

Institut für Maschinelle Sprachverarbeitung

Universität Stuttgart  
Pfaffenwaldring 5b  
70569 Stuttgart

Masterarbeit

**Characteristics of Neighbourhood  
Vector Spaces for Abstract and  
Concrete Words**

Maximilian Bräuninger

**Studiengang:** Informatik, M. Sc.

**Prüfer:** apl. Prof. Dr. Sabine Schulte im Walde

**Betreuer:** apl. Prof. Dr. Sabine Schulte im Walde  
Dr. Diego Frassinelli

**begonnen am:** 15.05.2019

**beendet am:** 15.11.2019

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Data and Methods</b>	<b>6</b>
2.1	Distributional Semantics . . . . .	6
2.2	Concreteness Score . . . . .	12
<b>3</b>	<b>Methodology</b>	<b>14</b>
3.1	Target, Neighbour and Context Dimensions . . . . .	15
3.2	Cosine Similarity . . . . .	18
3.2.1	Definition . . . . .	19
3.2.2	Context-Context Similarity . . . . .	20
3.2.3	Target-Context Similarity . . . . .	22
3.2.4	Target-Neighbour Similarity . . . . .	23
3.2.5	Results . . . . .	24
3.3	Entropy . . . . .	61
3.3.1	Definition . . . . .	61
3.3.2	Target Entropy . . . . .	63
3.3.3	Second Degree Entropy . . . . .	63
3.3.4	Results . . . . .	64
3.4	Dimensionality Reduction . . . . .	70
3.4.1	Singular Value Decomposition . . . . .	71
3.4.2	word2vec . . . . .	72
3.4.3	Results . . . . .	72
<b>4</b>	<b>Conclusion</b>	<b>80</b>

# 1 Introduction

Language and speech consist of a plethora of different word categories, meanings and concepts and many distinctions can be made across all kinds of different words and phrases. One important concept is the distinction of abstract and concrete words and phrases.

Concreteness refers to perceivability of a word. Words that are very concrete, or concrete words, are words that can be directly experienced by a human being using one of their five different senses, i.e seeing, hearing, smelling tasting and touching (Brysbaert et al., 2014). With other words, concrete words are words like, for example, "dog" or "banana". Both can be experienced pretty much with all five senses. One can hear a dog, one can see a dog, one can smell a dog and one can touch a dog. All of these things can also be said about a banana. Besides, it is also certainly possible to taste a banana. Technically this is also the case for a dog, however we will focus on the banana for this particular sense. Note, however that it is not necessary for a word to be experienceable through all five senses, especially it is not required that a word can be seen in order for it to be concrete. For example, you can not see the word "sweet", but you can experience it or explain it very easily by tasting something sweet like sugar. Another way to describe a concrete word is that it exists in reality (Brysbaert et al., 2014).

Abstract words on the other hand are words that can not be directly experienced or explained using one of the five senses. For example the word "love" can not be experienced using one of the five senses. In order to explain abstract words to someone it is easiest to describe them by using other words instead of "demonstrating" them (Brysbaert et al., 2014).

For example letting someone who is unfamiliar with the word "love" experience it will be pretty hard. There is no tangible instance of love that can be smelled, touched, heard, tasted or seen without the person already knowing what love means. Therefore, in order to explain love to someone, one would use other words that describe its meaning. Other examples for abstract words

include "justice", "freedom" or "dream".

Concreteness versus abstractness is an important topic in cognitive science, psycholinguistic and computerlinguistic, especially regarding the mechanisms behind remembering and processing concrete and abstract words (Naumann et al., 2018). Many works and studies deal with the difference in concrete and abstract concepts, trying to find differences in memorizing and processing of words (Brysbaert et al., 2014).

For instance, according to the "Dual-Coding Theory" (Paivio, 2013) concrete words should be more easy to remember than abstract words, as concrete words can be stored both visually, as an image in the head, as well as verbally, as a word. Abstract words on the other hand can only be stored verbally and hence, according to the "Dual-Coding Theory", are harder to remember. Various studies, amongst others by Yui et al. (2017), Schultz Jr. and Woodall (1980) and Paivio et al. (1994) show evidence of this actually being the case.

In their study, Yui et al. (2017) surveyed 298 participants. Each participant was either given a list of 30 concrete or 30 abstract words and asked to memorize as many of the words as possible in one minute. Then, participants were asked to perform a simple mathematical task of about 45 seconds. Afterwards participants were asked to write down as many of the words they were shown earlier as possible within a two minute time frame. The results of this study show a significant difference between abstract and concrete words with people that had to memorize concrete words being able to memorize more words on average than participants that were given abstract words (Yui et al., 2017).

Another important, and still unsolved question with regard to concrete and abstract words are the mechanisms playing a role in processing them (Barsalou and Wiemer-Hastings, 2005). Especially the question regarding the processing mechanisms behind abstract words remains unanswered (Vigliocco et al. (2013), McRae and Jones (2013)). The "Context Availability Theory" (Schwanenflugel, 2013) argues that, in order to evoke the meaning of a word,

finding the appropriate context of this word is crucial. It has been shown that creating an appropriate context for abstract words is more difficult than for concrete words, as abstract words lead to higher reaction times and a larger number of errors (Naumann et al., 2018). Regarding this, Hill et al. (2014) used a computational study in order to show that concrete words have a small, distinct context they appear in, whereas abstract words appear in a broader, more general context (Hill et al., 2014). These findings have been supported by Hoffman et al. (2013) as well as Hoffman and Woollams (2015), again finding that concrete words appear in specialized, very similar contexts in comparison to abstract words which seem to appear in more generalized, broad contexts (Hoffman et al. (2013) Hoffman and Woollams (2015)).

Further work regarding the topic of concrete and abstract words has for example been done by Frassinelli et al. (2017) and Bhaskar et al. (2017). In their work Frassinelli et al. (2017) focus on finding differences and similarities of concrete and abstract words by looking at the respective contexts of the words. According to previous theories, both abstract and concrete words are expected to mainly appear in a concrete context, i.e. they mainly co-occur with concrete words (Barsalou and Wiemer-Hastings, 2005). However, investigating the contexts of both concrete and abstract words, Frassinelli et al. (2017) find that, while concrete words mainly appear in concrete contexts, abstract words tend to appear within an abstract context, not aligning with the existing psycholinguistic theories (Frassinelli et al., 2017). Bhaskar et al. (2017) focus on multi-modal models consisting both of text and images, hoping to find differences in the information provided by visual and textual representations with regard to abstract and concrete words (Bhaskar et al., 2017). However, while producing a very successful concrete and abstract predictor model, they were not able to find any differences between textual or visual information or the combination of both (Bhaskar et al., 2017).

As illustrated, the topic of concreteness and abstractness offers a wide area of interesting and important questions and hypotheses to be examined. However, in order to provide a detailed overview and analysis, it is important

to focus on a smaller area of research. In this case this means choosing an already existing hypothesis and providing a thorough, in-depth analysis with regard to this hypothesis. The main hypothesis chosen as the underlying theory of this work deals with the different contexts of concrete and abstract words.

**Hypothesis** Concrete words predominantly occur within a specialized, small context. Abstract words on the other hand appear in a broader, more distinct set of contexts (Naumann et al., 2018).

At first, important models, like the Distributional Semantic Model, and methods will be introduced and described. Afterwards, different models of dimensionality and several semantic diversity measures will be presented and explained. Then, the different diversity measures will be employed on the different dimensionality models and the results will be presented and discussed in-depth. Lastly, different means of dimensionality reduction will be introduced and applied to the vector space models. A comparison between the regular dimensionality models and the dimensionality reduction models shows interesting differences in the results they provide.

## 2 Data and Methods

### 2.1 Distributional Semantics

Distributional Semantics aims to describe and quantify the semantic meaning or the semantic similarity of words and phrases that are used in different languages. While proficient speakers of a language usually can make sense of a word's meaning without thinking about it, describing semantics in a more general or mathematical way is less intuitive.

In order to quantify semantics, the distributional hypothesis (Harris, 1954) can be applied. In short, the distributional hypothesis states the suggestion

that words share a similar semantic meaning if they appear to often occur within the same contexts, or, as Firth put it: "a word is characterized by the company it keeps" (Firth, 1957).

Simply put, if two words appear with the same context words very often, the likelihood of these two words to be semantically closer to each other is higher than for two words that do not share context words. For example the words "banana" and "strawberry" supposedly appear within the context of eating rather often. Therefore context words like "to eat", "hungry" or "tasty", all hint at the semantic context of food which is both used for "banana" as well as "strawberry". On the other hand the word "book" is less likely to appear within a food context.

Therefore, the distributional hypotheses argues that "banana" and "strawberry" are more semantically similar to each other than "book" is to either of them, since they are "[...]keeping the same company"(Firth, 1957), i.e. appear with the same context words more often.

Distributional Semantic Models (DSM) (Landauer and Dumais, 1997) use

<b>The</b>	<b>furry</b>	<b>dog</b>	<b>barks</b>	<b>at</b>	<b>the</b>	<b>cat.</b>
------------	--------------	------------	--------------	-----------	------------	-------------

Figure 1: Exemplary sentence with target word "dog" and window size one to each direction.

the distributional hypothesis by modelling the semantics of words using the context words that appear with them. In order to obtain an extensive distributional semantics model, large text corpora are used. In order to create a DSM, for all selected target words, a certain window (for example 1, 10, 20...) of words is looked at. All words that occur within this window are then added to the list of context words for the given target word. Figure 1 illustrates this process for the target word "dog" and a window size of one for both directions. As a result, the words "furry" and "barks" or its lemma "to bark" are added to the context dimensions for the word "dog".

One by one, all occurrences of the target word within the text corpus are considered and context dimensions are added, i.e. set to one. If a context word appears again, the count of the corresponding word is increased by one.

Overall, the result of this process is a vector space model with high dimensional vectors for all target words Singhal (2001). Each dimension of a given target word vector corresponds to exactly one context word, i.e. a context dimension. If a target word does not appear in the context of a specific context word, the corresponding context dimension equals zero. Figure 2 depicts two

<b>dog =</b>	<b>3</b>	<b>walk</b>	<b>human =</b>	<b>1</b>	<b>walk</b>
	<b>4</b>	<b>eat</b>		<b>3</b>	<b>eat</b>
	<b>1</b>	<b>fruit</b>		<b>1</b>	<b>fruit</b>
	<b>1</b>	<b>pet</b>		<b>5</b>	<b>pet</b>
	<b>2</b>	<b>coat</b>		<b>0</b>	<b>coat</b>
	<b>6</b>	<b>bark</b>		<b>0</b>	<b>bark</b>
	<b>0</b>	<b>engine</b>		<b>3</b>	<b>engine</b>

Figure 2: Two exemplary word vectors for the target words "dog" and "human". The words next to each vector entry represent the context dimension the dimension corresponds to.

word vectors for the words "dog" and "human". As shown, both vectors have non-zero values in the same context dimension rows. These dimensions, like for example row one, could correspond to context words that appear both within the context of a dog and a human, like for example "walk". Others are zero for one of the two word vectors. These rows correspond to context words that appear within the context of one but not the other target word. For example row six could correspond to the word "to bark", which appears



within the context of a dog but not within the context of a human.

Another possible representation is a word-word matrix, with each row of the matrix corresponding to one specific target word and each column corresponding to a certain context word respectively context dimension.

	walk	eat	fruit	pet	coat	bark	engine
dog	3	4	1	1	2	6	0
human	1	3	1	5	0	0	3
banana	0	7	8	0	0	0	0
car	0	0	0	0	1	0	10

Figure 3: Exemplary word-word matrix for the four target words "dog", "human", "banana" and "car".

Figure 3 illustrates a word-word matrix for four target words and seven context dimensions. For example, in the underlying corpus the target word "dog" appears with the context word "walk" three times as shown in row one, column one. Overall, the target word "human" cooccurs 13 times with one of the context words, i.e. the sum of the second row.

Each row of the word-word matrix corresponds to a certain word vector, i.e. the word vector for the target word corresponding to the row. In figure 3, the first row, which corresponds to the target word "dog" is similar to the word vector of the target word "dog" in figure 2.

When using Distributional Semantic models, different values can be used in order to fill the model. One possibility is to use the regular co-occurrence frequency counts. However, while frequency counts will provide an accurate count of all context words that appear with the target word, frequency counts do not account for the different amount of information each specific context word holds. For example articles like "the" or "to" or prepositions like "on" are very likely to appear very often as context words for all different kinds of target words, while providing very little information regarding the tar-

get word itself. One possibility of dealing with this problem is to filter the context dimensions, automatically sorting out certain words that are known or at least assumed to appear often and provide little to no information on the actual target word. The words that are filtered out are called stopwords and lists containing a number of stopwords are called stoplists. Words that are commonly included on such stoplists are, amongst others, for example articles like "a", "an", "the", auxiliary verbs like "be" or prepositions like "over" or pronouns like "he", "they" or "my".

Another way is to alter the actual values within the matrix. In order to do this, the co-occurrence frequency counts are used to calculate different probability values. These probability values can then be used in order to describe the "Mutual Information" of a target word and the context word. For example, one possibility is to calculate "Local Mutual Information" (LMI) values. These values take into account the co-occurrence frequency of the target and context word as well as the frequencies of both the target and the context word separately. This results in values that weigh the amount of information provided by the context word regarding the target word.

LMI values are calculated similar to "Pointwise Mutual Information" values using the marginal probability  $P(A)$ ,  $P(B)$  of words  $A$  and  $B$  as well as the joint probability  $P(A \cap B)$  of words  $A$ ,  $B$ .

Remembering the word-word matrix, the marginal probability for target words can be calculated by summing up all entries in the row corresponding to the target word. This value is then divided by the sum of all matrix elements. This is illustrated in red in figure 4. In this case, the sum of the elements in the row corresponding to the target word "dog" equals 17. The sum of all elements within the matrix equals 56. Therefore the marginal probability for the target word "dog",  $P(\text{dog}) = \frac{17}{56}$ .

Calculating the marginal probability of a context word can be done likewise, with the only difference being that, instead of summing up the matrix row corresponding to the target word, now the matrix column corresponding to the context word is summed up. This is illustrated in blue in figure 4. The

	walk	eat	fruit	pet	coat	bark	engine
dog	3	4	1	1	2	6	0
human	1	3	1	5	0	0	3
banana	0	7	8	0	0	0	0
car	0	0	0	0	1	0	10

Figure 4: Calculating the marginal probability for the target word "dog",  $P(\text{dog})$ , the context word "eat",  $P(\text{eat})$  and the joint probability of both words,  $P(\text{dog} \cap \text{eat})$ .

sum of all column elements in the second row, which corresponds to the context word "eat", equals 14, while the sum of all matrix elements naturally remains at 56. Therefore  $P(\text{eat}) = \frac{14}{56}$ .

The joint probability of the target word "dog" and the context word "eat" is calculated by taking the matrix element in the row corresponding to "dog" and the column corresponding to "eat", i.e. the element at matrix position (1,2) and divide it by the sum of all matrix elements, i.e. 56. As highlighted in green in figure 4, for this model, this leads to a joint probability  $P(\text{dog} \cap \text{eat}) = \frac{4}{56}$ .

Using these three probabilities calculated using co-occurrence frequencies, LMI overall can be calculated like this:

$$LMI(A,B) = P(A \cap B) \cdot \log \left( \frac{P(A \cap B)}{P(A) \cdot P(B)} \right)$$

$$\text{Using the example, } LMI(\text{dog}, \text{eat}) = P(\text{dog} \cap \text{eat}) \cdot \log \left( \frac{P(\text{dog} \cap \text{eat})}{P(\text{dog}) \cdot P(\text{eat})} \right) = \frac{4}{56} \cdot \log \left( \frac{\frac{4}{56}}{\frac{17}{56} \cdot \frac{14}{56}} \right) \approx -0.004$$

Word vectors and word-word matrices can be used alternatively and can be easily converted into each other, by either transposing the word vectors or deriving the vectors by looking at the matrix rows.

**Corpus** As mentioned above, it is necessary to use a (large) corpus. This corpus is used to create the Distributional Semantic model respectively the vector space model.

In this case, the COW corpus (Schäfer, 2015), (Schäfer and Bildhauer, 2012) from 2016 in English language and sentence shuffled ENCOW16AX is used. Overall this web corpus contains over 9.5 billion tokens. POS tagging and lemmatisation is done using the TreeTagger (Schmid, 1994).

## 2.2 Concreteness Score

The goal of this work is to investigate the neighbourhood vector spaces of concrete and abstract words. In order to do this, it is important to have a comprehensible concreteness respectively abstractness rating of a large number of words.

Brysbaert et al. provide such an evaluation of roughly 40000 widely known English words and two-word expressions. The extensive list was created by polling over 4000 participants online. All participants had to be current US residents and were asked to rate a list of 300 words regarding the concreteness of each word on a scale from 1 to 5 (1 very abstract, 5 very concrete). Concreteness in this case was defined and explained similarly to section 1. If participants did not know a word they were asked to mark this word and not rate it. Overall 210 lists each containing 300 words were created and each list was annotated by 25 to 30 people. Afterwards words only known to at least 85% of the annotators were included in the final ratings. Overall this leads to a list of 37058 English words and 2896 English two-word expressions (Brysbaert et al., 2014). Validation of the results was done by correlating them to concreteness scores provided by Coltheart (1981), proving the validity of the obtained concreteness scores (Brysbaert et al., 2014) (Coltheart, 1981).

