# Comparing Distributional and Frame Semantic Properties of Words

**Thomas Kleinbauer** **Thomas Alexander Trost**
Spoken Language Systems
Saarland University
Saarland Informatics Campus, Bldg. C7 1
66123 Saarbrücken, Germany
{thomas.kleinbauer,thomas.trost}@lsv.uni-saarland.de

## Abstract

Word embeddings have gained a lot of traction in recent years since their multidimensional representation of words exhibits useful semantic properties while the process of obtaining them requires little to no manual interaction. At the same time, linguistic theories of lexical semantics have been under active research for many decades and have rich and well-established foundations. Have such theories now become obsolete with the advent of sophisticated statistical methods based on machine learning?

This paper investigates the relationship between a number of widely used variants of word embeddings and one specific theory of lexical semantics, viz. frame semantics. Based on the lexical units available in the FRAMENET corpus, we study some commonalities and differences in the way these are represented in each formalism. We comment on possible future directions that allow both worlds to learn from each other and thus may help improve practical aspects of word embeddings as well as frame semantics.

## 1 Introduction

*Frame semantics* is a semantic theory based on the idea that in communication, words perceived by a listener[1] evoke in their minds a mental image of a scene or situation together with the relevant constituents of and participants in that scene, and the roles they play (Fillmore, 1982; Petruck, 1996). For instance, in the sentence "Bob paid 15 dollars for a simple sandwich.", the word *paid* evokes a COMMERCIAL_TRANSACTION frame which draws on the listener's everyday experience of the exchange of goods for money at a certain price.

With the FRAMENET corpus[2], a manually annotated resource is available for English[3] that defines over 1200 frames, lists over 13,000 words and the frames they evoke, and contains over 200,000 example annotations based on material mainly taken from the British National Corpus[4].

However, in recent years, a different kind of semantic word representation has become widely used. *Word embeddings* map words to multi-dimensional vectors with the idea that words that are closely related in meaning should be mapped to vectors that are near each other in the underlying vector space. Such embeddings are typically derived using machine learning techniques on large textual corpora, a process that eliminates most of the manual efforts required for a hand-crafted resource like FRAMENET.

Considering these two opposing approaches, the question arises in how far linguistic insights expressed in Frame Semantics can be recovered in the geometric properties of word embeddings. We are interested in questions such as:

- Is the lexical material encapsulated in a semantic frame as a whole reflected in the vector space locations of the words belonging to that frame?
- How are words that evoke multiple frames treated by embedding methods?
- Can statistical methods underlying word em-

[1] Throughout the paper, we will refer to *speakers* and *listeners* as producer and consumer of linguistic material, even though our arguments are not restricted to spoken language but apply to other media as well.

[2] https://framenet.icsi.berkeley.edu

[3] FrameNets in other languages are also available but only English is considered in this paper.

[4] http://www.natcorp.ox.ac.uk/



Proceedings of the 14th Conference on Natural Language Processing (KONVENS 2018)
Vienna, Austria – September 19-21, 2018

bedding approaches aid in the development of a resource like FRAMENET, thus far created mostly by hand?

We present below four experiments that serve as a starting point to answer these and similar questions.

## 2 Background and Related Work

### 2.1 Word Embeddings

Word embeddings map each word in a given dictionary to a vector in a high-dimensional vector space where the specific vector chosen for a word depends on its distributional properties in comparison with other words. The often-cited rationale for this, as formulated by Harris (1968), states that similar distributional patterns tend to indicate similar word meanings.

Since the seminal work of Bengio and colleagues (Bengio et al., 2003), dense neural-network-based vector representations of words have become ubiquitous with the number of different approaches too large to address all of them in this work. The basic tenet is that of distributional semantics: a word's representation is sought to be highly predictable from the representation of the surrounding context words found in a corpus. The specific methods that we consider are given in Sec. 3.

