two meanings were kind of equally frequently attested, and we considered them both. (This did not, however, result in a different assignment of a type of *fake*-noun interpretation for these two uses of *title*; cf. Section 4.)

Figure 2 represents the different types schematically, showing the interpretational features making up each type of *fake*-noun combination. The features on the left correspond roughly to information relating to the AGENTIVE qualia role of the combination. The features on the right correspond roughly to its TELIC qualia role. Interpretative features enclosed by dotted lines are ones that do not hold for all relevant combinations and/or all contexts of use and/or all speakers. This figure also shows which types are clearly privative (the three types at the top, instantiated by *fake punt*, *fake blood* and *fake beard*) and which types are clearly non-privative (the three types at the bottom, instantiated by *fake article*, *fake version* and *fake person*): the former do and the latter do not include the feature 'Not (a(n)) x'. As discussed above (Sections 1 and 2.2), *fake passport* is not clearly privative nor clearly non-privative; this depends on which features of *passport* one considers essential.

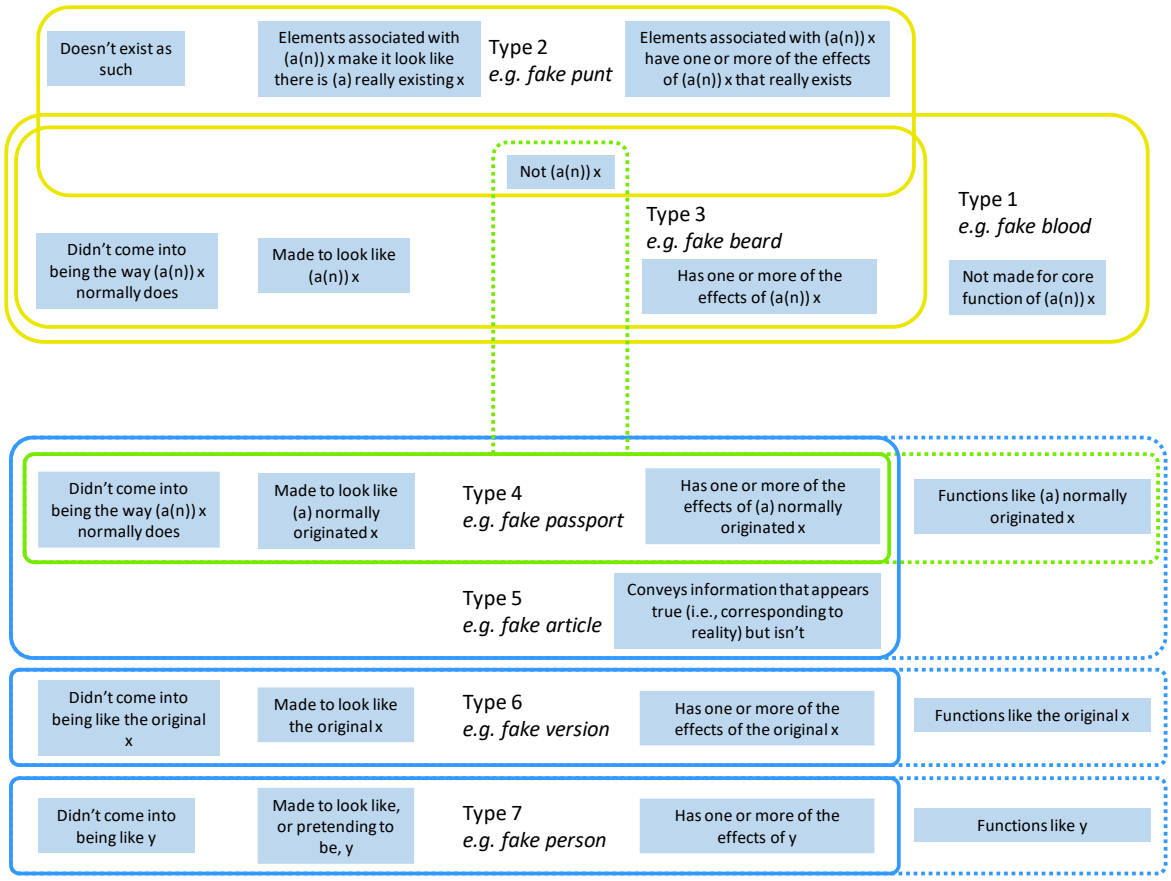



**Figure 2.** Interpretational features of seven different types of *fake*-noun combinations, used for a manual annotation of 100 frequent *fake*-noun combinations extracted from the English Wikipedia 2013 corpus.

### 3.2.2 Computation of semantic distances between nouns

The second part of hypothesis 2 is that the different interpretations assigned manually to *fake*-noun combinations are not random but correlate with semantic properties of the nouns involved. In other words, nouns with similar semantic properties yield similarly interpreted *fake*-noun combinations. To test this, we computed the cosine distance between the embeddings for each pair of nouns in the set of 100 most frequently used *fake*-noun combinations (cf. Section 3.2.1). In order to obtain a network of similarities between these nouns we then computed an affinity matrix by applying a Gaussian kernel (i.e., an exponential transformation) to these distances, setting a gamma value of 6.25. The effect of the Gaussian kernel is to emphasize the discrepancies, thus providing a high affinity value to words that are close and an affinity value closer to zero the further apart the words are. It was decided that two nouns were linked if their affinity was above a threshold (= 0.01). The gamma value (controlling how much to emphasize the discrepancies) and the threshold were chosen jointly in order to obtain a network not fully connected (in case of too low a threshold) but without disconnecting too many words (when the threshold chosen is too high).

Our next step in the procedure was to obtain a visually interpretable representation of the distributional similarity of each noun in the set of hundred nouns to each other noun in that set. To this end, we projected the network in 2D, using the algorithm ForceAtlas2. As explained by Jacomy et al. (2014: 3), ForceAtlas2 "simulates a physical system in order to

spatialize a network. Nodes repulse each other like magnets, while edges attract their nodes, like springs."[16]

Finally, we measured the degree of correspondence between the interpretational types manually assigned to *fake*-noun combinations and the visually recognizable clusters of nouns in the distributional network. To do so, we applied the *performance P* (Fortunato 2010). This is a quality function which allowed us to check for each pair of nodes belonging to the same class (i.e., nouns for which the same type of *fake*-noun interpretation was assigned) whether they are connected by an edge, and for each pair of nodes not belonging to the same class whether they are not connected by an edge.