**Targetlist** Using the concreteness scores by Brysbaert et al., target lists for nouns, verbs and adjectives are constructed. For each part of speech two

	noun - abstract	noun - concrete	verb - abstract	verb - concrete	adjective - abstract	adjective - concrete
min	1.07	4.86	1.19	3.31	1.19	3.45
max	1.96	5.0	1.93	4.8	1.67	4.69
mean	1.73	4.93	1.74	3.73	1.54	3.84
median	1.77	4.93	1.76	3.68	1.56	3.81

Table 1: Minimum, maximum, mean and median values of all six target lists.

lists are created. One list contains words that have a very high concreteness score i.e. very concrete words and the other one contains very abstract words i.e. a very low concreteness score. Producing two different lists per part of speech is necessary in order to examine the differences between concrete and abstract words. Using the same kinds of vector space models and measures on the two different targetlists, each representing either concrete or abstract targets, allows for a comparison in the different behaviour and characteristics exhibited by concrete and abstract words. The lists for abstract and concrete nouns contain 500 words each, whereas the four lists for verbs and adjectives each contain 200 words. As shown in table 1, no abstract target list contains a word with a concreteness rating higher than 1.96 and no concrete target list contains a word with a lower concreteness rating than 3.31. Overall the mean difference between the concrete and abstract nouns is the highest with a difference of 3.2. The mean difference between abstract and concrete verbs is the smallest, however still very big with a value of 1.99. The words with the highest concreteness score overall can be found in the concrete noun list with the highest possible concreteness score of 5.0. One of this words, for example, is "whisky". Exactly one word, "spirituality", with the lowest score (1.07) can be found in the abstract noun list. The highest rated verb is "sit" with a concreteness score of 4.8 and the highest rated adjective, amongst others, is "nutty" with a score of 4.8. On the other hand, the lowest rated, i.e. most abstract adjective in the list is "enlightening" and the lowest rated verb is "idealize" both with a score of 1.19.

### 3 Methodology

In order to find characteristics of abstract and concrete words, one can use various methods, options and settings. This thesis focuses on finding characteristics of the neighbourhood vector spaces for abstract and concrete words. The main idea is to inspect different neighbourhoods of the target words as described in section 2.2 using semantic diversity/similarity measures such as the cosine similarity or predictability measures such as entropy (Naumann et al., 2018). According to the hypothesis (see section 1) one would expect the abstract and concrete target words to show a different behaviour, i.e. concrete targets are assumed to have a higher semantic similarity in comparison to their abstract counterparts, as according to the hypothesis, concrete words appear in a smaller, more specified context whereas abstract words appear in a broader more general one.

As described in section 2.1, the underlying vector space model used is created using the ENCOW16AX corpus (Schäfer, 2015), (Schäfer and Bildhauer, 2012). The window size used to create the distributional semantics relation is set at 20. Employing the same corpus and similar semantic measures, Naumann et al. (2018) found that altering the window size did not change the outcome of the results.

Different types of vector space models can be created and used in order to calculate the semantic diversity of target words with regard to their neighbourhoods. For example, different context dimensions can be included or omitted to create different models. Using different models can lead to different results, as each model contains different context dimensions, probably masking or enhancing effects and characteristic behaviour. Generally three different models, regarding the context dimensionality, are used.

1. "Full Dimensionality Model": No restrictions on the context dimensions. All context dimensions associated with the target word vector using the ENCOW corpus are used. Note that also no stoplist is used for this model.

2. "Nouns Only Model": Context dimensions are restricted to all words tagged as nouns in the ENCOW corpus.
3. "Brysbaert Model": Context dimensions restricted only to words (no two-word expressions) contained in the Brysbaert norm (Brysbaert et al., 2014). This restricts the context dimensions for the vectors to 37,058 nouns, verbs and adjectives.

As mentioned above one goal of this work is to employ semantic diversity measures to learn more about the context of target words. In this case the target words consist of nouns, verbs and adjectives that are either strongly abstract or strongly concrete according to Brysbaert et al. (2014). According to section 1 abstract words appear in a broader range of contexts whereas concrete words appear in a limited number of distinct contexts (Naumann et al., 2018). By learning more about the semantic diversity of abstract and concrete neighbourhoods one can test this hypothesis (Naumann et al., 2018). Semantic diversity in this context describes the range of variety displayed by words connected to the word, like for example the context dimensions of the target word (Hoffman et al., 2013).

### 3.1 Target, Neighbour and Context Dimensions

Subsequently, it is important to distinguish different terms regarding the distributional semantics model as described earlier in section 2.1. The following sections will mention three different aspects of the Distributional Semantic model.

- Target: Target words are words which have been chosen using a targetlist as described in section 2.2. Each target word has a word vector associated with it. The word vectors of all target words make up the vector space model. With regard to the word-word matrix representation form of the distributional semantics model, each target word is

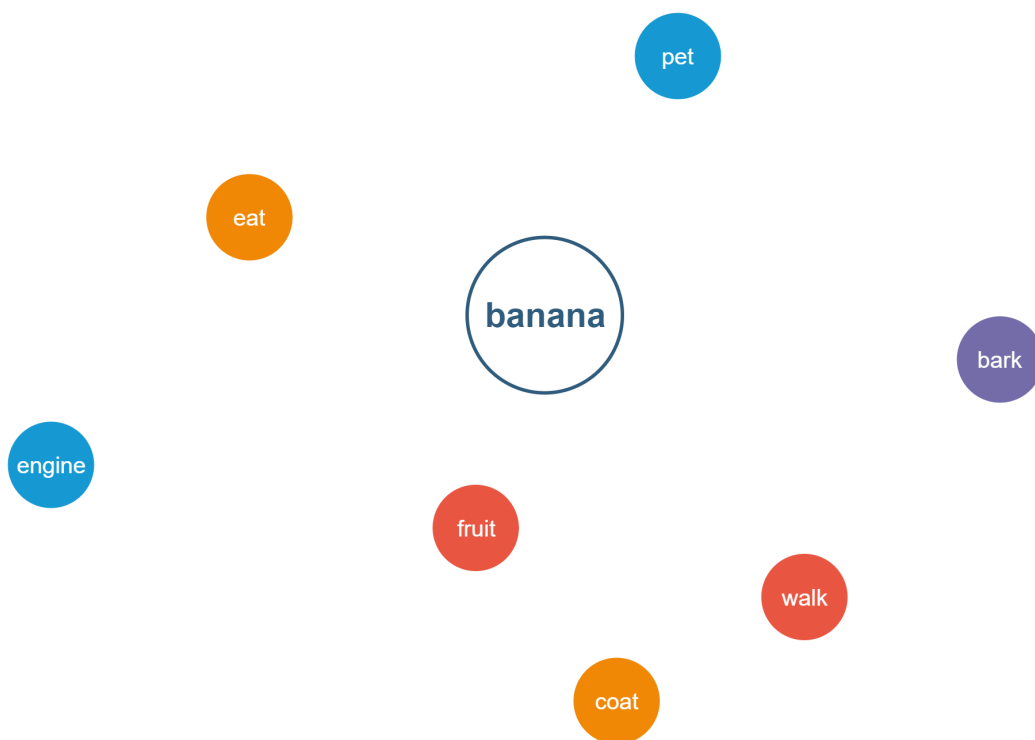


Figure 5: Exemplary representation of the target word "banana" and its seven context dimensions.

associated with exactly one row in the matrix, the target row. For the exemplary word-word matrix displayed in figure 3 and figure 4 respectively, the target words for the model are "dog", "human", "banana" and "car".

- Context dimensions: The context dimensions are derived using all words that appear with one of the target words for a given corpus and window size (see section 2.1), i.e. all contexts that target words appear with. In a word-word matrix, context dimensions are represented by the columns in the matrix, i.e. each column corresponds to one context dimension respectively context word. As displayed in figures 3 and 4, the exemplary model has seven context dimensions. These dimensions, "walk", "eat", "fruit", "pet", "coat", "bark" and "engine" are also displayed in figure 5. Note that word vectors for context words can



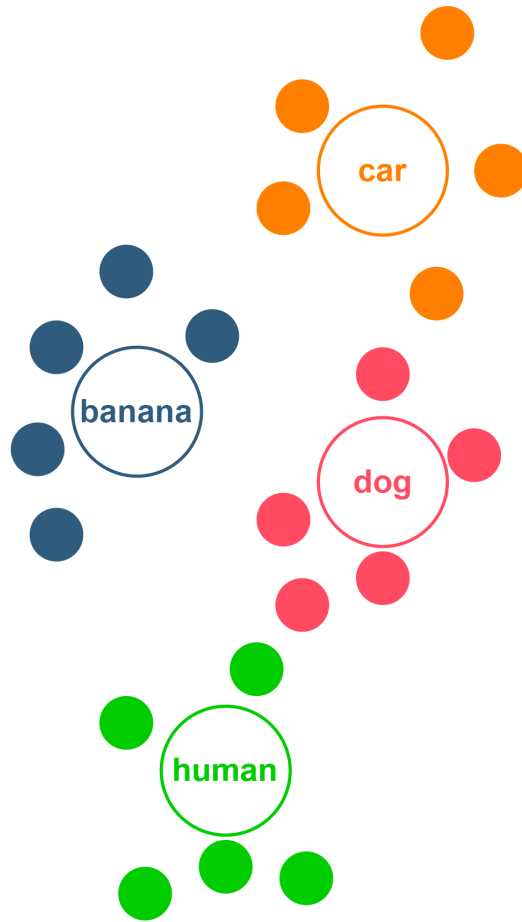


Figure 6: Exemplary representation of the neighbourhood of the four target words "banana", "dog", "car" and "human". Each target word has five context dimensions displayed around it.

be derived from the word-word matrix analogously to deriving target word vectors. However instead of deriving the vectors by using the rows of the matrix, in this case the columns of the matrix make up a context word vector.

- Neighbour: The neighbours of a specific target word consists of all other target words in the Distributional Semantic model at hand. Using a word-word matrix, this means all matrix rows which do correspond to the chosen target word represent the neighbours of this target word. Us-

ing a vector representation, all word vectors representing a target word which is not the actually chosen target word represent the neighbourhood. For example, using the exemplary model described in figures 3 and 4, the neighbours of the target word "banana" are "dog", "human" and "engine". The word banana and its neighbourhood is illustrated in figure 6. When looking at the neighbours of a certain target word, it is possible to find close neighbours by calculating the similarity of the target and its neighbour. The neighbour that has the highest similarity to the target word is called the nearest neighbour, the neighbour with the second highest similarity is the second nearest neighbour and so on.

Note that in this case, the exemplary model is very compact. Therefore, all seven of its context dimensions are mentioned in section 3.1 and displayed in figure 5. However, using realistic distributional semantics models, the number of context dimensions is distinctly higher. In the following, only a varying subset of the context dimensions will be used. In this case, the most relevant context dimensions are defined by their LMI scores, i.e. the context dimension with the highest LMI score is the most relevant.

Using the context dimensions and neighbours, different methods can be used to calculate various different similarities. This work focuses on three different methods, the "context-context similarity", the "target-context similarity" as well as the "target-neighbour similarity".

## 3.2 Cosine Similarity

A very commonly used semantic similarity measure is the cosine similarity (Singhal, 2001). The cosine similarity allows for an easy and intuitive way of calculating the semantic similarity of the target neighbourhoods. Since the "full-dimensionality model" includes all dimensions of the ENCOW corpus no stoplist is used. This expectably leads to some "typical stopwords" to be included in the target vectors with very high co-occurrence frequency counts.

For example the target noun "actuality" is associated with the verb "be" as one of the context words with the five highest co-occurrence frequency counts. Other examples include the target noun "technicality" which contains the verbs "be", "not" and "get" in the top five co-occurrence frequency counts. Using these rather un-descriptive or uninformative context dimensions alters the results of the calculations especially for cases where fewer context dimensions are considered, as they are not able to describe the target word as accurately as other words could.

To account for this as well as other frequency effects, and to obtain comparable results across all models "Local Mutual Information" (LMI) (see section 2.1) values can be used instead.

All results using cosine similarity measures presented in this work exclusively use LMI values instead of co-occurrence frequency counts.

### 3.2.1 Definition

The Cosine Similarity for two vectors  $A, B$  is defined as follows:

$$\text{CosineSimilarity}(A, B) = \cos(\phi) = \frac{A \cdot B}{\|A\| \cdot \|B\|}$$

$$\cos(\phi) \in [-1, 1]$$

Cosine Similarity values close to one represent smaller angles, i.e., in the case of semantic similarity, higher similarity.

Explicitly, for two given vectors  $A = \begin{pmatrix} 4 \\ 2 \\ 0 \\ 1 \end{pmatrix}, B = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 0 \end{pmatrix}$  the cosine similarity is calculated as  $\frac{4 \cdot 2 + 2 \cdot 3 + 0 \cdot 4 + 1 \cdot 0}{\sqrt{4^2 + 2^2 + 0^2 + 1^2} \cdot \sqrt{(2^2 + 3^2 + 4^2 + 0^2)}} = \frac{14}{\sqrt{21} \cdot \sqrt{29}} \approx 0.57$

Figure 7 illustrates three different word vectors for the words "walk", "sit" and "run". As shown in figure 7 the high semantic similarity of "walk" and

”run” results in two rather similar vectors, with a small angle between them. However the words ”walk” and ”sit” are not as semantically similar, therefore both vectors are less alike, resulting in a wider angle.

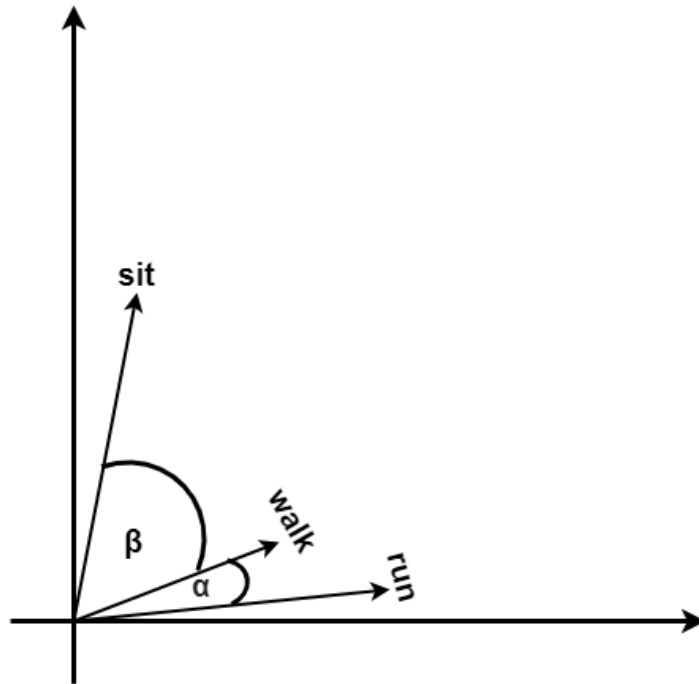


Figure 7: Word vectors of the words ”walk”, ”run” and ”sit”.

Having established a mathematical method to calculate the (semantic) similarity of given (word) vectors, the next step is to decide on which vectors to measure.

### 3.2.2 Context-Context Similarity

The ”context-context similarity” is a similarity measure based on Hoffman et al. (2013) and Sagi et al. (2009). It is also used by Naumann et al. (2018). ”context-context similarity” is calculated by taking the  $k$  most relevant (according to LMI values) context dimensions for a certain target word. Then,

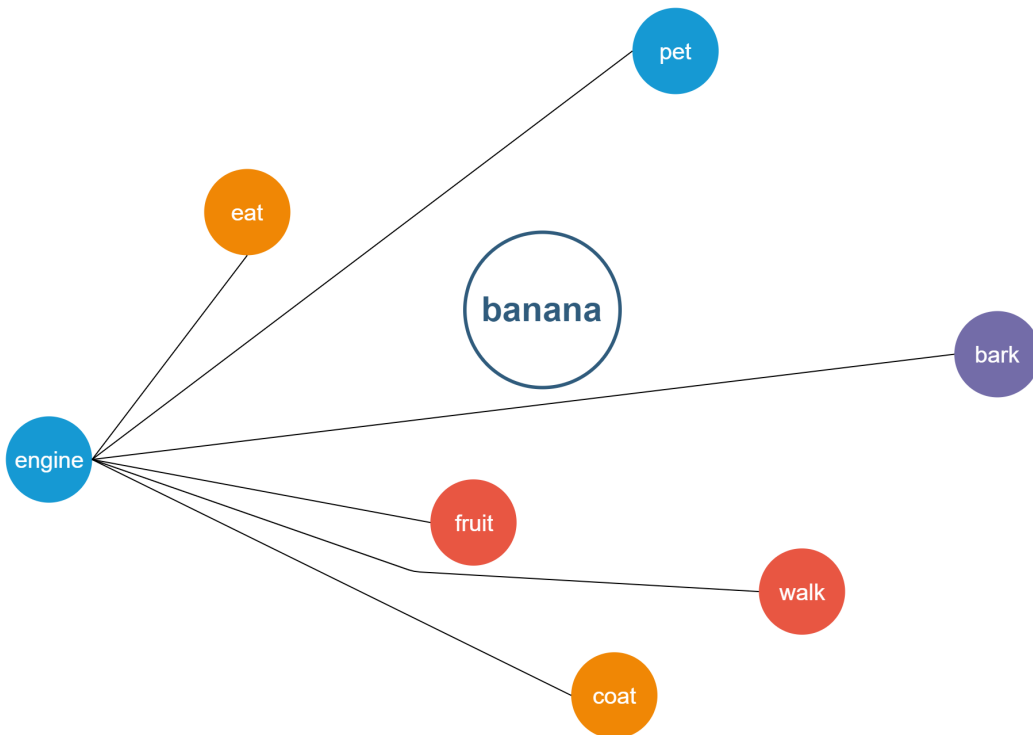


Figure 8: Target word "banana" and the seven context dimensions. Exemplary the six pairwise connections for the context word "engine" with the six other context dimensions are shown. Note that for larger models, the top  $k$  dimensions are chosen according to their LMI values.

the pairwise cosine similarity between each of the  $k$  context dimensions is calculated, i.e. for  $k=5$ , 10 ( $4 + 3 + 2 + 1$ ) different cosine similarity values. Since cosine similarity is commutative, for  $n$  context dimensions, this leads to  $\frac{(n-1) \cdot n}{2}$  context-context pairs that have to be calculated. This leaves the algorithm with  $\mathcal{O}(n^2)$  cosine similarity calculations. The average of those values then equals the average cosine similarity for the target word.

Figure 8 illustrates this for  $k=7$ . Overall, 21 pairs of cosine similarity values have to be calculated (first seven are shown by the lines) in order to get the average similarity for the word "banana" and  $k=7$ .

According to section 1, concrete words appear in a limited number of dis-

tinct contexts, whereas abstract words appear in a broader range of different contexts (Naumann et al., 2018).