### 2.2 Frame Semantics

Frame semantics is a theory of lexical semantics proposed by Fillmore (Fillmore, 1976; Fillmore, 1982) as an advancement of his earlier theory of case grammar. Its central posit is that the understanding of the meaning of words is intimately related to real-life experiences of the listener. In understanding, words are said to *evoke* in the mind of the listener a mental structure called a frame, a cognitive schematization of a situation or scene. The participants that contribute to such a scene are called frame elements. For instance, the previously mentioned COMMERCIAL_TRANSACTION frame alludes to a situation consisting of a BUYER, SELLER, GOODS, and MONEY. When a frame is evoked by a word, all of its frame elements become simultaneously available. Because of this, it is not possible to understand the meaning of the term *goods* without also knowing about the meaning of such terms as *buy*, *price*, etc. (Petruck, 1996). The pairing of a word and its meaning (expressed as a frame) is called *lexical unit* (LU). This study focuses on words and the frames they evoke and leaves other aspects of frame semantics for future work.

The FRAMENET project (Baker et al., 1998) provides an online resource consisting of a lexical database defining frames and listing frame-evoking lexical units with the aim of documenting the range of valences of each lexical unit through example annotations (Ruppenhofer et al., 2016). FRAMENET has been used as a key resource in a number of natural language processing tasks, including e.g. semantic parsing (Shi and Mihalcea, 2005), opinion and topic extraction (Kim and Hovy, 2006), and question answering (Shen and Lapata, 2007).

### 2.3 Relating Embeddings and Frames

One aspect of Mikolov et al.'s *word2vec* system that drew interest to the use of embeddings for lexical semantics is the fact that the learned representations of words and phrases allowed certain semantic properties to be expressed in terms of simple algebraic operations, e.g., vector addition (Mikolov et al., 2013). This motivated related research on consolidating such algebraic relationships, e.g. by Pennington et al. (2014), as well as on incorporating external resources in addition to textual corpora and on representing more abstract linguistic notions as embeddings. For instance, Iacobacci et al. (2015) learn continuous representations of word senses, based on the BABELNET[5] resource, a large network of synsets and semantic relations (Navigli and Ponzetto, 2012). Similarly, Bollegala et al. (2016) combine a text corpus and WORDNET (Fellbaum, 1998) to yield word representations that outperform previous methods in a semantic similarity prediction task and a word analogy detection task. They claim that other resources, such as e.g. FRAMENET, could be used with their method instead of WORDNET but do not report any results on this. Flekova and Gurevych (2016) recount that WORDNET senses have been criticized for being too fine-grained and thus experiment with coarser *supersenses* of words. They train a joint word+supersense embedding model and apply it to a number of text classification tasks.

At this point, it is worth mentioning that some ef-

[5] http://babelnet.org

forts have been made at integrating WORDNET and FRAMENET, e.g. by Ferrández et al. (2010), to leverage complementary information encoded in both. Similarly, techniques to leverage WORDNET and FRAMENET in order to improve the quality of word embeddings have been proposed, e.g. where relational information is derived from semantic lexicons to retrofit embedding vectors in a post-processing step (Faruqui et al., 2015) .

Automatic methods for predicting the frame evoked by a word and its frame elements have become a research topic in its own right, see, e.g., (Gildea and Jurafsky, 2002; Erk and Padó, 2006). The fact that the performance of the best semantic role labeler software at the time, SEMAFOR, could be further improved through the integration of word embeddings (Hermann et al., 2014), is an additional motivation to study the relation between frames and embeddings more closely. There is, however, little prior research in that area. One exception is the work by Botschen et al. (2017) who use frame embeddings trained in a fashion comparable to Hermann et al.'s approach to study whether the FRAMENET relations naturally arise from text in similar algebraic notions as in Mikolov et al.'s work. However, they find no such evidence.

Perhaps most similar to our own efforts is the work by Pennacchiotti et al. (2008) and by Roth (2008) who define embeddings for frames themselves to help discover new frame evoking words. However, our own interest is less task-driven: we focus on studying the general question in how far independently derived word embeddings exhibit features predicted by a non-distributional semantic theory.