### 3.3 Computation of semantic distances between adjective-noun combinations and nouns: privative vs. non-privative uses of *fake*

We also wanted to test the hypothesis – our third one – that the contexts of a *fake*-noun combination differ more greatly from the contexts of the noun as such when *fake* is used privatively than with it is used non-privatively. This hypothesis was tested by adopting the same method as described in section 3.1 (which we used for 'privative' and 'non-privative' adjectives), applied in this case to the hundred most frequent *fake*-noun combinations in the English Wikipedia 2013 corpus, for which the annotations (see section 3.2.1) allowed us to make a distinction between privative and non-privative uses, as well as ambivalent uses.

## 4. Results

### 4.1 Distance of adjective-noun combinations to nouns: privative vs. non-privative adjectives

As regards the semantic similarity between adjective-noun combinations and nouns, we found no difference between four privative and four non-privative adjectives. This is shown in Figure 3, where each coloured dot represents an adjective-noun combination, either with one of the 'privative' adjectives or one of the 'non-privative' ones. The position of an adjective-noun combination along the X-axis is relative to the frequency of its noun (NN counts) minus its own frequency as a bigram (adjxNN counts), while its position along the Y-axis represents the cosine similarity between its (observed) embedding as a bigram and the embedding of the noun (both based on the English Wikipedia 2013 corpus).

[16] In graph theory, *nodes* (also known as *vertices*) are the fundamental objects making up a network and their connections are referred to as *edges*, represented as lines between nodes. In our study, the nodes in the network represent nouns, and the edges between them are the links provided by their distributional similarity.

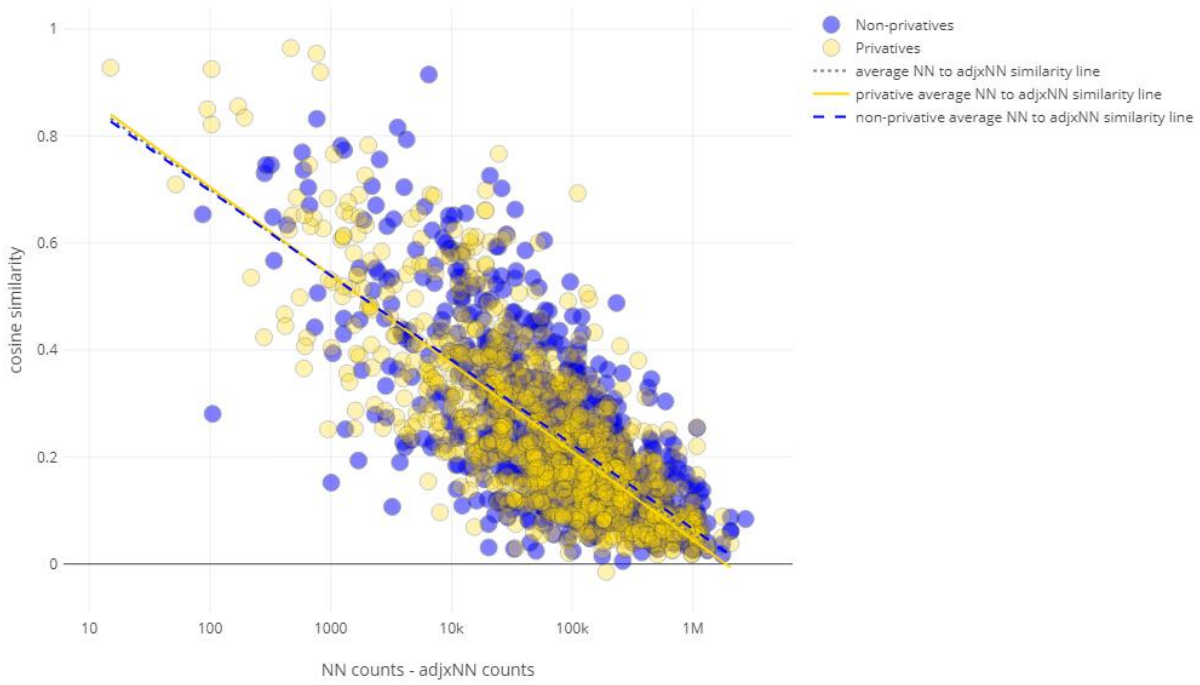



**Figure 3.** Distributional similarity between nouns and adjective + noun bigrams for a set of 'ordinary' adjectives (labelled here as 'non-privatives') and a set of adjectives with privative uses ('privatives').

The reason for presenting the data along a conceptually complex X-axis is that the distributional similarity between an adjective + noun bigram and the related noun depends on two factors that should be viewed together. First, a high-frequency noun tends to be general or polysemous (cf. Zipf 1949) – it appears in diverse contexts, all of which jointly make up its unique embedding – and an adjective will be likely to select only one of the many possible uses or meanings of that noun. Thus, as the corpus frequency of a noun increases, so the distributional similarity between the embedding of the noun and the embedding of the related adjective + noun bigram typically decreases. Secondly, the more often a noun co-occurs with a particular adjective, the more similar the contexts of the noun as such and the adjective + noun bigram will tend to be. That is why the frequency of the noun has to be 'corrected' with the frequency of the adjective + noun bigram.

Figure 3 shows most conspicuously that in a logarithmic scale there is a linear association between the cosine similarity of interest and the corrected number of occurrences of the noun indiscriminately across type of adjectives. So, no matter whether the adjective is privative or not, adjective + noun bigrams differ in their distribution more greatly from the noun when this noun has a rich 'life of its own' than when that noun is found almost exclusively with the adjective. For instance, the noun *part*, which is very general/polysemous, occurs in many more contexts than does, say, *loose part*, whose contexts make up only a tiny fraction of the noun's usage; hence, the contextual vectors of this noun and this adjective + noun bigram differ greatly, and their similarity, therefore, is close to 0. By contrast, *insemination* occurs almost exclusively in combination with *artificial*, so their contextual vectors are almost identical (i.e. the contexts in which one finds the noun are almost the same

as the contexts in which we find the adjective + noun bigram), and the similarity between the vector of *insemination* and that of *artificial insemination* is close to 1.