The "context-context similarity" is a computational method to quantify semantic diversity, different contexts and ambiguities. Overall, target words with a broad range of contexts, i.e. abstract words are expected to show higher semantic diversity, and therefore lower "context-context similarity" scores (Hoffman et al., 2013),(Sagi et al., 2009).

### 3.2.3 Target-Context Similarity

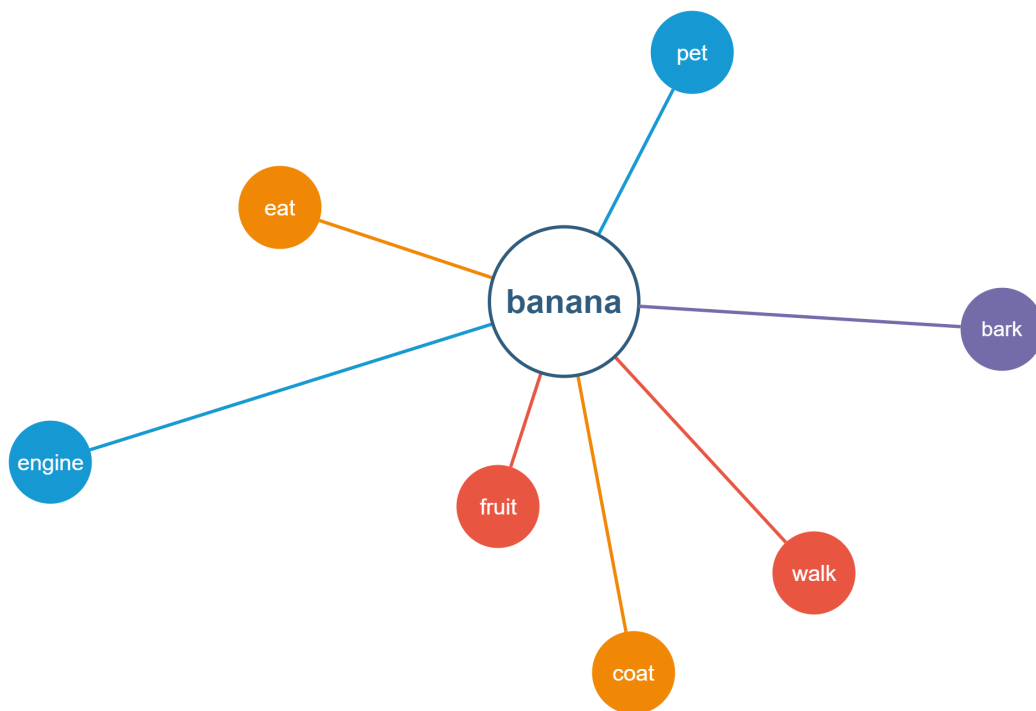


Figure 9: Target word "banana" and the seven context dimensions. The seven pairs of target word and context word are illustrated. Note that for larger models, the top k dimensions are chosen according to their LMI values.

Another possible similarity measure is the "target-context similarity". The basic idea is similar to the "context-context similarity" however, instead of calculating the pairwise similarity of the top k context dimensions,

the pairwise similarity of the target word and the context words is calculated. Calculating the "target-context similarity" gives an indication of how semantically similar the target word and its most relevant (also defined by the LMI scores) context dimensions are. Overall for  $n$  context dimensions,  $n$  pairs have to be calculated. The average of those  $n$  pairs then results in the average similarity. This leads to a complexity of  $\mathcal{O}(n)$  similarity calculations. Figure 9 illustrates the seven pairwise connections of the target word and the seven most relevant context dimensions. As will be shown in section 3.2.5 the behaviour of the "target-context similarity" regarding the abstract and concrete test sets resembles the "context-context similarity" especially using noun targets.

### 3.2.4 Target-Neighbour Similarity

"Target-Neighbour similarity" is different to the two previously mentioned in that it does not consider the context dimensions of the target word but rather its nearest neighbours (Cohen and Widdows, 2009). Given a distributional semantics model and a target word, the cosine similarities between the target word and its neighbours are calculated. The  $k$  nearest neighbours are the  $k$  neighbours with the highest similarity to the target word.

The "target-neighbour similarity" gives an indication about how similar a target word and its nearest neighbours are.

Since cosine similarity is commutative, for  $n$  target dimensions the pairwise similarity of all  $n$  target words has to be calculated. Overall this leads to  $\frac{(n-1) \cdot n}{2}$  pairs that have to be calculated, resulting in a complexity of  $\mathcal{O}(n^2)$ , similar to the "context-context similarity".

Figure 10 illustrates the target word "banana" and the three pairs it forms with the other target dimensions of the model, i.e. the pairs (banana,dog), (banana,car) and (banana,human). In order to complete the calculation three other pairs still have to be calculated. These pairs are (dog,car), (dog,human) and (car,human). So for an example using four target dimensions, overall six

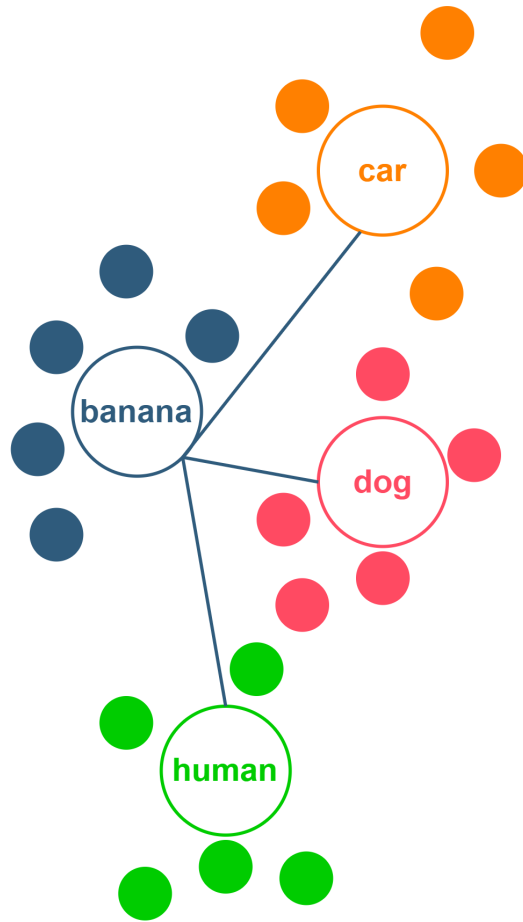


Figure 10: Target word "banana" and its three neighbours "human", "car" and "dog".

pairs of cosine similarity have to be calculated.

### 3.2.5 Results

**"Context-Context Similarity"** The results presented are split up by part of speech into nouns, verbs and adjectives. For each category, the results using both the "full dimensionality model" as well as the "brybaert model" and the "nouns only model" will be reported and discussed. All graphs are built in similar fashion. Results are reported for the top k LMI values for each target word. The chosen k values are 5, 10, 50 and 100.



Each graph shows the abstract and concrete results for either nouns, verbs or adjectives and one dimensionality model. Therefore overall nine graphs are presented. Within each graph, red boxes indicate concrete target words, blue boxes indicate abstract targets. Within each box, the dotted white line depicts the overall mean value and the black line illustrates the median value.

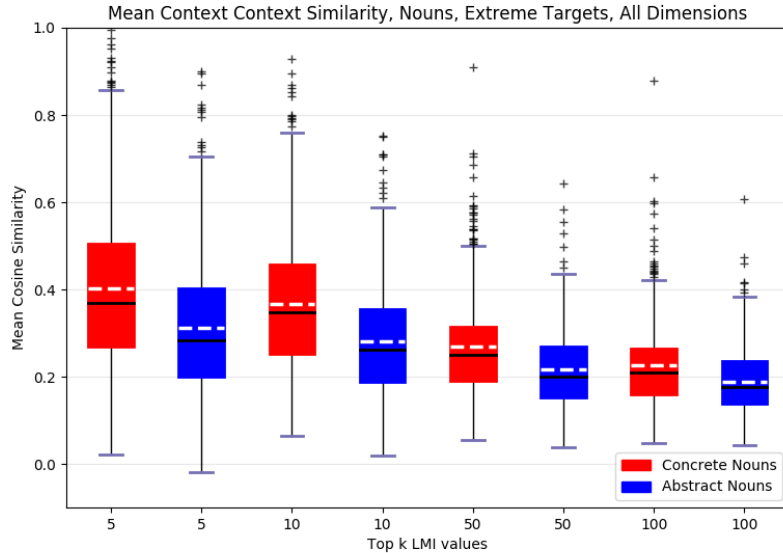


Figure 11: Mean cosine similarity for the 500 concrete and abstract target nouns using the "full dimensionality model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

**Nouns** Figures 11, 12 and 13 illustrate the mean cosine similarity using the 500 abstract nouns and the 500 concrete nouns target sets (see section 2.2). Figure 11 reports the results of the "full dimensionality model", figure 12 shows the same calculations for the "brysaert model" and figure 13 depicts the results using the "nouns only model".

The results are very similar for all three models. The "full dimensionality model" leads to slightly higher values overall however all relevant trends can be observed for the "brysaert model" in figure 12 as well as the "full dimensionality model" in figure 11 and the "nouns only model" in 13.

As expected, increasing the value of k decreases the mean and median values both for the concrete and the abstract target sets. For concrete nouns the mean value for  $k = 5$  is 0.4 for the "full dimensionality model" and 0.38

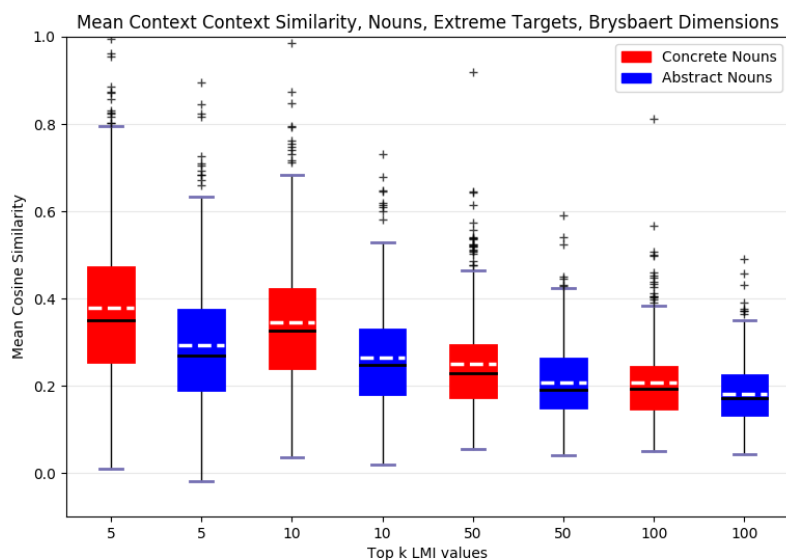


Figure 12: Mean cosine similarity for the 500 concrete and abstract target nouns using the "brysbaert model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

for both the "nouns only" as well as the "brysbaert model". For  $k=100$  these values decrease to 0.23 and 0.2 respectively. For abstract nouns the mean values decrease from 0.31 ("full dimensionality model" and "nouns only model") or 0.29 ("brysbaert model") for  $k = 5$  to 0.18 (all models) for  $k = 100$ . Furthermore, increasing  $k$  decreases the overall spread and variance of mean cosine similarity values. For  $k = 5$  very high values close to one as well as very low values are shown. Standard deviation for concrete nouns and  $k = 5$  is 0.19 ("full dimensionality"), and 0.18 ("brysbaert", "nouns only"). For  $k = 100$  the values drop to 0.09 or 0.08 respectively. For abstract nouns standard deviation drops from 0.17 for  $k = 5$  to 0.07 for  $k = 100$ . Exemplary, for  $k = 5$  the concrete target noun "dollar" achieves a very high mean cosine similarity of over 0.9. However other words like for example the abstract

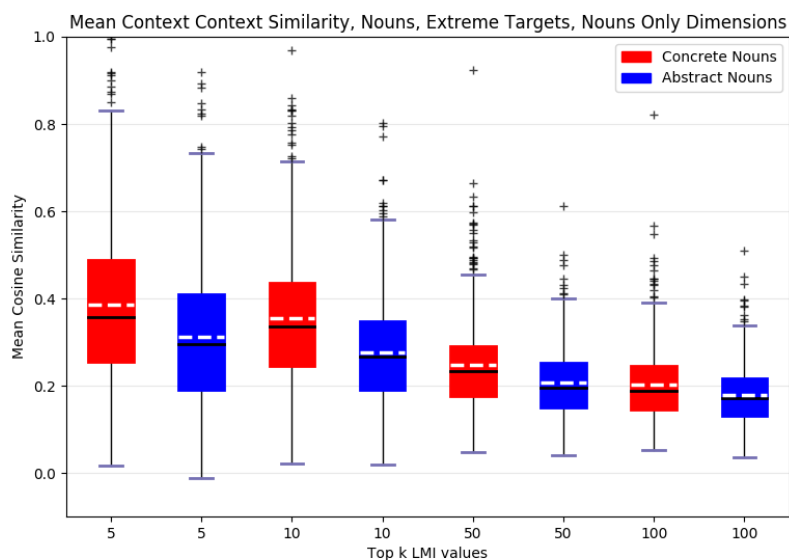


Figure 13: Mean cosine similarity for the 500 concrete and abstract target nouns using the "nouns only model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

target noun "generalization" possesses very low mean cosine similarities of under 0.05. Predictably these outliers are smoothed out for bigger k.

However, the far more interesting results can be observed when comparing the concrete and abstract results. For the same k the mean and median values are higher for the concrete nouns in comparison to the abstract nouns. Even though the gap between concrete and abstract values decreases for larger k it remains visible for all k. This fact supports the hypotheses (see section 1) that concrete nouns appear in a limited amount of contexts, therefore showing lower semantic diversity values than abstract words which appear in a broader field of contexts. These results are also in line with previous similar calculations, for example by Naumann et al. (2018)

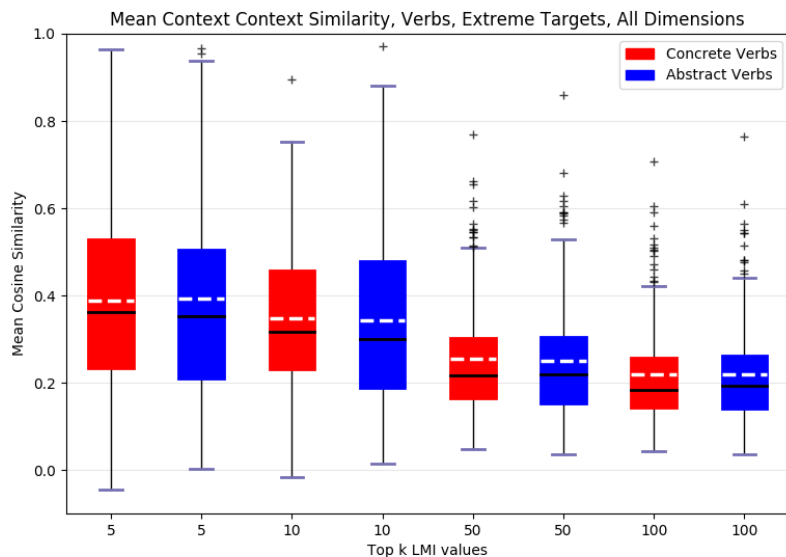


Figure 14: Mean cosine similarity for the 200 concrete and abstract target verbs using the "full dimensionality model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

**Verbs** Figure 14, figure 15 and figure 16 illustrate the mean cosine similarity values for the 200 concrete verbs respectively the 200 abstract target sets. The white dotted lines illustrate the mean values, the black lines show the median value, abstract verbs are depicted in blue, concrete verbs in red. For increasing k the values show the same effects discussed above i.e. an increase in mean values as well as variance. For concrete respectively abstract verbs and the full dimensionality model mean values decrease from 0.39 for  $k = 5$  to 0.22 for  $k = 100$ . The standard deviation decreases from 0.21 (concrete,  $k = 5$ ) or 0.23 (abstract,  $k = 5$ ) to 0.11 (concrete and abstract,  $k = 100$ ).

More importantly, the verb target sets do not show any differences for a given k between concrete and abstract verbs using either the "full dimension-

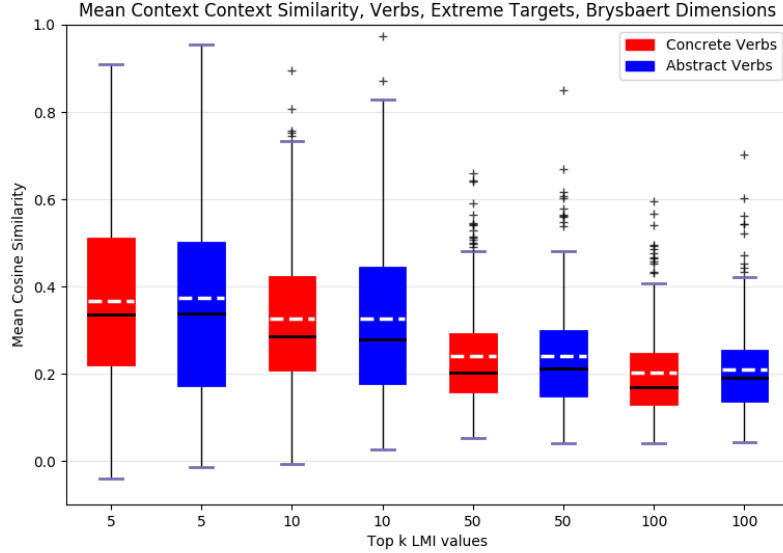


Figure 15: Mean cosine similarity for the 200 concrete and abstract target verbs using the "brysaert model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

ality" or the "brysaert model". For "full dimensionality" the mean for  $k = 5$  is exactly the same at 0.39. For "brysaert dimensions" and  $k = 5$  the mean for concrete verbs is 0.36 and for abstract words it is 0.37. This questions the underlying hypotheses (see section 1) since abstract and concrete verbs do not seem to show a difference regarding the contexts they are showcasing. These findings also concur with Naumann et al. (2018)

However using the "nouns only model" a gap between concrete and abstract verbs can be observed similarly to the nouns. For  $k = 5$  the concrete mean equals 0.41, for abstract it is 0.35. These "nouns only model" results contrary to the results for the "full dimensionality" and the "brysaert model" support the underlying hypotheses (see section 1).