## 3 Experiments

Prima facie, a conceptual difference between the two approaches is that the meaning representation of a word in the context of word embeddings is a singular point (i.e. a vector in a high-dimensional space), while in frame semantics, it is a rich structure: the frame. By "rich" we refer to the fact that a frame is a "system of concepts related in such a way that to understand any one concept it is necessary to understand the entire system; introducing any one concept results in all of them becoming available" (Petruck, 1996). At the same time, the vectors for the words under any embedding scheme are not assigned randomly, the mapping is designed such that the distance between two vectors gives an indication about the semantic similarity of the underlying words. But despite this geometric property, does the arrangement of word vectors give rise to the same, or at least similar, structures as can be found in frame semantics?

Table 1: The datasets used in this paper. See the text for more information on the column headers.

| dataset | *D* | *C* | *T* | corpus |
|---|---|---|---|---|
| `FT` | 300 | 96.7% | | Wikipedia |
| `G50` | 50 | 96.3% | 6G | Wikipedia & Gigaword 5 |
| `G100` | 100 | 96.3% | | |
| `G200` | 200 | 96.3% | | |
| `G300` | 300 | 96.3% | | |
| `G42` | 300 | 98.9% | 42G | Common Crawl |
| `G840` | 300 | 99.3% | 840G | Common Crawl |
| `GT25` | 25 | 90.2% | 27G | Twitter |
| `GT50` | 50 | 90.2% | | |
| `GT100` | 100 | 90.2% | | |
| `GT200` | 200 | 90.2% | | |
| `W2V` | 300 | 96.1% | 3G | Google News |

In order to get an overview in this matter, we first consider the question whether the lexical unit structure of FRAMENET can in any way be found in typical embedding spaces. For that we look at some statistics of a variety of embedding datasets for English words. Specifically, we consider neighborhood relations and geometric properties of embeddings and how they relate to frame membership.

A list of the datasets we use with some of their key properties can be found in Table 1. For easier referencing, each of them is given a short identifier (`FT`, `G50`, ...). The embeddings are reference datasets for the respective method and were pretrained on the given generic corpora by the developers: `W2V`[6] with *Word2Vec* (Mikolov et al., 2013), `FT`[7] with *fastText* (Bojanowski et al., 2016), and all others with *GloVe*[8] (Pennington et al., 2014). All datasets can freely be downloaded from the websites given in the footnotes where also additional information can be found. Further important differ-

[6] `https://code.google.com/archive/p/word2vec/`

[7] `https://fasttext.cc/docs/en/pretrained-vectors.html`

[8] `https://nlp.stanford.edu/projects/glove/`

Table 2: Statistics for the different embedding datasets. See the text for details.

| dataset | $n_1$ | $n_2$ | $n_3$ | $\hat{n}$ | $\alpha_u$ | $\alpha_s$ | $d_u$ | $d_s$ | $\ell$ |
|---|---|---|---|---|---|---|---|---|---|
| FT | 0.47 | 0.35 | 0.29 | 0.63 | $90\pm5$ | $80\pm10$ | $5.5\pm0.7$ | $5.0\pm0.8$ | $3.5\pm0.4$ |
| G50 | 0.37 | 0.27 | 0.23 | 0.52 | $90\pm12$ | $75\pm16$ | $6.2\pm0.9$ | $5.3\pm1.1$ | $3.7\pm0.6$ |
| G100 | 0.42 | 0.32 | 0.27 | 0.58 | $90\pm9$ | $78\pm13$ | $6.8\pm0.9$ | $6.0\pm1.1$ | $4.3\pm0.5$ |
| G200 | 0.43 | 0.32 | 0.27 | 0.59 | $90\pm7$ | $80\pm11$ | $8.1\pm0.9$ | $7.4\pm1.1$ | $5.2\pm0.7$ |
| G300 | 0.44 | 0.33 | 0.27 | 0.60 | $90\pm6$ | $81\pm10$ | $8.8\pm0.9$ | $8.1\pm1.2$ | $5.6\pm0.9$ |
| G42 | 0.49 | 0.37 | 0.31 | 0.65 | $90\pm6$ | $80\pm10$ | $8.7\pm0.9$ | $8.0\pm1.1$ | $5.6\pm0.7$ |
| G840 | 0.17 | 0.13 | 0.10 | 0.24 | $90\pm5$ | $87\pm7$ | $10.5\pm1.8$ | $10.3\pm2.1$ | $7.7\pm3.3$ |
| GT25 | 0.17 | 0.14 | 0.12 | 0.28 | $90\pm17$ | $77\pm20$ | $4.9\pm1.0$ | $4.4\pm1.1$ | $3.1\pm0.5$ |
| GT50 | 0.23 | 0.18 | 0.15 | 0.36 | $90\pm13$ | $79\pm16$ | $6.0\pm0.9$ | $5.4\pm1.1$ | $3.8\pm0.5$ |
| GT100 | 0.27 | 0.20 | 0.17 | 0.40 | $90\pm9$ | $81\pm12$ | $7.1\pm0.9$ | $6.6\pm1.0$ | $4.6\pm0.6$ |
| GT200 | 0.29 | 0.21 | 0.17 | 0.42 | $90\pm7$ | $83\pm10$ | $8.3\pm0.9$ | $7.8\pm1.0$ | $5.5\pm0.6$ |
| W2V | 0.51 | 0.38 | 0.32 | 0.66 | $90\pm5$ | $80\pm10$ | $4.3\pm0.5$ | $3.8\pm0.6$ | $2.6\pm0.2$ |