Crucially, though, Figure 3 reveals that, on average, the 'privative' adjectives and the 'non-privative' ones shift the distribution of nouns to a practically identical extent. Indeed, observe that both types' average N to Adj+N similarity lines lie very close to each other. This means that, counter to expectation (cf. hypothesis 1), when a noun is modified by a 'privative' adjective such as *false*, *artificial*, *fake* or *fictitious*, this combination typically does *not* find itself in contexts that are more different from the general contexts of that noun than when a noun is modified by a 'non-privative' adjective such as *stable*, *safe*, *loose* or *nasty*.

## 4.2 Context-sensitivity of *fake* senses

As can be seen in Figure 4, only for 10 *fake*-noun interpretations in our set does the type 1 interpretation, given above in (4), appear to hold. In other words, for all the other *fake*-noun combinations, the most typical interpretation is of a different type. Privative interpretations were only assigned to almost exactly half the combinations (to 51 out of 101; remember that *fake person* was considered twice, with different interpretations). Matching the colours used in Figure 2 above, bars indicating numbers of privative interpretations are rendered in yellow and bars indicating numbers of non-privative interpretations in blue, while green is used for the bar indicating the number of times we assigned the interpretational type that is ambivalent as to privativity.

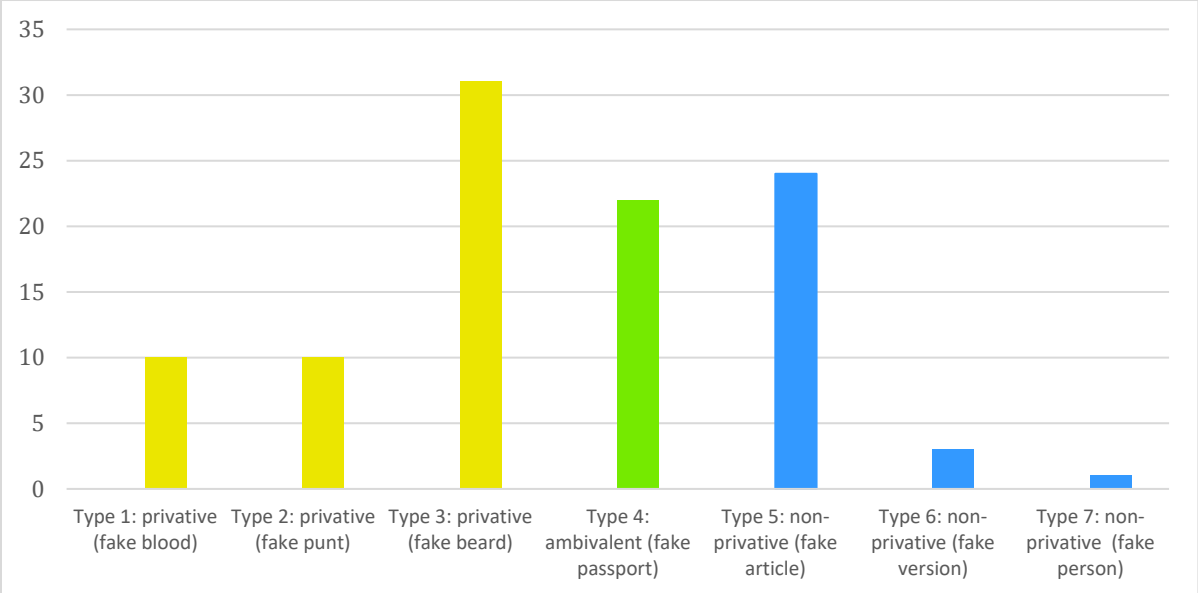



**Figure 4**. Number of *fake*-noun combinations (types) in our set of frequent combinations to which a particular interpretational type has been assigned.

Table 2 provides further details by mentioning for each interpretational type the combinations we assigned it to. Numbers following combinations indicate the number of tokens of that combination in the sample extracted from the English Wikipedia 2013 corpus.

**Table 2.** Interpretational types of *fake*-noun combinations and the combinations, extracted from the English Wikipedia 2013 corpus, to which these are assigned.

| Interpretational type | Combinations to which the type on the left was assigned |
|---|---|
| Type 1: not x, didn't come into being normally, not made for core function, made to look like x, has one or more of the effects of x | fake arm 8, fake blood 113, fake bomb 18, fake drugs 10, fake food 13, fake gun 15, fake hand 9, fake medicine 9, fake pistol 7, fake teeth 10 |
| Type 2: not x, doesn't exist as such, elements associated with x made it look like there is existing x, these elements have one or more of the effects of existing x | fake attack 8, fake band 9, fake burglary 9, fake charity 7, fake companies 7, fake company 23, fake coup 11, fake couple 12, fake death 21, fake encounter 36, fake engagement 9, fake eviction 10, fake evidence 16, fake fight 7, fake funeral 7, fake kidnapping 12, fake movie 9, fake newspaper 8, fake person 8, fake pregnancy 22, fake psychic 8, fake punt 70, fake reality 9, fake references 19, fake relationship 12, fake robbery 9, fake source 8, fake suicide 12, fake town 13, fake war 9, fake wedding 14 |
| Type 3: not x, didn't come into being normally, made to look like x, has one or more of the effects of x | fake beard 45, fake eyelashes 12, fake family 14, fake flowers 7, fake fur 20, fake money 32, fake mustache 25, fake tan 24, fake tree 7, fake vomit 10 |
| Type 4: didn't come into being normally, made to look like normally originated x, has one or more of the effects of normally originated x | fake account 18, fake advertisement 9, fake bills 9, fake commercials 17, fake credentials 14, fake currency 16, fake degree 8, fake goods 9, fake ID 75, fake identification 14, fake identity 42, fake marriage 42, fake name 81, fake notes 13, fake paintings 8, fake passport 83, fake persona 11, fake profiles 14, fake signature 10, fake tickets 16, fake title 8, fake trailer 25 |
| Type 5: didn't come into being normally, made to look like normally originated x, appears true but isn't, has one or more of the effects of normally originated x | fake article 72, fake blog 12, fake case 15, fake claims 8, fake documentary 16, fake documentation 7, fake email 16, fake footage 7, fake image 9, fake information 40, fake interview 24, fake letter 41, fake message 13, fake news 35, fake page 13, fake papers 11, fake photos 8, fake reason 32, fake report 14, fake story 44, fake telegram 8, fake video 10, fake warnings 9, fake website 19 |
| Type 6: didn't come into being like the original x, made to look like the original x, has one or more of the effects of the original x | fake copy 14, fake edition 11, fake version 26 |
| Type 7: made to look like, or pretending to be, something it isn't or somebody they aren't, has one or more of the effects of that thing or person | fake person 8 |