The difference between the "nouns only model" and the other models could

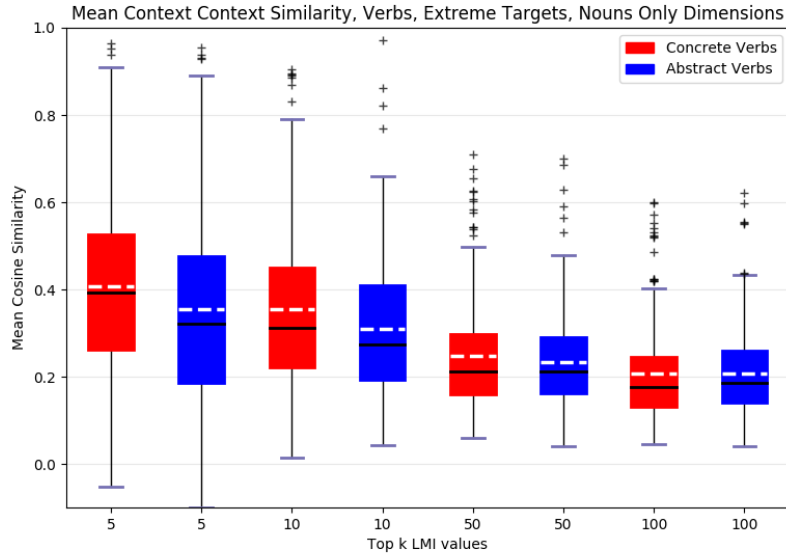


Figure 16: Mean cosine similarity for the 200 concrete and abstract target verbs using the "nouns only model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

indicate that restricting the dimensions to contain only nouns increases the overall information contained in the model regarding the abstractness and concreteness of target words i.e. co-occurring nouns holding the most information with regard to the abstractness and concreteness of words.

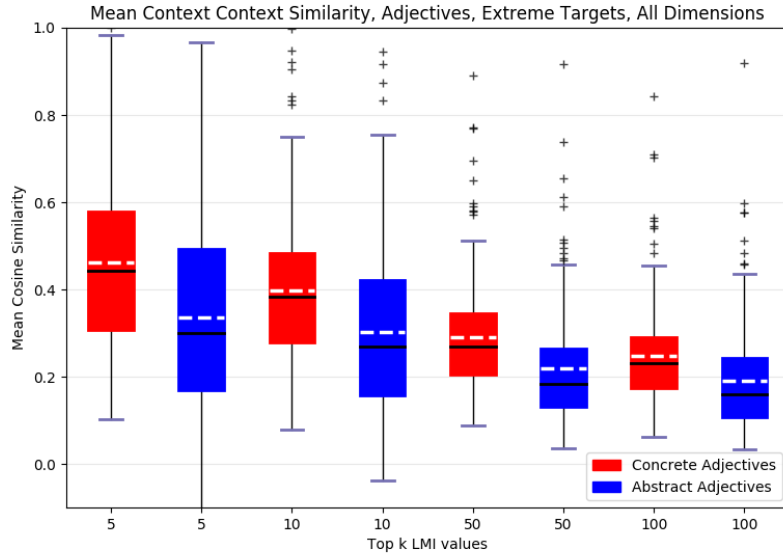


Figure 17: Mean cosine similarity for the 200 concrete and abstract target adjectives using the "full dimensionality model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

**Adjectives** Similar to the nouns and verbs, figure 17 reports the mean cosine similarity for the "full dimensionality model", figure 18 reports the same data for the "brysbaert model" and figure 19 shows the results for the "full dimensionality model". Again, results are reported for  $k = 5, 10, 50$  and  $100$ . Red boxes report the data for the 200 concrete adjectives, blue boxes correspond to the 200 abstract adjectives. White dotted lines in the boxes illustrate the mean values, the black lines show the median value.

Similar to the previously described results for the nouns and verbs, the data for adjectives shows the same effect of a decrease in variance and mean value with an increase of  $k$ . For the "full dimensionality model" mean values decrease from 0.43 (concrete,  $k = 5$ ) or 0.34 (abstract,  $k = 5$ ) to 0.23 (concrete,  $k = 100$ ) respectively 0.17 (abstract,  $k = 100$ ). Standard deviation decreases



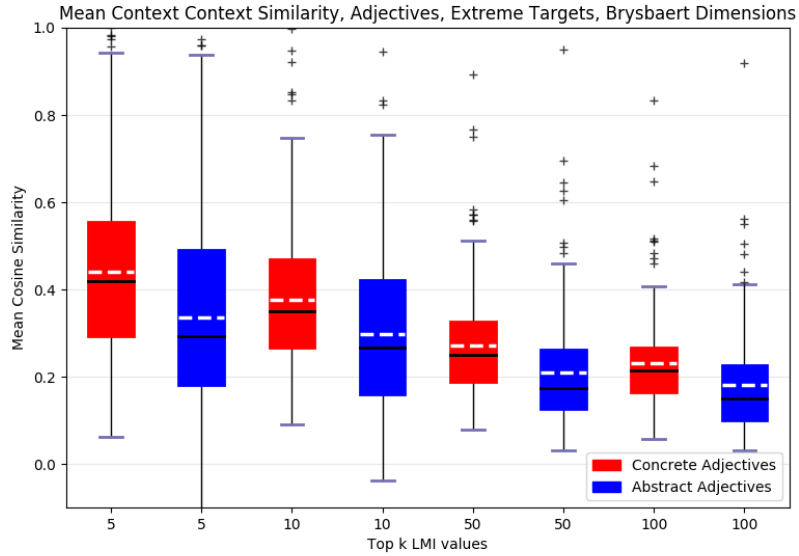


Figure 18: Mean cosine similarity for the 200 concrete and abstract target adjective using the "brysbaert model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

from 0.21 ( $k = 5$ ) to 0.11 ( $k = 100$ ).

Also, again in line with Naumann et al. (2018), the concrete adjectives show significantly higher mean cosine similarity values when compared to the abstract counterparts. This, in line with the results for nouns but contrary to the results for verbs, holds true for all three different dimensionality models. These findings support the underlying hypothesis (see section 1) as again, concrete adjectives seem to appear in a more limited set of contexts compared to abstract adjectives.

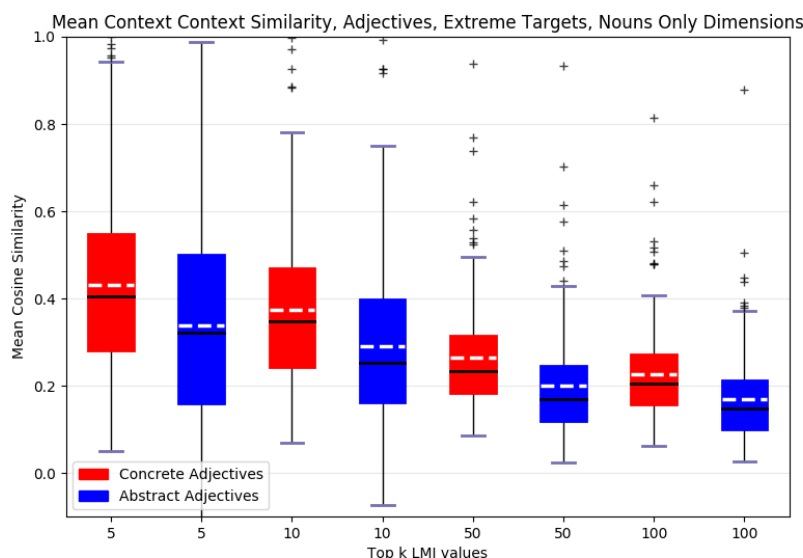


Figure 19: Mean cosine similarity for the 200 concrete and abstract target adjective using the "nouns only model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

**Target-Context Similarity** Figures 20-28 illustrate the "target-context similarities" by reporting the mean cosine similarity values for both the concrete and abstract noun, verb and adjective target sets. Both the results for the "brysbaert model", the "nouns only model" as well as the results for the "full dimensionality model" are reported. The layout for all the graphs is similar. The mean cosine similarity values for concrete targets and abstract targets are shown next to each other for one value of k. Overall the mean cosine similarity values are calculated for  $k = 5, 10, 50$  and  $100$ . Blue boxes indicate an abstract target set, red boxes indicate the concrete targets. Each box contains a black line, indicating the median value as well as a dotted white line indicating the mean value. For all parts of speeches, the calculated results for the "full dimensionality model", "nouns only model" and

the "brybaert model" are very similar, therefore the results discussed in the corresponding sections, unless differently indicated, always refer to all three models.

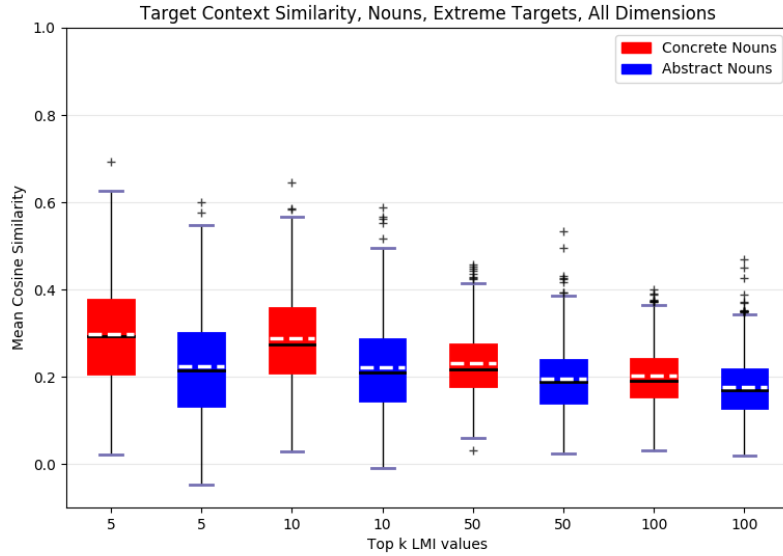


Figure 20: Mean cosine similarity for the 500 concrete and abstract target nouns using the "full dimensionality model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

**Nouns** Figure 20 depicts the mean cosine similarity for the 500 abstract and concrete target nouns calculated using the "full dimensionality model". Figure 21 reports the same values using the "brysbaert model" while figure 22 illustrates the results for the "nouns only model". Increasing the number of context dimensions considered for the calculation, i.e. increasing k overall has the same effect as for the "context-context similarity". Increasing k lowers the mean mean cosine values as well as the variance of the results. For the "full dimensions model" the mean value decreases from 0.30 (concrete, k = 5) to 0.20 (concrete, k = 5) respectively and from 0.22 (abstract, k = 5) to 0.18 (abstract, k = 100). Standard deviation decreases from 0.12 for k = 5 to 0.07 for k = 100 both for concrete and abstract targets. However, again, this is an expected behaviour since increasing the number of context dimensions overall

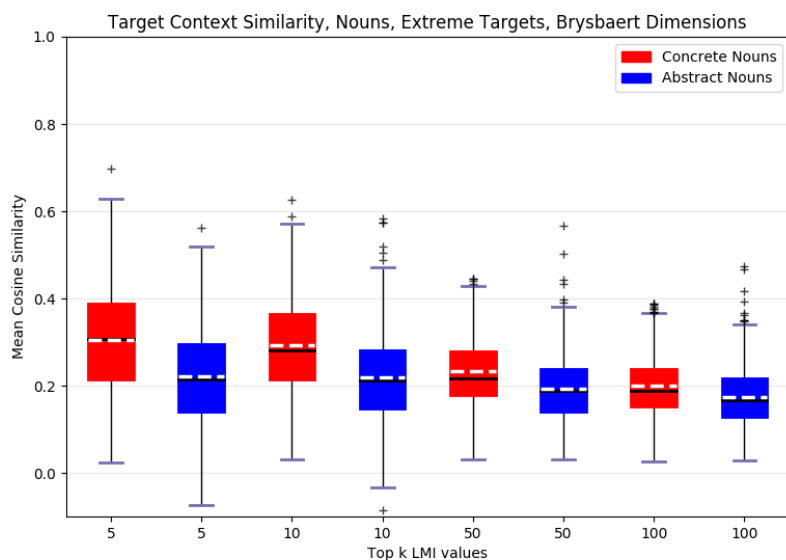


Figure 21: Mean cosine similarity for the 500 concrete and abstract target nouns using the "brysbaert model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

smooths the outliers which can heavily influence the values for smaller k. Some outlier values for example include the concrete noun "broccoli" which shares very high cosine similarity values with context words like "cauliflower" or "cabbage". For  $k = 5$  this leads to a mean cosine similarity of over 0.6. On the other hand the concrete noun "rubber" has very small cosine similarity values with context dimensions like "band" or "stamp". Overall the mean cosine similarity of "rubber" is lower than 0.1.

Of bigger interest however, again, is the discrepancy between the concrete and abstract nouns for a fixed k. The fact that the values for concrete nouns are higher means that overall, concrete nouns are more similar to their relevant context dimensions in comparison to abstract nouns, which are less similar to the relevant context dimensions. This could indicate, that concrete nouns

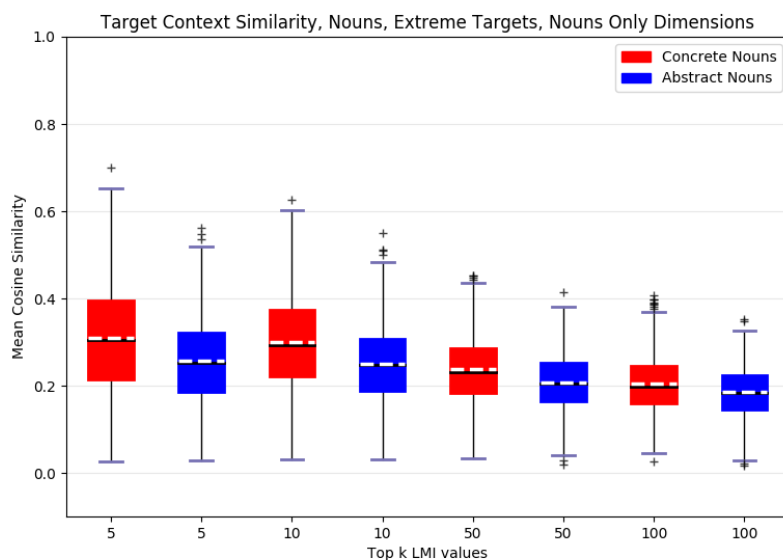


Figure 22: Mean cosine similarity for the 500 concrete and abstract target nouns using the "nouns only model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

appear within very distinct context dimensions which are semantically similar to the actual target words. Abstract nouns on the other hand seem to appear with certain context dimensions less similar to the actual target. This could indicate that the context of abstract nouns is less distinct.

Using this reasoning, the results, again, support the underlying hypothesis.

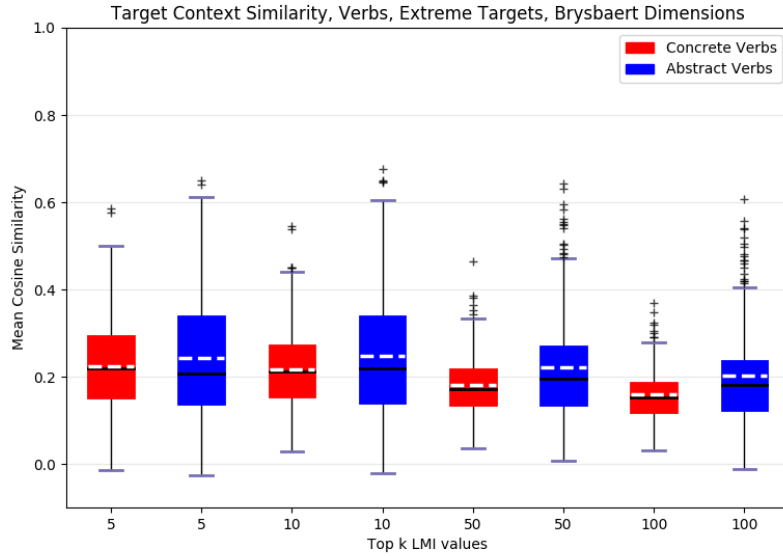


Figure 23: Mean cosine similarity for the 200 concrete and abstract target verbs using the "full dimensionality model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

**Verbs** Figures 23-25 depict the mean cosine similarity for the 200 abstract and concrete target verbs using the "full dimensionality", "brysbaert" and "nouns only model" respectively.

Similar to the nouns, increasing k expectedly smoothes out outlying values, decreasing the variance. Standard deviation for the "full dimensionality model" decreases from 0.11 ( $k = 5$ ) to 0.06 ( $k = 100$ ) for concrete verbs and from 0.14 ( $k = 5$ ) to 0.11 ( $k = 100$ ) for abstract verbs. Furthermore, as expected, mean values decrease from 0.23 to 0.16 (concrete) or 0.24 to 0.20 (abstract). Overall abstract verbs show a larger range of results in comparison to concrete verbs.

Similar to the "context-context similarity" values using the "full dimen-

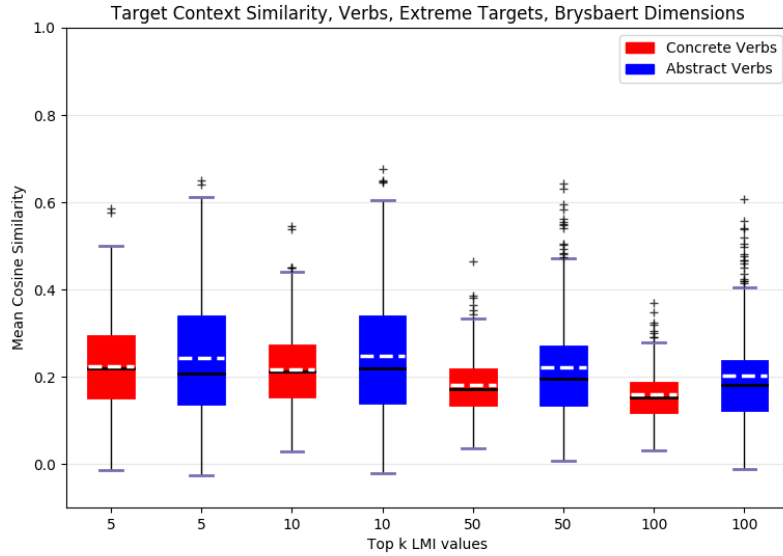


Figure 24: Mean cosine similarity for the 200 concrete and abstract target verbs using the "brysbaert model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

sionality" or the "brysbaert model" for verbs, there is no gap between abstract and concrete verbs for a given k. Furthermore, contrary to the "context-context similarity", no difference between abstract and concrete verbs can be observed using the "nouns only model".