ences between the embeddings lie in the number of components $D$ per vector and the vocabulary that is covered. The latter loosely depends on the the number of tokens $T$ in the corpus. For the comparison with FRAMENET we are only interested in those embedding vectors that correspond to one of the 8579 lexemes[9] that form the basis of the lexical units. The value $C$ in Table 1 is the percentage of FRAMENET lexemes that is represented in the respective corpus. This means that we extract about 8000 embedding vectors out of each dataset for our analysis, depending on the specific dataset. We use the lexemes from FRAMENET and the datasets as they are and without altering the case. Depending on the dataset, case variants may or may not be treated differently, so that mixing them might lead to inconsistencies, which we want to avoid. As a consequence, typical FRAMENET lexemes that do not exist in the embedding sets are capitalized words, uncommon words, and in particular certain spellings of compounds like *photoelectricity* or *medium-build*. Ambiguous frame memberships of lexemes do not play a role in our experiments because for the statistics we are interested in, it only matters whether lexemes share a frame but not which one.

As all experiments that are described in this section rely on geometric notions, we have to choose some kind of metric on the embedding spaces. First tests revealed that the overall results do not essentially depend on specific choices in this respect (while specific details do). With that in mind, we opt for the intuitive and well-known Euclidean metric. Compared to the often-used cosine similarity, the Euclidean distance is an actual metric and is thus better suited for discussing the interplay of distances and angles. Additional benefits of this choice are discussed in (Trost and Klakow, 2017).

### 3.1 Nearest Neighbor Relation

First, we consider the relation between the FRAMENET frame structure and the nearest neighbor (NN) relation in embeddings. For all datasets, we compute the three nearest neighbors in terms of the Euclidean distance for all words. Some statistics are given in Table 2. $n_1$ gives the probability of finding the lexemes of two embeddings that are connected by the NN relation within the same frame of FRAMENET. $n_2$ is the same quantity for the second nearest neighbor and so forth. $\hat{n}$ is the probability for finding any of the first three NNs in the same frame.

The values for $n_1$ vary from a low 0.17 for G840 and GT25 to a better-than-chance value of 0.51 for W2V. In general, $n_1$ increases with $D$ and also with $T$, with the exception of G840, which shows a deviating behavior in general (see below). The regular behavior makes sense if we expect a higher dimensionality to yield more flexibility in aligning the words and if more training data is expected to lead to better results. The values of $n_2$ and $n_3$ are consistently lower than $n_1$, while $\hat{n}$ must be higher (or equal) by definition. The ratio between these values is surprisingly similar across all datasets.

In order to get a better feeling for how these numbers come about, we can have a look at specific ex-

[9] For FRAMENET 1.7.