As can be seen at a glance in Figure 5, the interpretational types which we identified for *fake-noun* combinations by means of manual annotation (cf. Section 3.2.1) cluster relatively well together, reflecting the fact that they tend to belong to groups of nouns that have similar distributions (and, hence, similar meanings). In this figure, colours indicate the interpretational types of noun when modified by *fake*, based on our manual annotation: yellow hues represent interpretational types 1 to 3, green represents the ambivalent case (type 4) and blue hues represent interpretational types 5 to 7.

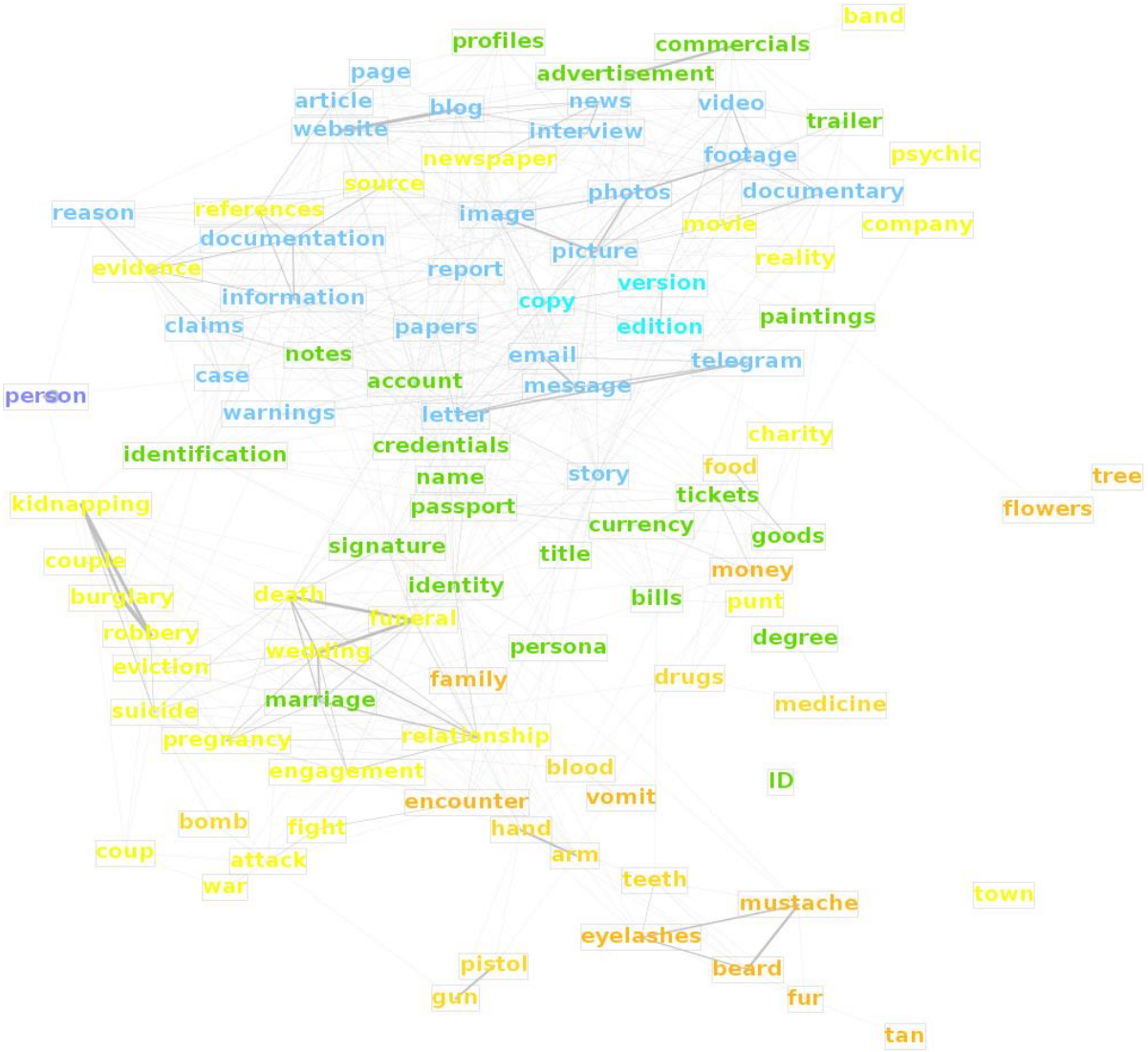



**Figure 5.** 2D network representation of nouns' distributional similarity, obtained with ForceAtlas2, with colours indicating interpretational types of *fake*-noun combinations.

We obtain a well-above-chance *performance P* of 0.78, that is, pairs of nouns to which we assigned the same *fake*-noun interpretation tend to be connected by an edge, and pairs of nouns to which we assigned different *fake*-noun interpretations tend not to be connected by one. This finding supports hypothesis 2 about the context-sensitivity of *fake*: the way we

interpret a *fake*-noun combination depends to some non-negligible extent on semantic properties of the noun concerned.