These results across all models indicate the possibility of concrete and abstract verbs not differing in the number of contexts they are used in, since both concrete and abstract target words are equally similar to the relevant context dimensions.



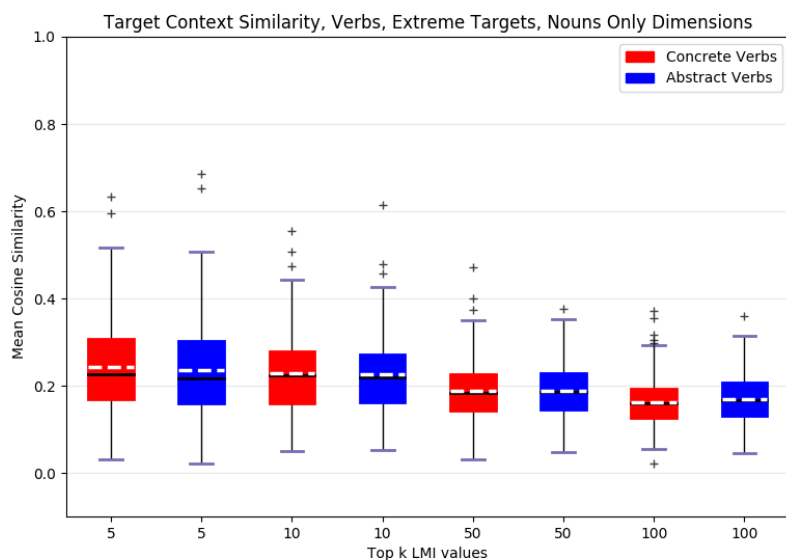


Figure 25: Mean cosine similarity for the 200 concrete and abstract target verbs using the "nouns only model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

**Adjectives** The mean cosine similarities for the 200 concrete and abstract adjectives are shown in figure 26 ("full dimensionality model"), figure 27 ("brysaert model") and figure 28 ("nouns only model"). Similar to all other results, increasing the number of relevant context dimensions, i.e. increasing  $k$  leads to a decrease in variance and mean mean cosine similarity. The mean decreases from 0.23 (concrete,  $k = 5$ ) to 0.18 (concrete,  $k = 100$ ) or from 0.24 (abstract,  $k = 5$ ) to 0.20 (abstract,  $k = 100$ ). Standard deviation decreases from 0.11 to 0.06 for concrete adjectives and from 0.14 to 0.11 for abstract adjectives. As for verbs, abstract adjectives show a bigger variance than the concrete adjectives.

Comparing abstract and concrete mean cosine similarity values for a fixed  $k$ , no differences can be seen for either the "brysaert" or the "full dimensions"

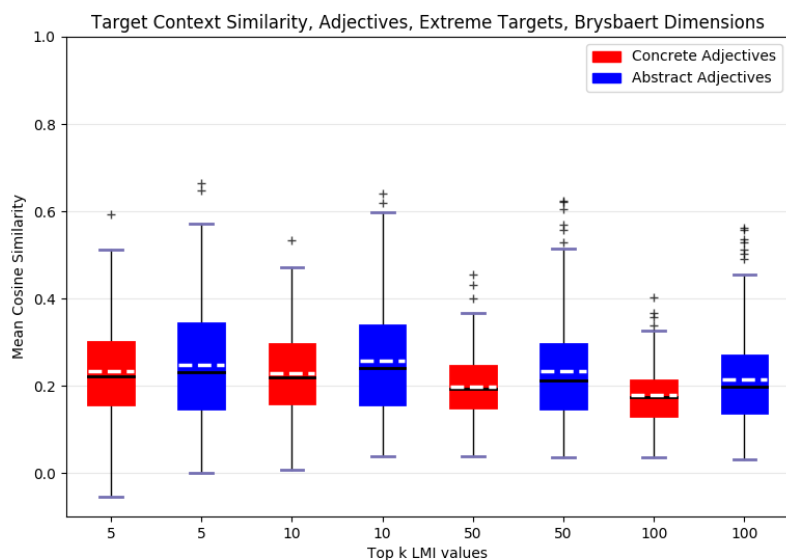


Figure 26: Mean cosine similarity for the 200 concrete and abstract target adjectives using the "full dimensionality model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

dimensionality models. This could hint at the possibility that abstract and concrete target adjectives do not differ in the number of used contexts. However, similar to the "context-context Similarity" using the "nouns only model" a gap between abstract and concrete adjectives can be seen. For  $k = 5$ , the mean value of concrete adjectives equals 0.24, the mean value of abstract adjectives is 0.20. This could further strengthen the possibility mentioned before i.e. reducing the context dimensions of the model to nouns only improves the amount of information with regard to the abstractness and concreteness of target words.

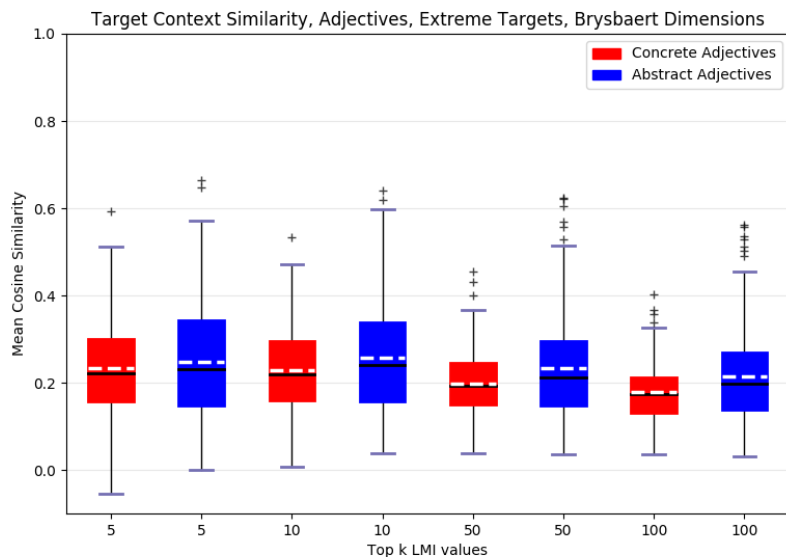


Figure 27: Mean cosine similarity for the 200 concrete and abstract target adjectives using the "brysbaert model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

**Target-Neighbour Similarity** This section illustrates and discusses the mean cosine similarity for the target words and their top k neighbours. In this case, top k neighbours refers to the k neighbours that have the highest cosine similarity with the given target word, i.e. if  $CosineSimilarity(dog, human) = 0.7$  and  $CosineSimilarity(dog, banana) = 0.2$  then "human" is a closer neighbour to "dog" than "banana" is to "dog". Similar to the previous results, each graph illustrates the results for one dimensionality model, one part of speech and k values of 5, 10, 50 and 100. Concrete words are depicted in red, abstract words are depicted in blue. In each box, a white dotted line indicates the mean value for that specific result, and the black line shows the median value.

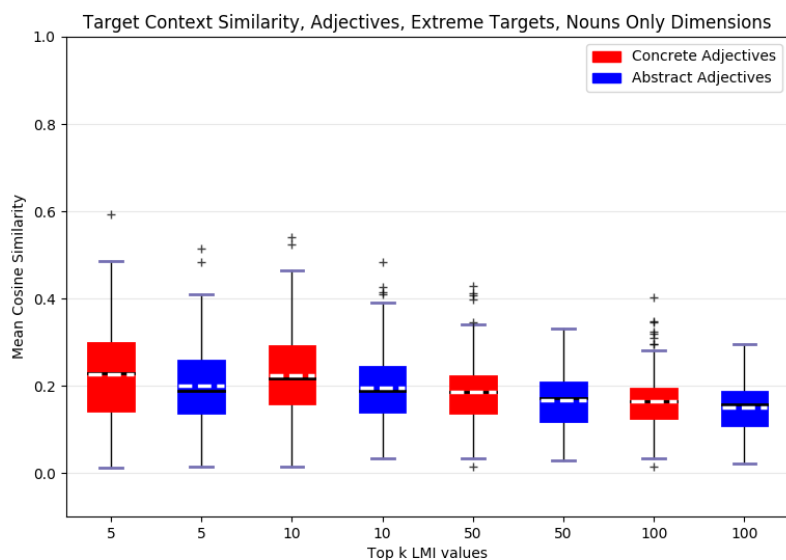


Figure 28: Mean cosine similarity for the 200 concrete and abstract target adjectives using the "nouns only model". The average is calculated using the top 5, 10, 50 and 100 most relevant context dimensions according to the LMI values. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

**Nouns** Figure 29 shows the results for "target-neighbour similarity" calculated using the "full dimensionality model". Figures 30 and 31 show the same results for the "brysbart" and the "nouns only model". Overall the results for all three dimensionality models are rather likewise. Therefore, all descriptions, unless indicated differently, refer to all three models. Overall the mean cosine similarity naturally drops for increased k. This is expected behaviour, since increasing k means including neighbours with a lower cosine similarity value to the target word. Furthermore, also expectedly, the range of the results decreases for larger k. Looking at the "full dimensionality model", for  $k = 5$  the range reaches from mean cosine similarity values of over 0.8, for example for the target word "reason" to mean values of under 0.15 ("impurity"). This range significantly decreases for  $k = 100$ . Overall

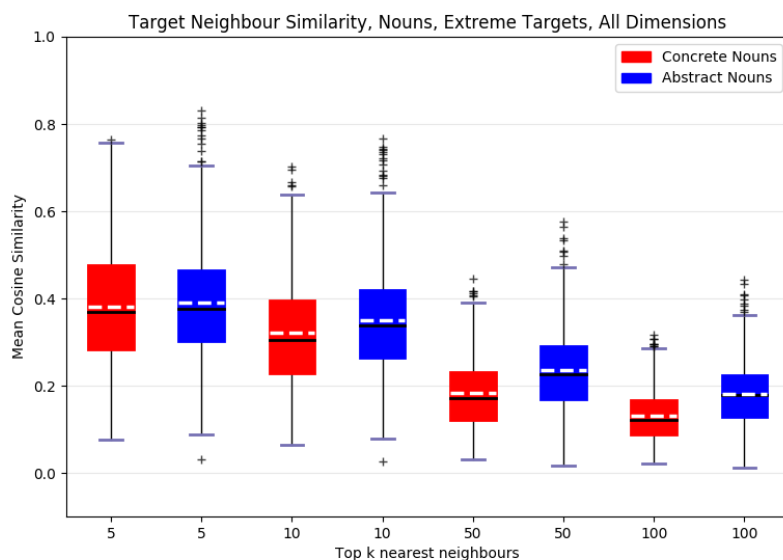


Figure 29: Mean cosine similarity for the 500 concrete and abstract target nouns using the "full dimensionality model". The average is calculated using the top 5, 10, 50 and 100 nearest neighbours. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

for concrete nouns ("full dimensionality model") the mean value is 0.38 and the standard deviation is 0.14. These values change to 0.13 (mean) and 0.06 (standard deviation). For abstract nouns, the mean decreases from 0.39 ( $k = 5$ ) to 0.18 ( $k = 100$ ) and the standard deviation decreases from 0.14 for  $k = 5$  to 0.07 for  $k = 100$ .

Comparing the results of concrete and abstract nouns for a fixed  $k$ , for smaller  $k$  the results are rather similar. However for larger  $k$ , the mean for abstract words is higher than the mean for concrete words, forming a gap between concrete and abstract words.

Regarding the hypothesis (see section 1) on the one hand it might be expected that concrete words have nearest neighbours with higher cosine similarity values to each other than abstract words have since these nouns appear only in very distinct contexts and therefore should be very similar to some of their

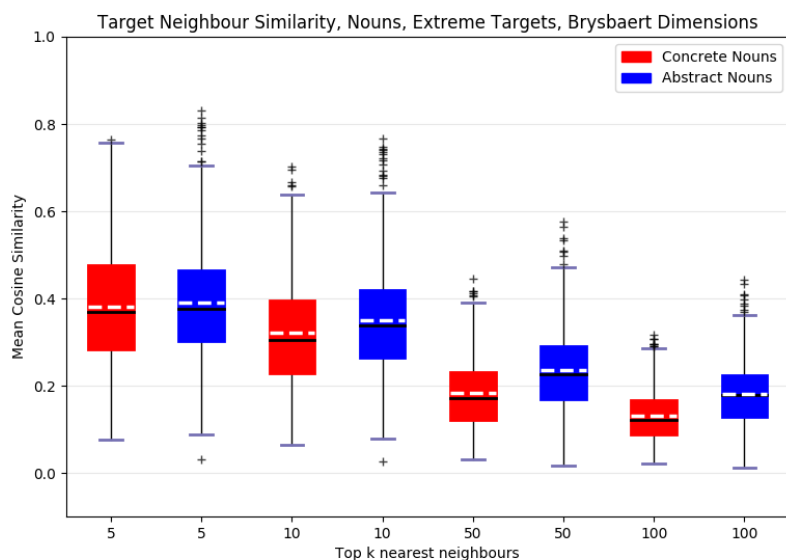


Figure 30: Mean cosine similarity for the 500 concrete and abstract target nouns using the "brysbaert model". The average is calculated using the top 5, 10, 50 and 100 nearest neighbours. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

neighbours. This behaviour is not supported by the data. However, on the other hand, for larger k, it could be argued that abstract words should have higher average similarity to their neighbours than concrete words, since they appear in a broad spectrum of contexts, therefore showing some similarity to a higher number of neighbours. Contrary to the previous two similarity measures, the "target-neighbour similarity" only takes into consideration target vectors. The number of target dimensions is significantly lower than the number of context dimensions. This means that for larger k the ratio of k to dimensions is higher for this similarity. Therefore this effect is only likely to appear here.

This is supported by the fact that for larger k, a gap between abstract and concrete words begins to form in the data, even while overall the deviation and outliers of the data are smoothed out for larger k. This could imply that

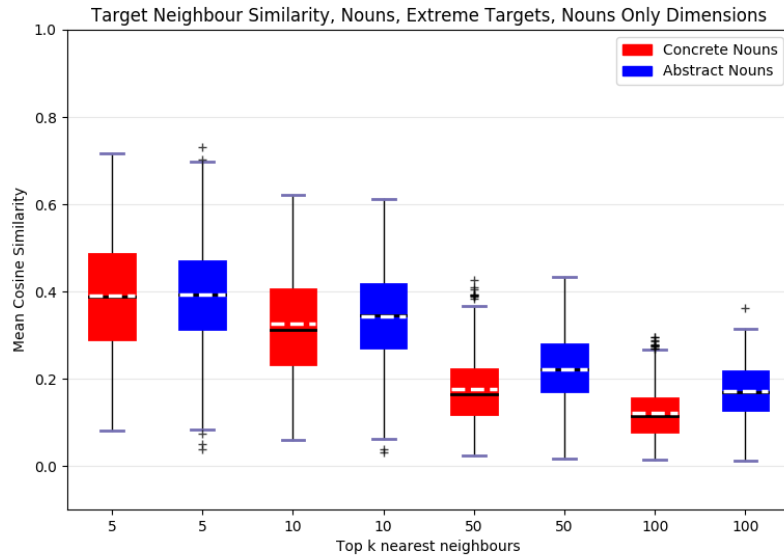


Figure 31: Mean cosine similarity for the 500 concrete and abstract target nouns using the "nouns only model". The average is calculated using the top 5, 10, 50 and 100 nearest neighbours. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

this is an actual effect caused by the hypotheses mentioned above.

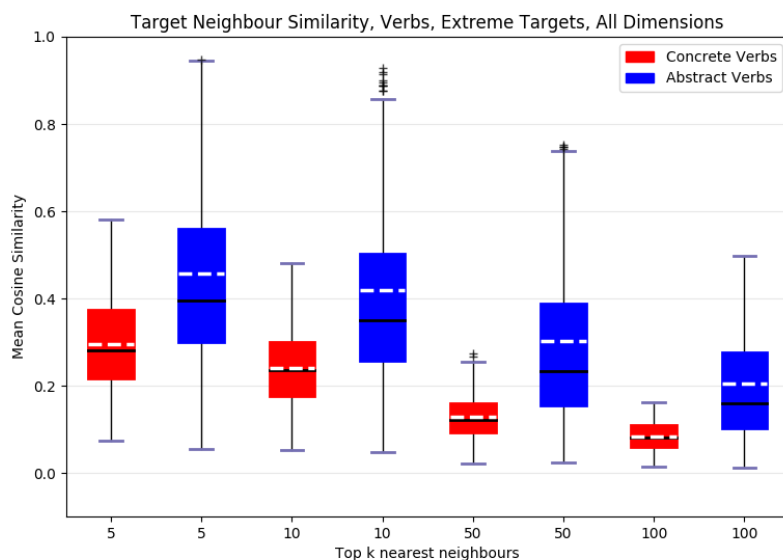


Figure 32: Mean cosine similarity for the 200 concrete and abstract target verbs using the "full dimensionality model". The average is calculated using the top 5, 10, 50 and 100 nearest neighbours. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

**Verbs** Figures 32-34 show the results calculating the "target-neighbour similarity" using the "full dimensionality", "brysbart" and "nouns only model". Again, the results for all three models are very similar and therefore the discussion, unless stated differently, refers to all models.

Again, predictably, mean values and similarity range decrease for larger k. For concrete verbs, "full dimensionality model", standard deviation decreases from 0.10 to 0.07 for abstract verbs it decreases from 0.22 to 0.14 for  $k = 5$  and  $k = 100$ . Mean values decrease from 0.30 to 0.09 (concrete) respectively and from 0.46 to 0.20.