Proceedings of the 14th Conference on Natural Language Processing (KONVENS 2018)
Vienna, Austria – September 19-21, 2018

Table 3: Example of NNs from FT. Neighbors that share a frame with the word are underlined.

| word | $NN_1$ | $NN_2$ | $NN_3$ |
|---|---|---|---|
| ablaze | <u>alight</u> | <u>conflagration</u> | <u>fire</u> |
| able | <u>unable</u> | could | would |
| ablution | prayer | ritual | pray |
| advance | advanced | <u>proceed</u> | push |

amples of NN relations (Table 3). The table shows examples where all, some, and no NNs share some frame. The general observation is that the NN relations confirm *some* kind of connection between the words as predicted by the distributional hypothesis. However, the type of these connections varies a lot and might be due to any syntactical or semantical structure in the data. The examples show, however, that the simple NN search already yields interesting candidates for enriching frames with new words and for further connections between lexemes.

### 3.2 Angles and Distances

Next we examine how the vectors within a frame are related to each other geometrically. As there are only about 8000 vectors in our datasets (see above) we can consider all possible pairs of embeddings $\mathbf{x}$, $\mathbf{x}'$ and calculate the angle

$$\alpha = \arccos\left(\frac{(\mathbf{x}-\bar{\mathbf{x}})\cdot(\mathbf{x}'-\bar{\mathbf{x}})}{\|\mathbf{x}-\bar{\mathbf{x}}\|\,\|\mathbf{x}'-\bar{\mathbf{x}}\|}\right) \tag{1}$$

and the distance

$$d = \sqrt{(\mathbf{x}-\mathbf{x}')\cdot(\mathbf{x}-\mathbf{x}')} \tag{2}$$

between the two for each of them. We choose the mean $\bar{\mathbf{x}}$ of the overall dataset as the point of reference for calculating the angles in order to eliminate misleading and irrelevant effects that stem from non-centered data.

For both the angles and the distances we divide our samples in two groups, depending on whether there exists a frame that contains both of the corresponding lexemes or not. Quantities with the **s**ame-frame relation get the subscript s. Quantities for **u**nlinked lexemes get the subscript u.

In order to get a first understanding of what the distributions of $\alpha$ and $d$ look like, we use Gaussian kernel density estimation with Scott's rule (Scott, 2015) for getting a sketch of the probability density functions. Two examples are given in Fig. 1 and Fig. 2. For these and also for the other datasets we observe bell-shaped curves that are close enough to Gaussians for giving their mean values and standard deviations a proper interpretation.

The means of $\alpha$ and $d$ with the respective standard deviations are shown in Table 2. Due to the sufficiently large amount of data, the hypothesis that the means are the same can clearly be rejected on the basis of Welch's $t$-test with a $p$-value that is orders of magnitude below 0.01 and a statistical power that is close to 1. For G840 the difference is still significant but not as obvious as for the other datasets. It is striking that both for $\alpha$ and for $d$ the same-frame value is about one standard deviation smaller than its counterpart (with the exception of G840). The standard deviation of the same-frame values are typically slightly larger than the others. This can be interpreted on the same basis as the figures: The maximum value is similar for both populations but the minimum and the mean are shifted towards smaller values for the same-frame pairs. This means that the same-frame pairs are on average closer to each other, but there are still many samples that do not reflect the FRAMENET structure, so that the overall variance is higher.

The distribution of the angle $\alpha_u$ is centered at 90° with a standard deviation that decreases with the dimensionality of the vectors. Embeddings spaces appear to be isotropic (Trost and Klakow, 2017) and high-dimensional random vectors are more likely to be orthogonal, so this behavior is expected. Conversely, this means that the lower values of $\alpha_s$ indicate that the lexemes of a frame tend to be clustered in a specific region of the embedding space.

The main message from these statistical experiments is that there is clear evidence for a connection between the embedding structure and the frame structure but that it is difficult to exploit without further investigations because in general, specific samples may be completely unrelated to their FRAMENET role. However, for very small values of $\alpha$ or $d$ a same-frame relation is almost certain (cf. Fig. 1 and Fig. 2). G840 behaves differently from the other datasets in all experiments. The reason for that might lie in the extremely large corpus on which these embeddings were trained and might be due to technical intricacies that occur for such large training sets only.