## 4.3 Distance of *fake*-noun combinations to nouns: privative vs. non-privative uses

Remember from Section 4.1 that we found no difference in 'distributional shift' between adjective+noun bigrams and nouns when we compared 'privative' and 'non-privative' adjectives. What if we now focus on the hundred most frequent *fake*-noun combinations extracted from the English Wikipedia 2013 corpus and take into account the observation, confirmed in Section 4.2, that not all uses of this so-called 'privative' adjective are, in fact, privative? In that case, we do find that the privative uses of *fake* engender a slightly larger shift than the non-privative uses do, with the ambivalent uses taking a middle position. Figure 6 plots the distributional similarity of *fake*-noun combinations to the respective nouns forming part of them. As in Figure 3, the position of combinations along the X-axis varies according to the corpus frequency of their noun corrected for the frequency of the *fake*-noun combination itself, as we want to represent how much of a 'life of its own' the noun has, i.e., how often it occurs without *fake*. Again, the more often the noun occurs without *fake*, the larger the distributional distance between that noun and its combination with *fake*, as could be reasonably expected. More to the point, Figure 6 shows how far nouns are 'shifted' from their distributional position depending on the main interpretational types of *fake*-noun combinations: with *fake* being privative (types 1 to 3 grouped together), ambivalent (type 4), or non-privative (types 5 and 6). (Type 7, which also involves a non-privative use of *fake*, has not been included, as it represents only one member with a low number of tokens.)

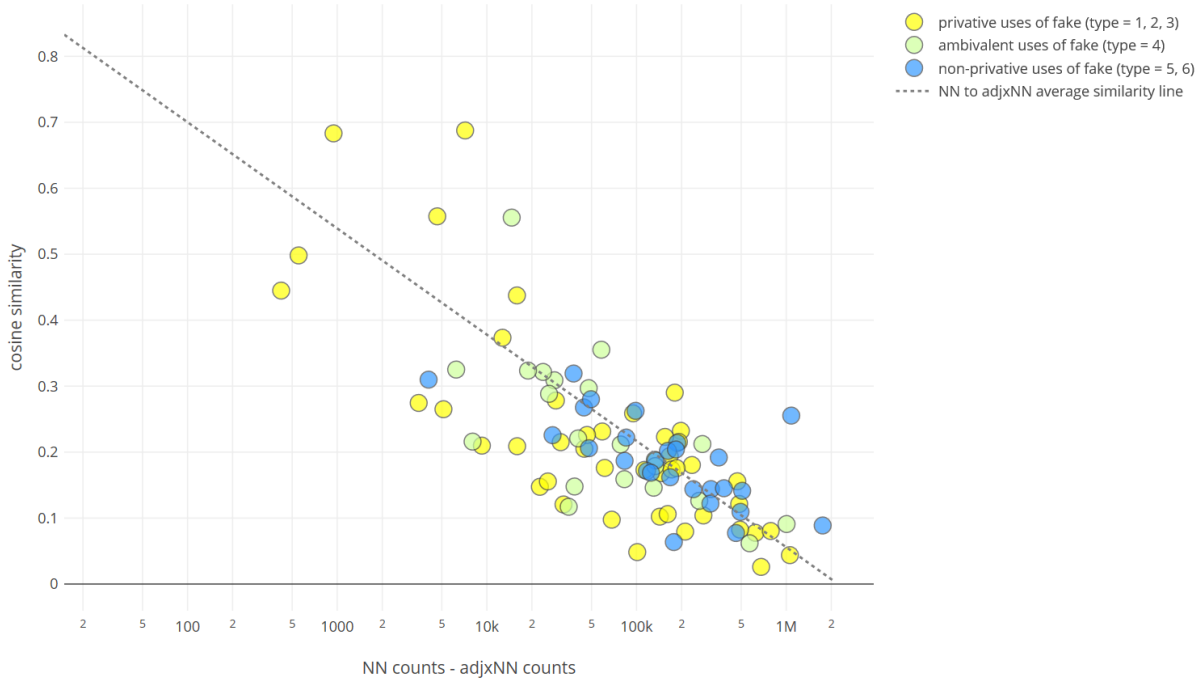



**Figure 6.** Distributional similarity between nouns and *fake* + noun bigrams for three main uses of *fake*.

*Fake*-noun combinations with a non-privative use of *fake* have a tendency to stay closer to the average N to Adj+N similarity line obtained earlier for a larger set of adjectives (cf. Section 4.1) than do combinations with an ambivalent and privative use of *fake*. Although this may be hard to make out in Figure 6, the data suggest that privative and ambivalent uses of *fake* tend to shift the *fake*-noun combination further away from the distributional position of the noun than do non-privative uses. This observed difference, though small and only suggestive, is shown in Figure 7. Each boxplot in this figure depicts for each use of *fake* (privative, ambivalent, non-privative) the range of distances between the average N to Adj+N similarity line and the N to *fake*+N similarity across relevant pairs of vectors. Positive values along the Y-axis indicate that the N to *fake*+N similarity is higher than the N to Adj+N average. Negative values indicate a lower-than-average similarity, in other words, a larger-than-average distributional shift caused by *fake*. The rectangles represent the second and third quartiles, with the line in the middle indicating the median value. Horizontal lines at the ends of the whiskers above and below each rectangle show the upper and lower values that still lie within 1.5 times the length of the rectangle, and values outside that range are shown as points. The median distance to the average N to Adj+N similarity line is at -0.024 for the privative uses of *fake*, at -0.022 for the ambivalent uses and at -0.006 for the non-private uses.