Comparing concrete and abstract words, abstract words show an overall higher similarity to their nearest neighbours compared to concrete words, at least for smaller k, contradicting the hypotheses of concrete words appearing in a limited, specialized context in comparison to abstract words



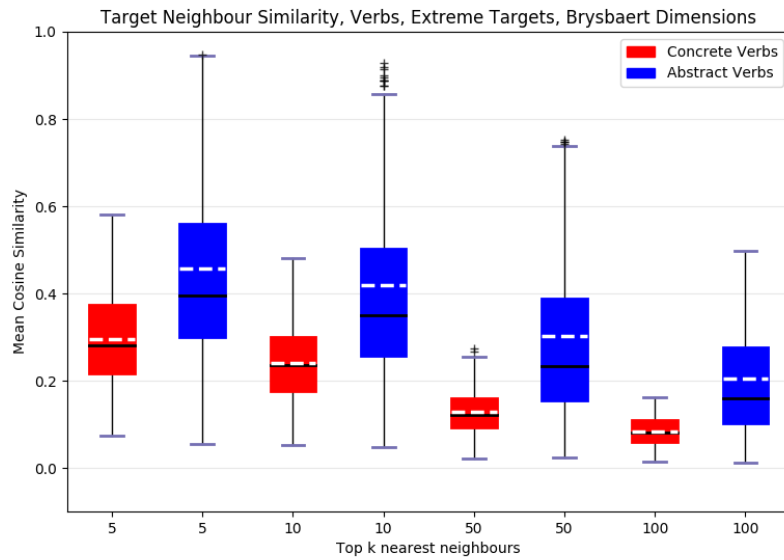


Figure 33: Mean cosine similarity for the 200 concrete and abstract target verbs using the "brysbaert model". The average is calculated using the top 5, 10, 50 and 100 nearest neighbours. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

appearing in a much broader context.

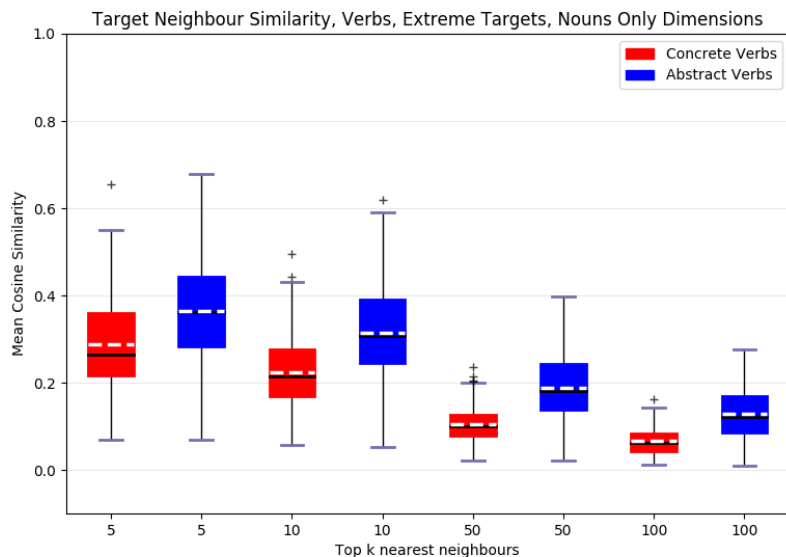


Figure 34: Mean cosine similarity for the 200 concrete and abstract target verbs using the "nouns only model". The average is calculated using the top 5, 10, 50 and 100 nearest neighbours. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

**Adjectives** The results for the concrete and abstract adjectives regarding the "target-neighbour similarity" are shown in figure 35 ("full dimensionality model"), figure 36 ("brysbaert model") and figure 37 ("nouns only model"). Similar to the results for nouns and verbs, the different dimensionality models produce rather likewise results and are thus discussed together. Also in line with all previous results increasing  $k$  streamlines the results, decreasing the overall mean values and the variance. Concrete adjectives using the full dimensionality model have a mean of 0.31 and a standard deviation of 0.14 for  $k = 5$ . For  $k = 100$ , the mean decreases to 0.08, the standard deviation to 0.04. Likewise, increasing  $k$  decreases the mean value for abstract words from 0.49 ( $k = 5$ ) to 0.29 ( $k = 100$ ) and the standard deviation changes from 0.24 ( $k = 5$ ) to 0.20 ( $k = 100$ ).

Again, comparing concrete and abstract adjective results for a fixed  $k$  shows

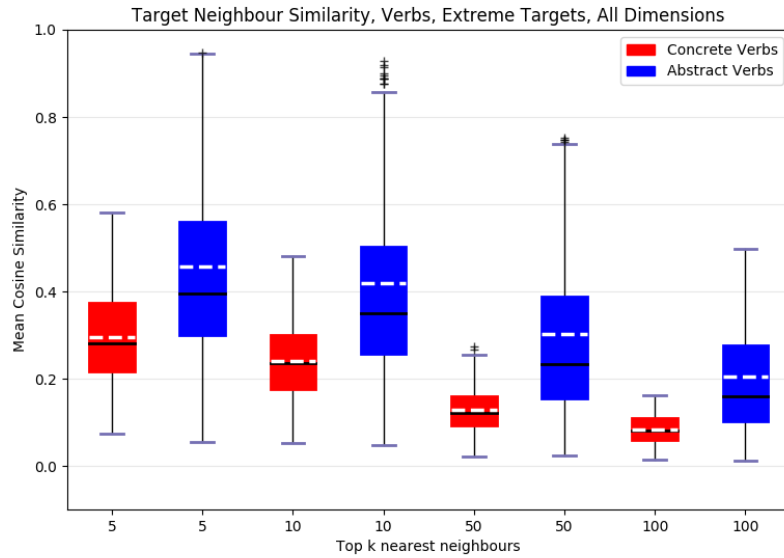


Figure 35: Mean cosine similarity for the 200 concrete and abstract target adjectives using the "full dimensionality model". The average is calculated using the top 5, 10, 50 and 100 nearest neighbours. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

that abstract verbs seem to be more similar to their nearest neighbours in comparison to concrete verbs, thus not supporting the underlying hypothesis (see section 1).

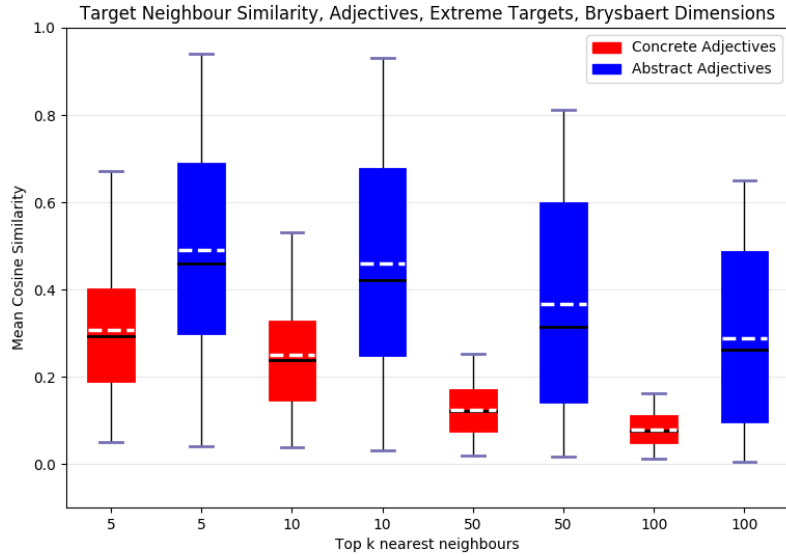


Figure 36: Mean cosine similarity for the 200 concrete and abstract target adjectives using the "brysbaert model". The average is calculated using the top 5, 10, 50 and 100 nearest neighbours. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

**Combination** The previous sections thoroughly presented the results for three different similarity models in showing and discussing the results for the "context-context similarity", the "target-context similarity" as well as the "target-neighbour similarity" across three different dimensionality models and for nouns, verbs and adjectives.

This section tries to combine the results and possible indications of the three different similarity measures by integrating the various results into two-dimensional scatterplots. By combining the results and illustrating them in a combined fashion, possibly further trends and indications in the data can be observed.

In order to improve the visibility of possible patterns and clarity overall, only combined plots using the target noun sets and the "full dimensionality model" are discussed. This choice was made since overall nouns seem to show

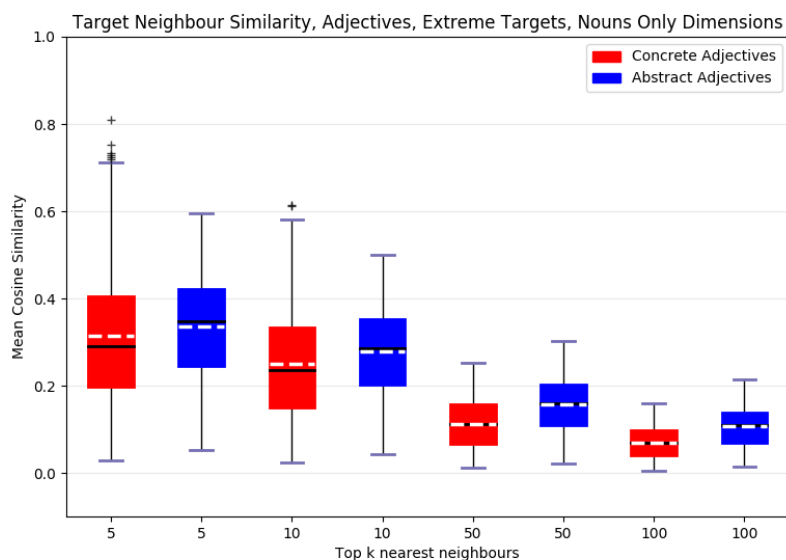


Figure 37: Mean cosine similarity for the 200 concrete and abstract target adjectives using the "nouns only model". The average is calculated using the top 5, 10, 50 and 100 nearest neighbours. Values are depicted in red and blue for the concrete target set and the abstract target set respectively.

the clearest trends and results, and overall, results for different dimensionality models are very similar, especially for noun targets. In order to account for outliers and data variance, the presented results are limited to  $k = 100$ , unless specified differently.

**Context-Context and Target-Context** The combination of "context-context similarity" and "target-context similarity" is shown in figure 38. The x-axis indicates the "context-context similarity", the y-axis corresponds to the "target-context similarity". Concrete target nouns are depicted in red, abstract targets are illustrated in blue. The red and blue line are regression lines for the concrete, respectively abstract data points. At first glance there does not seem to be a huge visible distinction between abstract and concrete words. This first impression is supported using the regression analysis.

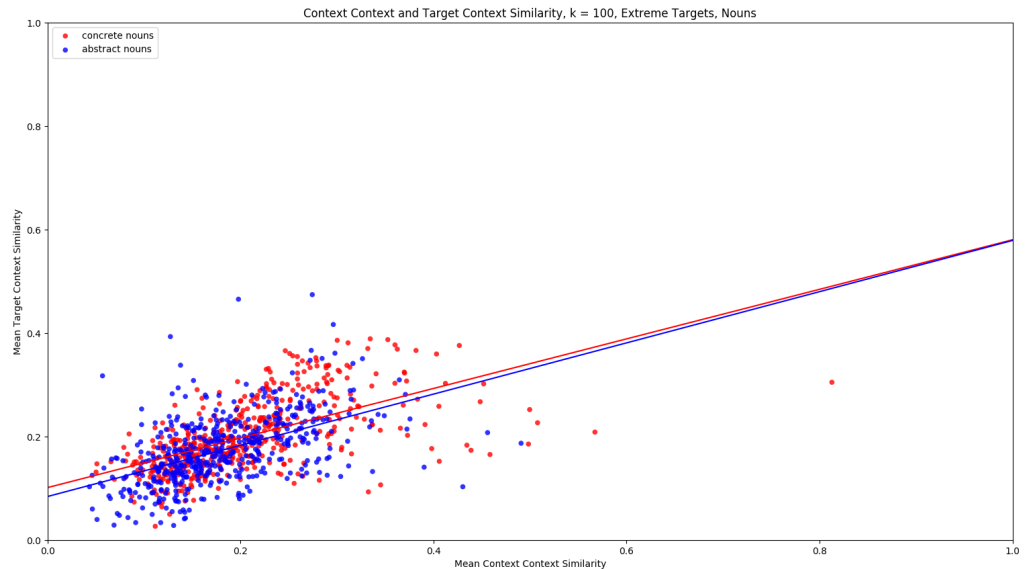


Figure 38: Combination of the "context-context similarity" on the x-axis and the "target-context similarity" on the y-axis. Blue dots indicate abstract words, red words indicate concrete words. The plot includes regression lines for both word classes.

Considering the disparity between concrete and abstract words, there is no significant difference regarding the coefficient values between concrete and abstract words (adjusted  $R^2 = 0.3298$ ).

Looking at the data points most of them seem to behave in a very linear fashion i.e. they closely mirror the regression lines. Values with higher "context-context similarity" also have higher "target-context similarity" values. This could make sense, since target words that have a highly similar context could arguably also be expected to be more similar to this context as well, when compared to words with a lower similarity within their context. This reasoning could be made both for abstract as well as concrete words, which is also supported by the regression analysis showing no significant dif-

ference in the coefficient value for abstract and concrete words.

However, looking at figure 38 shows a number of data points that do not fit the aforementioned hypotheses, in that they appear to have a high "context-context similarity" but a low "target-context similarity", i.e. the context of these targets is very similar to itself, however the target word itself is not very similar to this context.

Hoping to further strengthen this effect it was decided to have a closer look at the same data using a k value of 5 instead of 100, since decreasing the k value has been shown to increase the variance and number of outliers in the previous sections. Figure 39 illustrates the same target nouns for said value of  $k = 5$ .

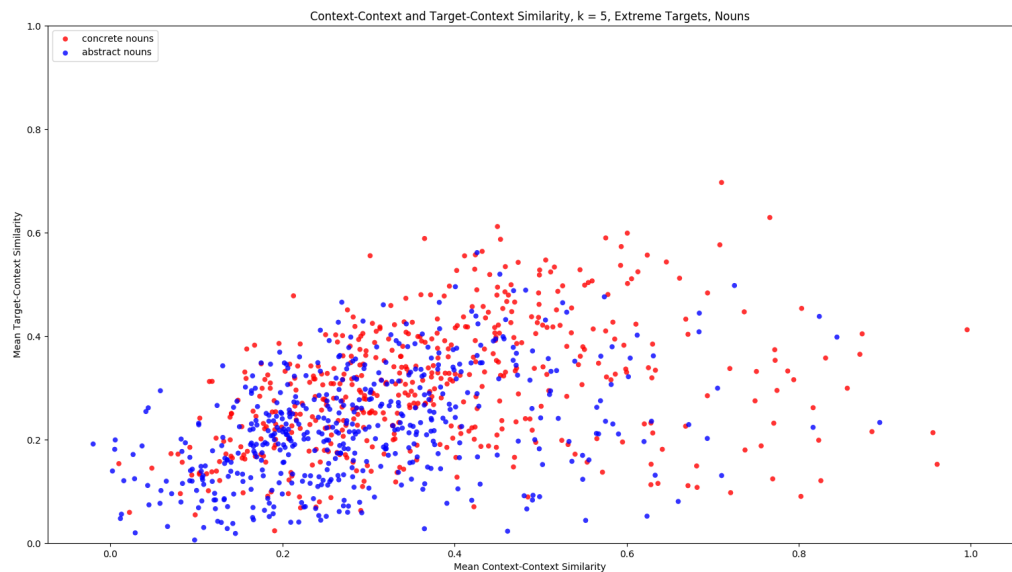


Figure 39: Combination of the "context-context similarity" on the x-axis and the "target-context similarity" on the y-axis. Blue dots indicate abstract words, red words indicate concrete words. K values are set to five.

Decreasing the k value visibly strengthens the effect as expected with figure 39 depicting a lot of datapoints, both abstract and concrete, in the bottom right quadrant, i.e. data points that have a higher "context-context similarity" coupled with a lower "target-context similarity". Going one step further, it was decided to have a closer look at the actual target words that appear in that quadrant, their most relevant context dimensions as well as their overall frequency. Figure 40 shows the labeled data for all words with a "context-context similarity" of at least 0.6 and at least twice as high as the "target-context similarity". The frequency of the annotated words is coded by color. For concrete words, words with lower frequency (compared to the entire target list) are depicted in orange, high frequency words are depicted in green. Abstract words with low frequency are shown in a lighter blue, darker blue indicates an increase in frequency.

Regarding the frequency, there does not seem to be a correlation. Both for concrete and for abstract words both highly frequent words (for example the word "student" on the top right is the most frequent word of the entire list) as well as words with lower frequency, like "luck", "fraud" or "marijuana" seem to show the effect.

Considering the context dimensions of the words however there seems to be one group of words that predominantly shows the effect. This group of words appears with context dimensions that are common collocations for the target word like for example the target word "luck" appears with the context dimensions "good", "bad" and "have", the target word "effect" appears with the context dimensions "side" and "negative". Overall roughly 20% of the labeled words appear to include at least one common "collocation context dimension".





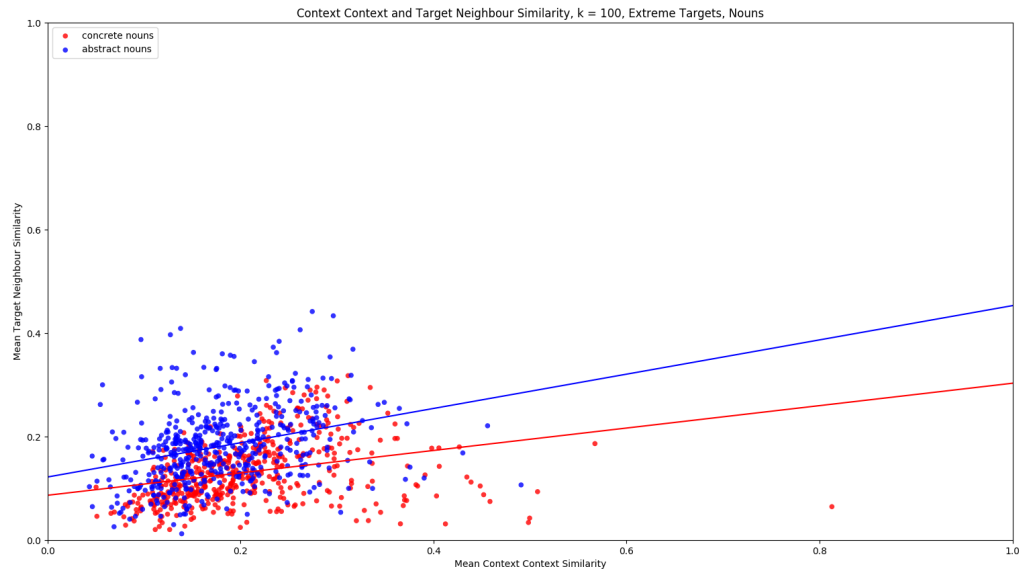


Figure 41: Combination of the "context-context similarity" on the x-axis and the "target-neighbour similarity" on the y-axis. Blue dots indicate abstract words, red words indicate concrete words. The plot includes regression lines for both word classes.