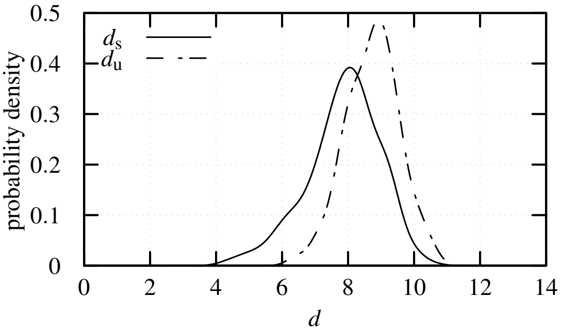


Figure 1: Inter-vector distance distributions for `G42`.

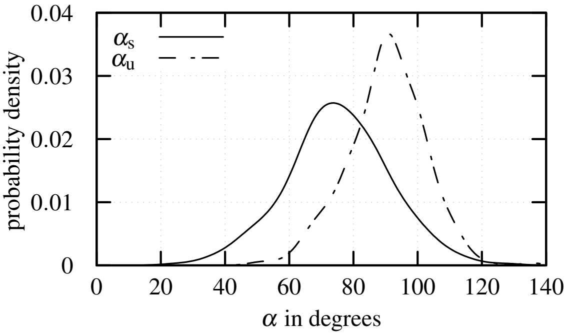


Figure 2: Inter-vector angle distributions for `G50`.

### 3.3 Center of Mass of Frames

In a third experiment, we investigate how word vectors are positioned relative to the centers of mass of the words in the frames to which they belong. In particular, we consider all lexemes that are part of more than one frame. For each of these lexemes, we calculate the Euclidean distance $\ell$ between its vector and the mean of all the embedding vectors of the other lexemes in these frames. The mean and the standard deviation of $\ell$ are listed in Table 2. A comparison with the values of distances $d$ between embeddings reveals that $\ell$ is typically around two standard deviations smaller than $d_s$ and thus smaller than the vast majority of inter-embeddings distances. This means that lexemes that are part of different frames are positioned in such a way that their distance to all the involved lexemes is very small. This is further evidence for the relevance of frame semantical relations in embedding spaces.

### 3.4 Implicit Properties

Some aspects of a frame may be so self-evident that they literally "go without saying", i.e., they are such a matter of course that speakers will rarely find the need to verbalize them explicitly. Such cases may not align well with the distributional hypothesis because the embedding might fail to map terms close to each other if the semantic connection is rarely made explicit in language. For instance, consider a sentence such as the following, referring to a new dress worn for the first time:

*William said I look like a banana.*

The image evoked by the word *banana* allows the listener to deduce that the dress in question is yellow in color. However, it is not true that bananas are always yellow–for instance, in Central America, a red variety is very common–and even the yellow specimen's peel is green before the fruit has fully ripened and will turn dark brown as it decays. Yet, there is hardly any need to refer to a banana explicitly as "a yellow banana" since this is prototypically assumed to be its default color.

Table 4: The five color terms nearest to the term *banana* across the different datasets.

| dataset | $NN_1$ | $NN_2$ | $NN_3$ | $NN_4$ | $NN_5$ |
|---|---|---|---|---|---|
| `FT` | black | orange | indigo | pink | green |
| `G50` | orange | indigo | ebony | red | yellow |
| `G100` | orange | indigo | pink | puce | ebony |
| `G200` | orange | puce | pink | green | mauve |
| `G300` | orange | violet | green | brown | purple |
| `G42` | orange | purple | brown | violet | yellow |
| `G840` | beige | maroon | blue | purple | green |
| `GT25` | orange | purple | green | yellow | pink |
| `GT50` | green | orange | red | purple | yellow |
| `GT100` | orange | green | red | yellow | pink |
| `GT200` | orange | green | yellow | purple | red |
| `W2V` | orange | brown | red | pink | white |

We would thus expect that word embeddings, being based mostly on term co-occurrence, will give greater (Euclidean and thus semantic) proximity to the color terms *green*, *brown*, and even *red* with respect to *banana* than *yellow*. In that case, understanding the implication of a sentence like the example above correctly would be more difficult using word embedding semantics than with frame semantics.