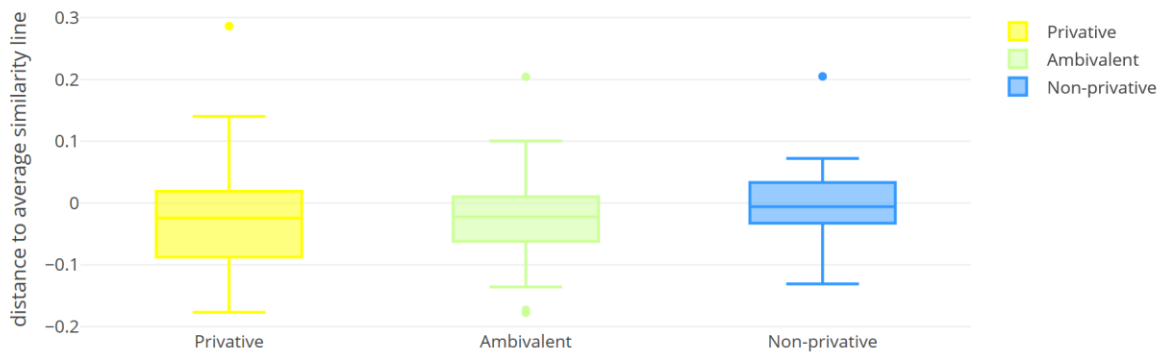



**Figure 7.** Boxplots comparing three main uses of *fake* with respect to how far their 'distributional shift' is removed from the average N to Adj+N similarity line.

This result is based on too small a data set for it to reach statistical significance. Nevertheless, it suggests that our third hypothesis might be correct: privative uses of *fake* (as in *fake blood*) cause there to be a larger shift between the contexts of the N and those of the Adj+N combination than do non-privative uses of *fake* (as in *fake article*).

# 5. Discussion

## 5.1 No noticeable difference between 'privative' and 'non-privative' adjectives in distributional shift

When an adjective modifies a noun, it stands to reason that the resulting Adj+N combination will mean something else than just the noun taken on its own. Consequently, we can expect the Adj+N combination to occur in different contexts from the contexts in which we find the

N. This difference in co-occurrence patterns between Adj+N and N can be conceived of as a quantifiable 'distributional shift', namely, as a measurable distance between their contextual vectors. So-called 'privative' adjectives, which are claimed to allow the statement '(An) Adj x is not (an) x', could be expected to trigger a bigger distributional shift than ordinary adjectives, as the effect of privative modification on the denotation of an Adj+N combination (compared to the denotation of the N) seems intuitively more drastic than that of ordinary modification. Yet, our comparison of four adjectives that are standardly considered to be privative ones (*artificial*, *fake*, *false*, *fictitious*) and four adjectives that are not (*loose*, *nasty*, *safe*, *stable*) yielded a negative result: no noticeable difference in distributional shift was found between these two sets.

This negative result confirms that of Boleda et al. (2013), who also set out to find a distributional difference between the vectors of nouns and observed vectors of adjective-noun combinations, but found no difference whatsoever in distance between ordinary (subsective) adjectives and so-called 'intensional' (non-subsective ones). However, as we explained in Section 2, the class of intensional adjectives not only includes privative adjectives such as *fake* but also 'plain' non-subsective adjectives such as *alleged*, *former*, *hypothetical* and *likely*, which do not necessarily have as an effect that the denotation of the Adj+N combination is not within the denotation of the N. So, even if the privative adjectives which we chose for our study could have been expected to produce a less 'diluted' picture than the larger set of intensional adjectives used by Boleda et al. (2013), these privative adjectives did not prove to cause a greater distributional shift than ordinary adjectives.

In fact, our result fails to confirm Boleda et al.'s (2012) earlier findings, which revealed a *greater* similarity of Adj+N bigrams to the related Ns for intensional (i.e. non-subsective) adjectives than for non-intensional colour terms (intersective ones such as *white* in *white dress* and otherwise subsective ones, such as *blue* in *blue state*). They comment on this by suggesting that "intensional adjectives do not restrict the descriptive content of the noun they modify, in contrast to both the intersective and subsective color ANs [i.e., adjective-noun combinations]. Restriction of the nominal description may lead to significantly restricted distributions (e.g. the phrase *red button* may appear in distinctively different contexts than does *button*; similarly for *green politics* and *politics*), while we do not expect the contexts in which *former bassist* and *bassist* appear to diverge in a qualitatively different way because the basic nominal descriptions are identical, though further research will be necessary to confirm these explanations" (Boleda et al. 2012: 1230). There may be something about colour term adjectives that they tend to cause a greater distributional shift than other adjectives do, including non-subsective ones.

## 5.2 *Fake*: semantic variation and unity

Our study has shown that not all uses of *fake* are privative, in the sense of allowing the statement '(A) fake x is (an) x', and that how we interpret *fake* in combination with a noun depends to a large extent on the semantics of the noun. Our identification of seven distinct interpretational types of *fake*-noun combinations therefore testifies to the (largely) context-dependent nature of interpreting *fake*. For now, the seven interpretational types, which we then reduced to three (privative, non-privative and ambivalent), may feel somewhat *ad hoc*. It would be worthwhile, therefore, to use distributional-semantic methods to find out how many specialized senses *fake* actually has. That there *are* distinct context-dependent senses is hardly

beyond doubt. Note, for example, that with event nouns such as *fight*, *robbery*, *wedding* or *suicide*, we can replace *fake* by *mock* or *staged*, while these alternatives would sound odd with nouns such as *tan*, *eyelashes* or *smile*. These different (near-)synonyms for *fake* depending on the noun that follows strongly suggests that *fake* has more than one sense.

This view should not be seen as an endorsement of rampant polysemy of *fake* or as a denial that there is an invariant aspect of meaning to *fake*. On the contrary, we firmly believe that *fake* has something semantically in common across all *fake-noun* combinations, namely the idea of deception: something/someone is deliberately presented as something/someone that looks like something/someone else in order to mislead. It is precisely because of this deception that some effects associated with the concept denoted by the unmodified noun are intended to be achieved, and in some cases, that the deceptive entity even functions like the 'real' thing. For instance, a *fake passport* is produced for the exact same purpose as a genuine passport, namely allowing the holder to get in or out of a country. Even *fake blood*, which obviously does not have the purpose of carrying oxygen to parts of the body, has one or more of the effects usually associated with real blood. Collins Dictionary defines blood as "the red liquid that flows inside your body, which you can see if you cut yourself"[17]. Consequently, on seeing *fake blood* and taking it for real blood, one can be led to assume that someone is hurt and one can experience a sense of horror at this. It merits further investigation to ascertain whether or not the notion of deception targets the AGENTIVE qualia role in those cases in which the conceptual content of a noun profiles, in Langacker's (1987) sense, information on the coming into being, provenance or fabrication of an entity (e.g. *fake Swiss army knife*) while it targets the telic qualia role when the functionality of an entity is a more central ingredient of the noun's conceptual structure (e.g. *fake knife*).