**Context-Context and Target-Neighbour** Figure 41 illustrates the combined results for the "context-context similarity" and the "target-neighbour similarity". The x-axis illustrates the "context-context similarity" for the target nouns, the y-axis corresponds to the "target-neighbour similarity". Again, concrete target nouns are depicted as red dots, abstract target nouns as blue dots. Shown in red is the regression line for all concrete data points, the blue line is the regression line for all abstract points.

Looking at the graph, there appears to be a distinction between abstract and concrete target nouns. Abstract nouns seem to overall have a higher "target-neighbour similarity" for a given "context-context similarity" in comparison to concrete nouns, i.e. the blue dots in figure 41 are higher on the y-axis for a given x-axis section. This is supported by the regression analysis. Regarding

the y-axis intercept, the analysis shows a significant difference for abstract and concrete nouns (p-value  $\approx 0$ ). This is also the case for the coefficient value (p-value 0.034, adjusted  $R^2 = 0.204$ ).

This could indicate that abstract words, while having equally similar context dimensions, are more similar to their neighbours.

This could support the underlying hypotheses in section 1, i.e. abstract words appear in a broader context, therefore they are more likely to overlap with neighbours, increasing their similarity to the nearest neighbours in comparison to concrete nouns. Furthermore looking at the context dimensions it appears that most concrete nouns appear with very concrete neighbours as their nearest neighbours as well. For example the five nearest neighbours of the concrete target noun "tv" (concreteness score 5.0) all also have a concreteness rating in the range of 4.9 to 5. Amongst these neighbours are the neighbours "sofa", "popcorn" and "theater". Abstract target nouns on the other hand appear to also have very abstract nearest neighbours. The five nearest neighbours of the abstract target noun "generosity" (concreteness score 1.84) all are within a concreteness range of 1.38 to 1.79, e.g. "goodness" or "humility".

**Target-Context and Target-Neighbour** The results shown in figure 42 illustrate the combination of the "target-context similarity" (x-axis) and "target-neighbour similarity" (y-axis). Again, red dots depict the concrete target nouns, the red line is the concrete regression line, blue dots illustrate abstract target nouns and the blue line shows the abstract regression line. Similar to the previous combination of similarites, there seems to be a visible difference between concrete and abstract targets. Again abstract targets appear to have a higher "target-neighbour similarity" than concrete nouns if they have the same "target-context similarity". Again, this is supported using a regression analysis. P values both for the difference in y-axis intercept (p value  $\approx 0$ ) as well as the slope of the regression lines (p value = 0.304, adjusted  $R^2 = 0.5012$ ) indicate a significant difference between concrete and

abstract targets.

Similar to the "Context-Context and Target-Neighbour" similarity, this supports the underlying hypotheses, since it appears that again abstract targets have a higher similarity to their nearest neighbours in comparison to the concrete targets when being equally similar to their context. This could be caused by the broader, more generalized context of abstract words, that is more likely to overlap with other neighbours, i.e. making them more similar, in comparison to the specialized, narrower contexts of concrete words.

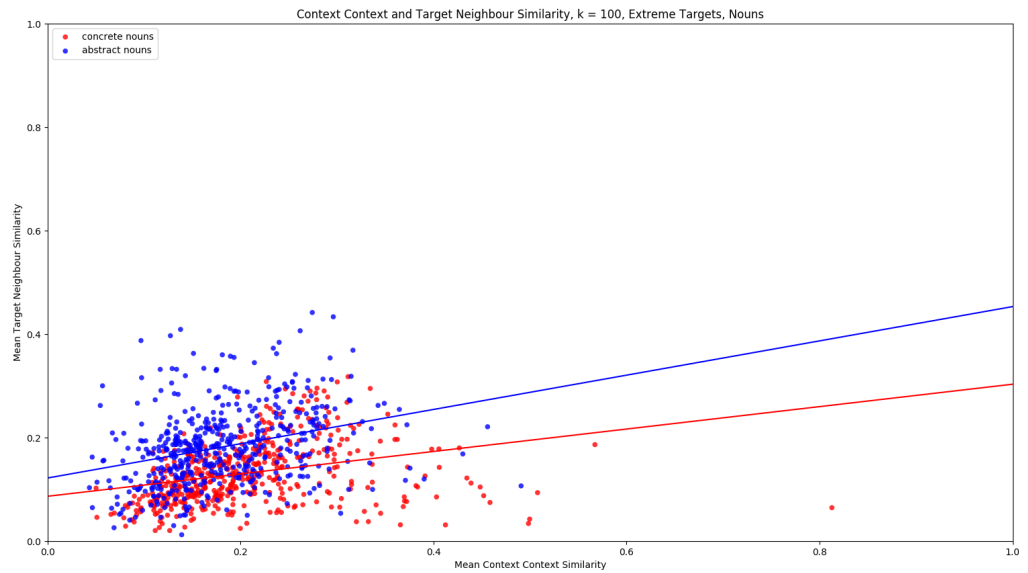


Figure 42: Combination of the "target-context similarity" on the x-axis and the "target-neighbour similarity" on the y-axis. Blue dots indicate abstract words, red words indicate concrete words. The plot includes regression lines for both word classes.

### 3.3 Entropy

Another possible way to look at characteristic behaviour of concrete and abstract words in vector space models, rather than looking at the similarity of certain context words and neighbours, as has been done above using cosine similarity, is to focus on the predictability of abstract and target words (Naumann et al., 2018).

In order to rate the predictability of words one can look at the amount of information conveyed by the words context (Frank, 2013), i.e. if the context a word appears in allows for an easy prediction of the actual word. For example if the word "dog" is strongly associated with the three context words "bark", "pet" and "fetch" it is rather likely that the word "dog" can be predicted. However often a word is not as strongly associated with a limited number of related contexts but rather a larger amount of different contexts like for example the word "nostalgia". Assuming the word nostalgia is best described by the three context dimensions "past", "memory" and "sense" predicting the word is not as easy as before. This can be accounted to the fact that the word "dog" is very strongly related to a smaller amount of context dimensions it is used with. However the word "nostalgia" seems to be associated with more, but weaker context dimensions. In this case, predicting the word is harder i.e. more context dimensions have to be taken into consideration, since every single dimension does provide less information about the word. One possibility to measure the amount of conveyed information is to use entropy (Shannon, 1948). Calculating the entropy of a words context gives an indication to the predictability of the target word (Naumann et al., 2018).

#### 3.3.1 Definition

The entropy as defined by Shannon can be calculated like this (Shannon, 1948):

$$H(X) = - \sum_{x \in X} \Pr(x) \cdot \log_2 \Pr(x)$$

$X$  is a random variable with possible random outcomes  $\{x_1, \dots, x_n\}$  and corresponding outcome probabilities  $\Pr(x_i)$ .

With regard to word vectors, the vector can be interpreted in the following manner in order to calculate the entropy according to Shannon. The entire word vector can be interpreted as the random variable with each context dimension describing one possible random outcome. The outcome probability can then be calculated, somewhat similar to the probabilities described in section 2.1, by taking the value of the corresponding context dimension and dividing it by the sum of all vector entries.

For example given the word vectors  $A = \begin{pmatrix} 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \end{pmatrix}$  and  $B = \begin{pmatrix} 5 \\ 1 \\ 5 \\ 1 \\ 5 \\ 1 \end{pmatrix}$  the en-

trophy can be calculated accordingly. Both vectors consist of six dimensions and for both of them the sum of all vector entries is 18. For vector A all context dimensions  $c_1, \dots, c_6$  have the same outcome probability  $\Pr(c_1) = \dots = \Pr(c_6) = \frac{3}{18}$ . For vector B the probability of dimensions  $c_1, c_3, c_5$ ,  $\Pr(c_1) = \Pr(c_3) = \Pr(c_5) = \frac{5}{18}$  and the probability of dimensions  $c_2, c_4, c_6$ ,  $\Pr(c_2), \Pr(c_4), \Pr(c_6) = \frac{1}{18}$  are the same. Therefore  $H(A) = -6 \cdot \left(\frac{3}{18} \cdot \log_2 \frac{3}{18}\right) \approx 1.42$  and  $H(B) = -3 \cdot \left(\frac{1}{18} \cdot \log_2 \frac{1}{18}\right) + 3 \cdot \left(\frac{1}{18} \cdot \log_2 \frac{5}{18}\right) \approx 1.07$ .

Obviously all context dimensions are equally associated with vector A, for vector B context dimensions 1,3 and 5 are strongly associated while context dimensions 2,4 and 6 are associated less strong. Therefore, as mentioned above, the entropy of vector A should be higher than the entropy of vector B, since vector A appears in a broader range of contexts compared to the more specified contexts of vector B, which indeed proves to be true. Note that in order to calculate entropy values of word vectors instead of using the LMI values described in section 2.1 co-occurrence frequencies are now used instead, since the values are used to calculate probabilities.

Using the entropy of word vectors, again two different measures will be used to investigate characteristics of concrete and abstract words.

### **3.3.2 Target Entropy**

The first way to measure the entropy in this work is to simply calculate the entropy for all target word vectors in a target list and then compare it to the corresponding target list, i.e. calculate the mean entropy for all 500 target abstract nouns and compare them to the 500 concrete target nouns and so on. Again the entropy values are calculated for a range of different context dimensions. Here the top k context dimensions according to cooccurrence frequency are taken into consideration. Contrary to the previous results k values are chosen to be way bigger in this case, ranging from 100 to the all context dimensions included in the ENCOW16AX corpus. Overall for j target words, j entropy values have to be calculated.

Calculating and comparing the entropy for abstract and concrete target words allows for an analysis of the contexts of concrete and abstract words. According to the hypotheses (section 1) abstract words appear in a broader context whereas concrete words appear in a smaller, more specialized context. In order to prove the hypotheses it should therefore be assumed that abstract words have a higher entropy value in comparison to the concrete words.

### **3.3.3 Second Degree Entropy**

Another possible way to use the entropy is to take target vectors, and look at their top k context dimensions. Taking these k dimensions, the entropy of the corresponding vectors is then calculated. For these vectors all context dimensions of the given dimensionality model regardless of the k value are taken into account. This means for j target words and k context dimensions (for the target word) overall j·k entropy values have to be calculated.

The second degree entropy gives an indication of the predicatbility of the

highest rated context dimension of the target words, i.e. if the highest ranked context dimensions of target words are predictable themselves. According to Barsalou and Wiemer-Hastings both concrete and abstract words appear in a concrete context, since the processing of words requires concrete information (Barsalou and Wiemer-Hastings, 2005).

According to the theory of Barsalou and Wiemer-Hastings (2005) it could therefore be expected to see no difference in the second degree entropy of concrete and abstract words.

However it is worth noting that parts of this hypotheses have been found to possibly be wrong by, amongst others, Naumann et al. (2018), as in their research they found that while concrete words seem to appear in a concrete context, abstract words on the other hand do not appear within a concrete but within an abstract context (Naumann et al., 2018).

### 3.3.4 Results

The results both for the "target entropy" as well as the "second degree entropy" will be presented and discussed in this section. Results are presented using all three different dimensionality models i.e. the "full dimensionality", "brysbaert" and "nouns only model". Since there was no visible difference between nouns verbs and adjectives it was decided to focus on the results of the target nouns, since in previous results nouns seem to draw the clearest picture.

Again all concrete results will be shown in red, blue boxes correspond to abstract nouns. Black lines indicate the median value, white dotted lines indicate the mean for each box.

Contrary to previous results it has to be noted that, depending on the dimensionality model used, different amounts of context dimensions are taken into consideration. For example using the "full dimensionality model" entropy values are calculated using all dimensions, which is not possible for the other models, since they restrict the amount of context dimensions.



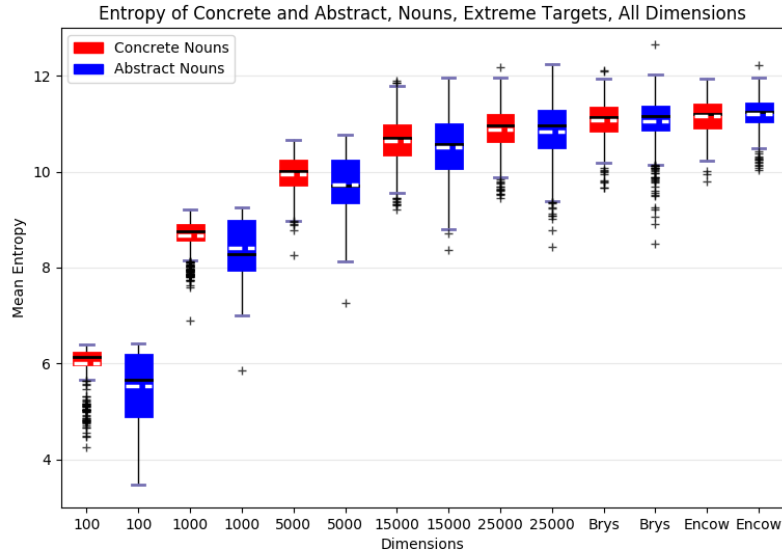


Figure 43: "Target Entropy" for the 500 target nouns calculated using the "full dimensionality model".

**Target Entropy** Figures 43 to 45 illustrate the "target entropy" values for the 500 abstract and 500 concrete target nouns. Depending on the dimensionality models the upper boundary of the context dimensions taken into consideration varies.

For all dimensionality models increasing the number of relevant context dimensions, expectedly increases the entropy values. Relevancy in this case is defined by the co-occurrence frequency of the context dimensions, i.e. context dimensions with higher co-occurrence frequency are considered more relevant. Including less relevant context dimensions is likely to increase the entropy of a given vector due to the definition of the entropy as described in section 3.3.1.

The mean entropy values for concrete target nouns using the "full dimensionality model" increase from 6.01 for the 100 most relevant context dimensions to 11.12 for all context dimensions. However there appears to be a limited growth to the entropy values as increasing the number of context dimensions

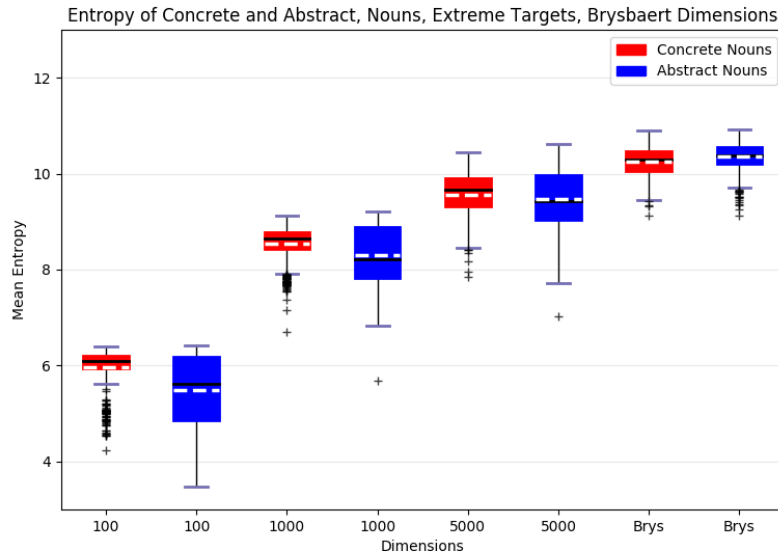


Figure 44: "Target Entropy" for the 500 target nouns calculated using the "brysbaert dimensionality model"

from 15000 (mean 10.69) to 25000 (mean 10.83) only slightly changes the entropy values in comparison to results using a smaller number of context dimensions. This could indicate that after a certain threshold of context dimensions, roughly 15000-20000 context dimensions in this case, the entropy values are only slightly changing.

Regarding the differences of abstract and concrete nouns, looking both at the "brysbaert" and "full dimensionality model" in figures 44 and 43 respectively abstract nouns have a lower entropy value than their concrete counterparts. For 100 context dimensions the mean value for concrete nouns is 6.01 for abstract nouns it is 5.49 (both for the full dimensionality model). Increasing the number of context dimensions closes the gap between concrete and abstract targets, for all dimensions the results are pretty much equal (concrete 11.12, abstract 11.14).

Looking at the results for the "nouns only model" in figure 45 the entropy of abstract and concrete nouns is roughly the same for the same number of

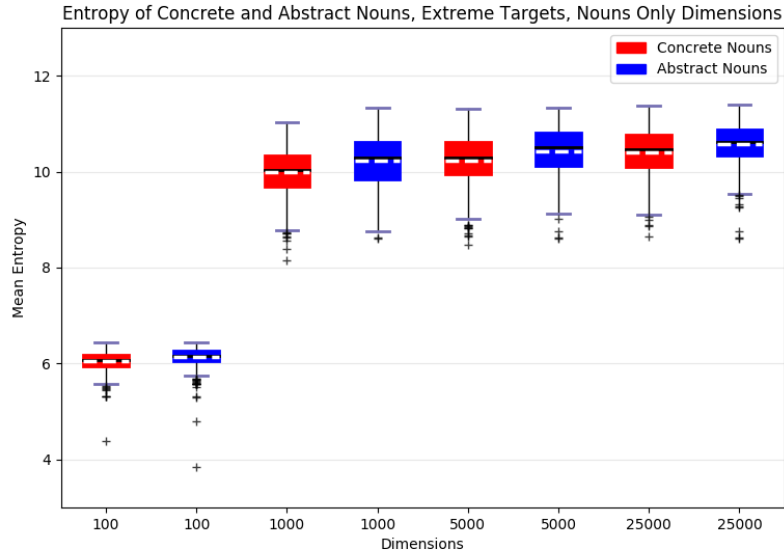


Figure 45: "Target Entropy" for the 500 target nouns calculated using the "nouns only dimensionality model"

context dimensions. For 100 context dimensions the mean value for abstract nouns is 6.02 and 6.0 for concrete targets. Increasing the number of context dimensions slightly widens that gap, for 25000 context dimensions the abstract mean increases to 10.50 and the concrete mean increases to 10.41. However overall there does not seem to be a huge distinction between concrete and abstract targets.