Here, we look at the distance between all color terms that are present in our data and the word *banana* and check the five closest color terms across all datasets. As predicted, the word *yellow* is never the nearest in distance (Table 4): for *fastText*, the five nearest color terms are *pitchblack*, *orange*, *indigo pink*, *green*, while for `W2V`, the terms are *orange*, *brown*, *red*, *pink*, *white*. For the various

*GloVe* datasets, *orange* is consistently the nearest color term, except for `G840`, where it is *beige*, and for `GT50`, where it is *green*. For *orange*, it is worth pointing out that this word can also refer to the fruit (as a noun) which may explain its prominence in the ranking. Embedding algorithms do not distinguish between different lexical units with the same surface form, and since proximity of embeddings is a result of similar contexts, fruit terms are expected to exhibit high similarity to each other. Further, it is remarkable that in almost all cases where *yellow* does appear among the five nearest color terms, *brown* or *green* are found to be even closer, with the exception of `GT50`. At the same time, though, we also observe a number of exotic color terms where the proximity to *banana* is difficult to explain.

To demonstrate the validity of this singular experiment, we also tested terms other than *banana* that are typically associated with a specific color. For example, and in contrast to the above findings, the term *ocean* has *blue* or *turquoise* as the nearest color term across all datasets. One possible explanation is that people are more likely to talk about the color of the ocean, for instance when describing a holiday location, than the color of a banana, and thus the connection can be captured in a corpus-based approach. Similarly, for the word *grass*, we find *green* to be the nearest color term in almost all cases, except for `G840` (*beige*, again hinting at a possible idiosyncrasy of that dataset), `GT50` (*yellow*), and `W2V` (*brown*).

## 4 Discussion and Future Work

In this paper, we looked at four selected comparisons between frame semantics and distributional semantics in order to investigate in how far a linguistically motivated theory shares properties with a word representation based purely on statistical information of word distributions. To this end, we first performed three experiments concerning the relative location of term vectors in the embedding space and whether words that belong to the same frame in FRAMENET map to vectors that show locational coherence. We investigated three measures: the nearest neighbors of all words in the considered embedding space, the Euclidean distance, and the angles between pairs of word vectors.

The nearest neighbor experiment suggests an interesting application for this type of research, namely the use of embedding representations to discover new words that were previously not associated as lexical units with a given frame. However, further research is required to better determine what kind of relationship exactly is expressed by the neighboring relation.

Both the Euclidean distance and angle experiments show that frame semantics and embedding representations agree to a certain extent on which words should be grouped together semantically. However, the distinction is not clear-cut in the embedding space as some frame-external word vector pairs do show smaller distances than some frame-internal word vectors. As we only look at the aggregated statistics over all word samples considered, we propose a more fine-grained analysis as a worthwhile direction for future work.

So far, we explored further similarities between frame semantics and distributional semantics by investigating the special case of words that evoke more than one frame. Here, we show that this frame semantic property can be found in the embedding space as well.

In a fourth experiment, we also observe some potential differences as far as frame-related information is concerned that is not typically verbalized. Although being a central part of the underlying experience or scene, such cases provide a challenge for methods that derive the embeddings via corpus statistics and co-occurrence measures only.

There is quite a rich body of prior work related to our own which provides broad shoulders to stand on. However, to our knowledge, this is the first attempt at comparing commonalities between these two very different approaches to lexical semantics.

In this work, we have only looked at the target words that evoke frames, but FRAMENET offers additional information, for instance on frame elements or on various relationships between frames, that lend themselves for additional studies along similar lines to the presented work. Similarly, the relationships of different words belonging to the same frame deserve further, more detailed examination in the future.

Deriving a good understanding of the semantic properties of embedding spaces may help improving existing approaches for the discovery of new lexical units for a given frame. In the long run, it might even become a useful resource for creating new FrameNets for other languages.

Likewise, the rich structures present in frame semantic treatments might offer useful additional information not currently present in embedding approaches, especially as far as semantic information is concerned that cannot easily be derived from corpora.