## 5.3 The privativity effect may still be a real thing

Remember that we found no noticeable difference between so-called privative and ordinary adjectives in the distributional shift they bring about between N and Adj+N. We also observed that *fake* is not an across-the-board privative adjective and can only be said to have privative and non-privative *uses* (as well as uses that are ambivalent). By restricting ourselves to the privative uses of *fake* and by controlling for number of N occurrences, we see what Boleda et al. (2013) had failed to detect: considering the median value, the cosine similarity between N and fake+N is smaller than the average cosine similarity between N and Adj+N. This suggests that the distributional effect of privative modification may be a real thing after all.

We still need to explain properly why the difference in distributional shift between privative and non-privative modification, if it is indeed real, is so small. Perhaps it is in part due to the fact that speakers don't always have to precede a noun with a privative adjective in order to make it clear that the referent of the noun phrase is not to be included within the noun's extension. For instance, in the sentence *That gun looks so fake*, the noun remains unmodified and it is the verb phrase *following* the noun phrase that makes it clear that the

[17] https://www.collinsdictionary.com/dictionary/english/blood, accessed 11 August 2018.

referent of that noun phrase is not within the noun's extension. Such a sentence and many others like it may attenuate the distributional difference between *fake gun* and *gun*.

## 6. Conclusion and prospects

In this study, we have combined methods in distributional semantics (involving the calculation of distances between vector representations of words and phrases) and a more traditional corpus linguistic approach (involving manual annotation of a set of extracted sentences) to answer this question: Is the semantic difference between an N and an Adj-N combination larger if the adjective is privative than when the adjective is an ordinary one? We did not find such a difference when we compared four so-called privative adjectives and four ordinary ones. However, after demonstrating that the privative adjective *par excellence*, *fake*, is used privatively in only half of the most frequently used *fake*-noun combinations, we restricted our attention to the privative uses and found that these did produce a slightly larger distributional 'shift' between N and Adj + N bigrams. That is, *fake*-noun combinations with a privative use of *fake* tend to appear in contexts that diverge (slightly) more largely from the contexts of the noun than do combinations with ordinary adjectives or combinations with a non-privative use of *fake*.

That there may not be any 'privative' adjectives that are privative in all their possible uses was acknowledged when the term was first introduced by Kamp (1975). However, to the best of our knowledge, the existence of non-privative uses for privative adjectives has since then only been mentioned sporadically in the literature and has even then only been presented as exceptional. In our study, we have put this usage variability fact centre-stage, attempting to make more precise the many interpretations of *fake-noun* combinations and showing that the interpretation of *fake* is context-dependent. In particular, semantic properties of the noun (or nominal) play a crucial role in the interpretation of the combination of *fake* with a noun (or nominal).

A further hypothesis to be tested in future research is that when *fake* modifies a very specific noun, it is privative, but when it modifies a very general noun, it is not. A preliminary indication that this may be the case comes from compound nouns such as *Rolex watch* or *Van Gogh painting*: with these, *fake* is privative with respect to the entire compound (e.g., a fake Rolex watch is not a Rolex watch) but not with respect to the head (e.g., it *is* a watch). With very general-concept nouns such as *thing*, *stuff*, *object* and *goods*, the interpretation of *fake* as a premodifier is automatically non-privative – *fake stuff* is obviously still stuff but it is not the specific kind of stuff that is under discussion in the context. A complication in this regard is that discourse often allows language users to use lexical items that are in themselves already quite specific but that in other contexts may be understood as shorthand descriptions for something yet more specific. Compare, for instance, two possible uses of *fake gun*. In a report of a bank robbery, a *fake gun* will be interpreted as an object that only looks like a gun but lacks the function of shooting bullets, while in the context of an antique and vintage firearms fair, *fake guns* may more likely be understood as referring to (fully functioning) cheap

replicas of, for instance, sought-after Smith & Wesson or Colt revolvers. In the former case, *fake* is privative with respect to the dictionary meaning of *gun*, in the second case, it is not.

It is clear, therefore, that for a correct understanding of *fake*-noun combinations, we cannot rely solely on the dictionary meaning of the noun but need to include encyclopaedic aspects of meaning, which involve speakers' cultural knowledge. As just stated, many nouns can in a given discourse context be used as hypernyms for a much more specific concept. To give another example, when there is discussion of *fake handbags*, what is meant is not things that look like handbags but that in fact lack the function of handbags; rather, *fake handbags* is more likely to be used in the sense of 'real handbags that are passed off as being manufactured by a well-known designer brand'. This is based on our world knowledge about handbags, namely that they are not just functional objects but, perhaps in the first place, accessories whose make is important. Likewise, the combination *fake paintings* usually does not refer to objects that look like paintings but that are really mass-produced posters, but to real paintings sold by fraudsters who trick people into believing the paintings were made by famous artists.

The challenge to predicting when *fake* is privative or not is further complicated by the possibility of adjectives to be used non-restrictively, as in *my sick mother* or *the talented Mr. Ripley*. In the newspaper headline *Ashley Madison admits tricking men with fake fembots*, the adjective *fake* is used in exactly such a way: the interpretation is not that men are tricked by 'things that look or sound like fembots but that are not fembots' – something the men would not complain about – but, of course, that they had unawarely been chatting up 'fembots (feminine-gendered chatbots), which are fake things'. We have some preliminary indications that this non-restrictive use may be more common than one could have expected. Similar example sentences we encountered in a follow-up study are *Phantasus was known for creating fake dreams and dreams full of illusion* (not: 'things that look like dreams but are not') and *These parkas featured fake synthetic fur on the hoods after an outcry from the fur lobby* (not: 'something that looked like synthetic fur but was not synthetic after all').

In sum, it is remarkable how effortlessly we, as humans, interpret *fake* correctly when it modifies a noun. The precise interpretation of *fake* is highly sensitive to the kind of noun it combines with and speakers also have to bring to bear their knowledge of the world and of the wider discourse context. Cognitive-linguistic, usage-based approaches are well equipped to face these facts about *fake*.

## Acknowledgments

This research was supported by a funded part-time research stay ("Délégation") at Inria (Institut National de Recherche en Informatique et Automatique) Lille – Nord Europe (2017-2018) awarded to Bert Cappelle for the project FACT (*Fake*-type Adjectives: a Computational Tool), with the aim of collaborating with Inria researchers to find ways of predicting the semantic impact of so-called privative adjectives. Our thanks also go to Chris Biemann, Hans Boas, Aurélie Herbelot and Chris Pinon, with whom we have had the opportunity to talk about the intricacies of *fake* modification at various points during our research. We are extremely grateful to the editors of this volume for their valuable comments on a former version of our paper. The usual disclaimers apply.

## References


Boleda, Gemma, Marco Baroni, Nghia The Pham, Louise McNally. 2013. Intensionality was only alleged: On adjective-noun composition in distributional semantics. In Proceedings of IWCS 2013, Potsdam.

Boleda, Gemma, Eva Maria Vecchi, Miquel Cornudella and Louise McNally. 2012. First order vs. higher order modification in distributional semantics. In *Proceedings of EMNLP*, Jeju Island, Korea, 1223–1233.

Chatzikyriakidis, Stergios and Luo Zhaohui. 2017. Adjectival and adverbial modification: The view from modern type theories. *Journal of Logic, Language and Information* 26(1), 45–88.

Coulson, Seana and Gilles Fauconnier. 1999. Fake Guns and Stone Lions: Conceptual Blending and Privative Adjectives. In Barbara Fox, Dan Jurafsky & Laura Michaelis (Eds.) *Cognition and Function in Language*, 143-158. Palo Alto, CA: CSLI.

Del Pinal, Guillermo. 2015. Dual Content Semantics, privative adjectives, and dynamic compositionality. *Semantics & Pragmatics* 8, article 7, 1-53.

"fake, n.2". *OED Online*. June 2018. Oxford University Press. http://www.oed.com.kuleuven.ezproxy.kuleuven.be/view/Entry/67776?rskey=wtLa3o&result=2&isAdvanced=false (accessed July 24, 2018).

Firth, John Rupert. 1957. *A synopsis of linguistic theory 1930-1955* (Studies in Linguistic Analysis). Oxford: Philological Society, 1–32. Reprinted in Frank Robert Palmer (ed.) 1968. Selected Papers of J.R. Firth 1952-1959. London: Longman.

Flach, Susanne. 2014. Kandidaten für den Anglizismus 2013: Fake-. *Sprachlog*. http://www.sprachlog.de/2014/01/06/kandidaten-fuer-den-anglizismus-2013-fake/, last accessed 28 June 2018.

Fortunato, Santo. 2010. Community detection in graphs. *Physics Reports* 486: 75–174.

Harris, Zellig. 1954. Distributional structure. *Word* 10(23), 146–162.

Huddleston, Rodney, and Geoffrey K. Pullum *et al*. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Jackendoff, Ray. 1977. *X-bar-Syntax: A Study of Phrase Structure* (Linguistic Inquiry Monograph 2). Cambridge, MA: MIT Press.

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann and Mathieu Bastian. 2014. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLoS ONE* 9(6): e98679. doi:10.1371/journal.pone.0098679

Kamp, Hans. 1975. Two theories about adjectives. In: Edward L. Keenan (eds.), *Formal Semantics of Natural Language*, 123-155. Cambridge: Cambridge University Press.

Kamp, Hans, and Barbara Partee. 1995. Prototype theory and compositionality. *Cognition* 57: 129-191.

Lakoff, George & Mark Johnson. 1980. *Metaphors we live by*. Chicago, IL: The University of Chicago Press.

Langacker, Ronald W. 1987. *Foundations of Cognitive grammar. Volume I, Theoretical prerequisites*. Stanford: Stanford University Press.

McDonald, Scott and Michael Ramscar. 2001. Testing the distributional hypothesis: The influence of context on judgements of semantic similarity. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, 611–616.

Morzycki, Marcin. *Modification* (Key Topics in Semantics and Pragmatics). Cambridge: Cambridge University Press.

Nayak, Neha, Mark Kowarsky, Gabor Angeli, Christopher D. Manning. 2014. A Dictionary of Nonsubsective Adjectives. Stanford, CA: Stanford University Reports. http://hci.stanford.edu/cstr/reports/2014-04.pdf

Partee, Barbara. 1995. Lexical Semantics and Compositionality. In: Lila Gleitman and Mark Liberman (eds.), *An Invitation to Cognitive Science (Second Edition)*. Volume 1: Language, 311– 360. Cambridge, MA: MIT Press.

Pavlick, Ellie and Chris Callison-Burch. 2016. So-called non-subsective adjectives. Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics (*Sem 2016), Berlin, Germany, August 11-12, 2016, 114–119.
Pustejovsky, James. 1995. *The generative lexicon*. Cambridge, MA: MIT Press.
Sahlgren, Magnus. 2008. The Distributional Hypothesis. *Rivista di Linguistica* 20(1): 33–53.
Siegel, Muffy. 1976a. Capturing the Russian adjective. In Barbara H. Partee (ed.), *Montague Grammar*, 293–309. New York: Academic Press.
Siegel, Muffy E. A. 1976b. Capturing the Adjective. Ph.D. dissertation. University of Massachusetts.
Stefanowitsch, Anatol. 2017. Laudatio zum Anglizismus des Jahres 2016: Fake News. *Sprachlog*. http://www.sprachlog.de/2017/01/31/laudatio-zum-anglizismus-des-jahres-2016-fake-news/, last accessed 28 June 2018.
Taylor, John R. 2003. *Linguistic Categorization.* Third edition. Oxford: Oxford University Press.
Turney, P. and P. Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37, 141–188.
Zipf, George K. 1949. *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.