## References


Collin F. Baker, Charles J. Fillmore, and B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, volume 1, pages 86–90, Montreal, Quebec, Canada, August 10–14.

Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. 2003. A neural probabilistic language model. *The Journal of Machine Learning Research*, 3:1137–1155.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2016. Enriching Word Vectors with Subword Information. *arXiv preprint arXiv:1607.04606*.

Danushka Bollegala, Alsuhaibani Mohammed, Takanori Maehara, and Ken ichi Kawarabayashi. 2016. Joint word representation learning using a corpus and a semantic lexicon. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*, pages 2690–2696.

Teresa Botschen, Hatem Mousselly-Sergieh, and Iryna Gurevych. 2017. Prediction of frame-to-frame relations in the framenet hierarchy with frame embeddings. In *Proceedings of the 2nd Workshop on Representation Learning for NLP*, pages 146–156, Vancouver, Canada, August 3.

Katrin Erk and Sebastian Padó. 2006. SHALMANESER–A Toolchain For Shallow Semantic Parsing. In *Proceedings of LREC 2006*, Genoa, Italy.

Manaal Faruqui, Jesse Dodge, Sujay K. Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. 2015. Retrofitting word vectors to semantic lexicons. In *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*, pages 1606–1615, Denver, Colorado, May 31–June 5. Association for Computational Linguistics.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Óscar Ferrández, Michael Ellsworth, Rafael Muñoz, and Collin F. Baker. 2010. Aligning FrameNet and WordNet based on Semantic Neighborhoods. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Charles J. Fillmore. 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, volume 280, pages 20–32.

Charles J. Fillmore. 1982. Frame semantics. In *Linguistics in the Morning Calm*, pages 111–137. Hanshin Publishing Co., Seoul, South Korea.

Lucie Flekova and Iryna Gurevych. 2016. Supersense embeddings: A unified model for supersense interpretation, prediction, and utilization. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2029–2041, Berlin, Germany, August. Association for Computational Linguistics.

Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3).

Zellig Harris. 1968. *Mathematical structures of language*. John Wiley & Sons, New York, US.

Karl Moritz Hermann, Dipanjan Das, Jason Weston, and Kuzman Ganchev. 2014. Semantic frame identification with distributed word representations. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1448–1458, Baltimore, Maryland, June. Association for Computational Linguistics.

Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli. 2015. SENSEMBED: Learning sense embeddings for word and relational similarity. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 95–105, Beijing, China, July 26–31.

Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pages 1–8, Sydney, July. Association for Computational Linguistics.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.

Roberto Navigli and Simone Paolo Ponzetto. 2012. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Marco Pennacchiotti, Diego De Cao, Roberto Basili, Danilo Croce, and Michael Roth. 2008. Automatic induction of FrameNet lexical units. In *Proceedings* *of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 457–465, Honolulu, Hawaii, October.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

Miriam R.L. Petruck. 1996. Frame semantics. In *Handbook of pragmatics*, pages 1–13. John Benjamins Publishing Co.

Michael Roth. 2008. Representing FrameNet in semantic spaces. Master's thesis, Saarland University, Germany.

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, Collin F. Baker, and Jan Scheffczyk. 2016. FrameNet II: Extended Theory and Practice, November 1. `https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf`.

David W Scott. 2015. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

Dan Shen and Mirella Lapata. 2007. Using semantic roles to improve question answering. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 12–21, Prague, June. Association for Computational Linguistics.

Lei Shi and Rada Mihalcea. 2005. Putting pieces together: Combining framenet, verbnet and wordnet for robust semantic parsing. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing. CICLing 2005.*, volume 3406 of *Lecture Notes in Computer Science*, Berlin, Heidelberg. Springer.

Thomas Alexander Trost and Dietrich Klakow. 2017. Parameter Free Hierarchical Graph-Based Clustering for Analyzing Continuous Word Embeddings. In *Proceedings of TextGraphs-11: the Workshop on Graph-based Methods for Natural Language Processing*, pages 30–38, Vancouver, Canada, August. Association for Computational Linguistics.