Despite the previous assumption made in section 3.3.2 the observed results differ from the expected results, thus not supporting the underlying hypotheses.

**Second Degree Entropy** The "Second Degree Entropy" for the target nouns is illustrated in figures 46-48 for the "full dimensions", "brysaert" and "nouns only model". Unless further noted, results discussed are valid for all three dimensionality models. Increasing the number of context dimensions that in turn are then considered for entropy calculation increases the overall

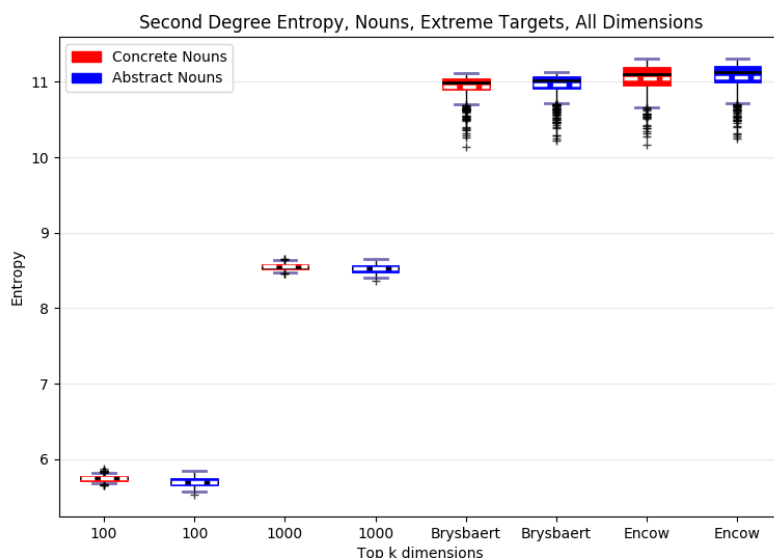


Figure 46: "Second Degree Entropy" for the 500 target nouns calculated using the "full dimensionality model"

entropy values of the target nouns. For 100 context dimensions the "Second Degree Entropy" has a mean value 5.74 for concrete nouns and 5.70 for abstract target nouns with regard to the "full dimensions model". Taking into consideration all context dimensions present in the ENCOW16AX corpus leads to entropy values of 11.1 both for concrete and abstract target nouns. However values seem to grow limited, since entropy values considering only the brysbaert dimensions are at 10.9 for concrete and 11.1 for abstract nouns, i.e. increasing the number of context dimensions after this point leads to very small changes in "Second Degree Entropy" values.

Evaluating the differences between concrete and abstract target nouns it is obvious that "Second Degree Entropy" values for the same number of context dimensions are rather similar for concrete and abstract targets, when looking at the results for the "full dimensionality" and "brysbaert model". Considering that according to Barsalou and Wiemer-Hastings (2005) both concrete and abstract words appear (mostly) in a similar context it would

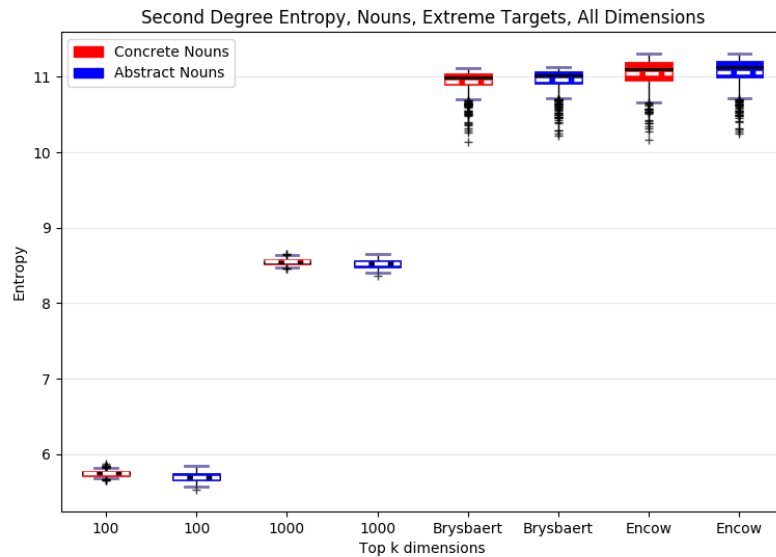


Figure 47: "Second Degree Entropy" for the 500 target nouns calculated using the "brysaert model"

be expected to achieve similar "Second Degree Entropy" values, hence these results could support this theory (Barsalou and Wiemer-Hastings, 2005). However the results shown in figure 48 representing the "nouns only model", show that when looking only at context nouns, concrete targets appear to include context nouns with a lower entropy than the abstract targets. This could indicate that the noun context of abstract nouns is harder to predict than the noun context of concrete words.

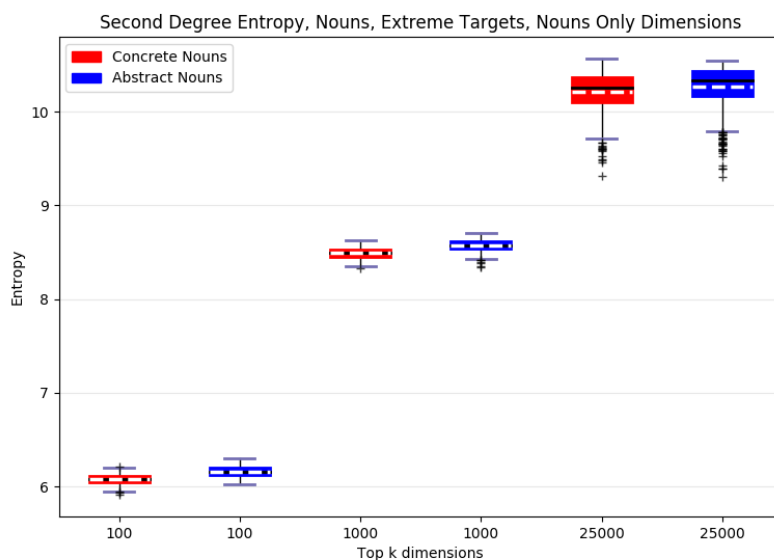


Figure 48: "Second Degree Entropy" for the 500 target nouns calculated using the "nouns only model"

### 3.4 Dimensionality Reduction

Up to this point three different dimensionality models (see section 3) have been used in order to calculate semantic diversity measures, searching for characteristics of both concrete and abstract words in the resulting vector space models.

Even though all three different dimensionality models produce different vector spaces, all three models rely on dimensions that each represent a single context word and just differ in the amount of context words taken into consideration.

Another possibility to reduce the dimensions of a vector space model or a matrix is to perform a dimensionality reduction technique that focuses on describing the characteristics of vectors and matrices while condensing them into a matrix or vectors with fewer dimensions. However, during this process, the possible interpretation of each dimension changes. As a result it is no longer possible to connect a context dimension to a specific context word.

While losing the connection of the context dimensions and their respective context words, dimensionality reduced matrices can condense the specific characteristics and properties of the matrix or the environment described by that matrix. In doing so, specific defining characteristics can be emphasized and outlined using dimensionality reduction techniques. Subsequently, two different dimensionality reduction techniques will be employed in the hope of discovering or improving characteristic behaviour for concrete and abstract words.

### 3.4.1 Singular Value Decomposition

The singular value decomposition (SVD) is an algebraic procedure. SVD factorizes a given matrix, i.e. it decomposes the matrix into a product of other matrices (Banerjee and Roy, 2014). More formally, given a  $m \times n$  matrix  $M$  SVD factorizes  $M$  into three factors  $U$ ,  $\Sigma$  and  $V^*$  such that  $M = U\Sigma V^*$  where  $U$  is unitary matrix with dimensions  $m \times m$ ,  $\Sigma$  is a diagonal  $m \times n$  matrix containing real numbers that are not negative and  $V^*$  is the conjugate transpose of  $n \times n$  unitary matrix  $V$ .

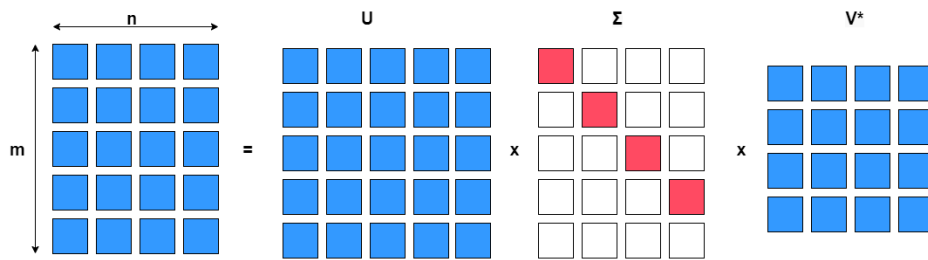


Figure 49: Exemplary illustration of a singular value decomposition.

Figure 49 illustrates the singular value decomposition of a  $m \times n$ .

Using the resulting matrices of a SVD it is now possible to perform a dimensionality reduction and as a result produce a dimension reduced approximation of the original matrix (Baker, 2005). This can basically be done by choosing a value  $t$  and then only taking into consideration the rows of  $U$

as well as columns of  $V^*$  corresponding to the  $t$  highest values of diagonal matrix  $\Sigma$ , leading to the approximated matrix  $\tilde{M}$  of rank  $t$ .

### 3.4.2 word2vec

Another method that results in word vectors with a lower number of dimensions are word embeddings. The basic idea is to find a mapping that maps the high dimensional word vectors to vectors with much fewer dimensions, while preserving its structure. Finding such a mapping is called an embedding in mathematics (Bishop and Crittenden, 2011). While finding such an embedding can be done using dimensionality reduction techniques, another promising possibility is to use neural networks (Mikolov et al., 2013).

One particular approach as described by Mikolov et al. uses neural networks in order to obtain lower dimensional representations for word vectors (Mikolov et al., 2013). Using this approach, a pre-trained set of word vectors was published. This pretrained set contains word vectors for 3,000,000 words and phrases. Each vector has 300 dimensions. The vectors were trained using parts of the "Google-News dataset" ( $\approx 100,000,000,000$  words) <sup>1</sup>

### 3.4.3 Results

Both the SVD as well as word2vec embeddings alter the vector spaces by reducing their dimensionality. Possibly by reducing the vector spaces in these ways characteristics of concrete and abstract words can be preserved or even emphasized, as has been shown to be the case for other semantic applications where these approaches vastly outperform the traditional approach (Mikolov et al., 2013).

Since, as already mentioned, in the process of dimensionality reduction the direct correlation of a context dimension and context word gets lost and hence the meaning of each specific context dimension is hard to impossible to

---

<sup>1</sup><https://code.google.com/archive/p/word2vec/> , last accessed: 11-01-2019, 14:53



understand, it is likely that measures that deal with the context dimensions interpreted as context words will not perform as previously expected for the original approach, i.e. "Target-Context Similarity" as well as "Context-Context Similarity" are prone to not react in the way they previously have. This thesis is underlined by looking at the "Context-Context Similarity" for the word2vec model and the 500 target nouns shown in figure 50. As can be seen, the results using the word2vec model clearly obscure the previously visible trends and indications for the "Context-Context Similarity". Hence it was decided to focus on the similarity measure that does not rely on the interpretation of the context dimensions, the "Target-Neighbour Similarity". Similar to previous figures, red boxes indicate concrete targets, blue boxes indicate abstract targets. The black line shows the median value for each box, the dotted white line shows the mean value.

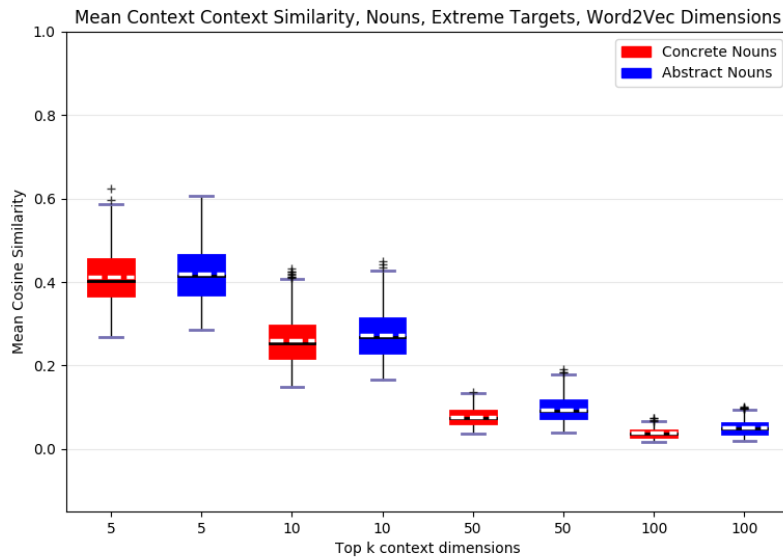


Figure 50: The "Context-Context Similarity" for the 500 concrete and abstract target nouns using the word2vec pre-trained model.

## word2vec

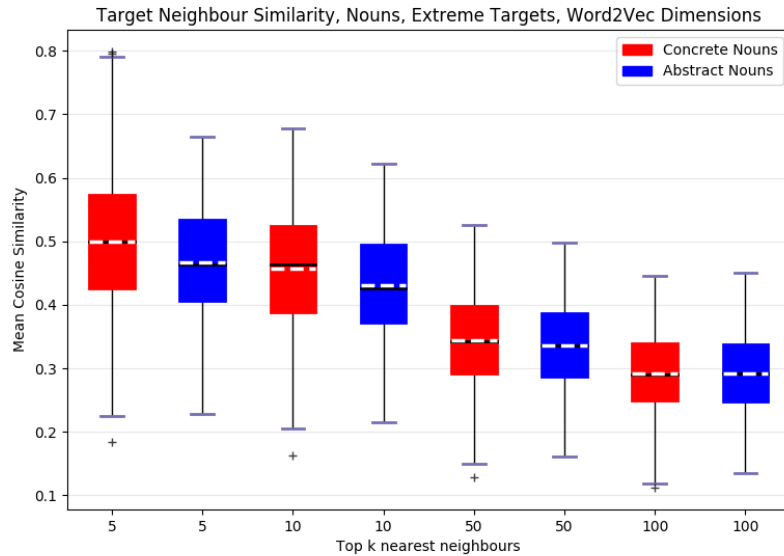


Figure 51: "Target-Neighbour Similarity" of the 500 concrete and abstract target nouns calculated using the pre-trained word2vec model.

**Nouns** Figure 51 illustrates the "Target-Neighbour-Similarity" for the 500 concrete and the 500 abstract target nouns, calculated using the pre-trained word2vec model. Similar to previous results using other dimensionality models, increasing the number of nearest neighbours naturally decreases the mean values of the results. For concrete nouns, the mean values decrease from 0.5 for the 5 nearest neighbours to 0.29 for 100 nearest neighbours. For abstract nouns the mean is 0.42 for 5 neighbours and 0.29 for 100. Comparing the similarity of concrete and abstract targets and a set number of nearest neighbours, contrary to the previously used three dimensionality models there is a visible gap between concrete and abstract target for a smaller k, i.e. a mean of 0.5 for concrete nouns and 0.46 for abstract nouns for  $k = 5$ . For increasing k this gap closes, however, these results could indicate that changing the vector space to a word2vec approach, characteristics

of concrete and abstract words are preserved and even emphasized in comparison to other models.

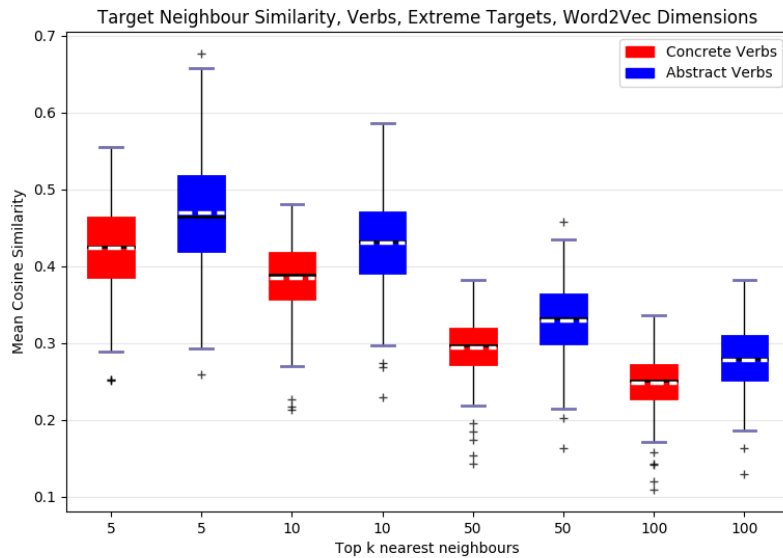


Figure 52: "Target-Neighbour Similarity" of the 200 concrete and abstract target verbs calculated using the pre-trained word2vec model.

**Verbs** Shown in figure 52 are the results for the "Target-Neighbour Similarity" calculated using the 200 abstract as well as concrete target verbs and the pre-trained word2vec model. Again, increasing the number of nearest neighbours taken into consideration expectedly decreases the overall cosine similarity. Abstract mean values change from 0.46 for  $k = 5$  to 0.27 for  $k = 100$ . Concrete values decrease from 0.43 ( $k = 5$ ) to 0.25 ( $k = 100$ ). In accordance with all previous results for "Target-Neighbour Similarity" and the verb targets, the abstract verbs score higher overall similarity values compared to their concrete counterparts, thus not supporting the underlying hypothesis described in section 1.

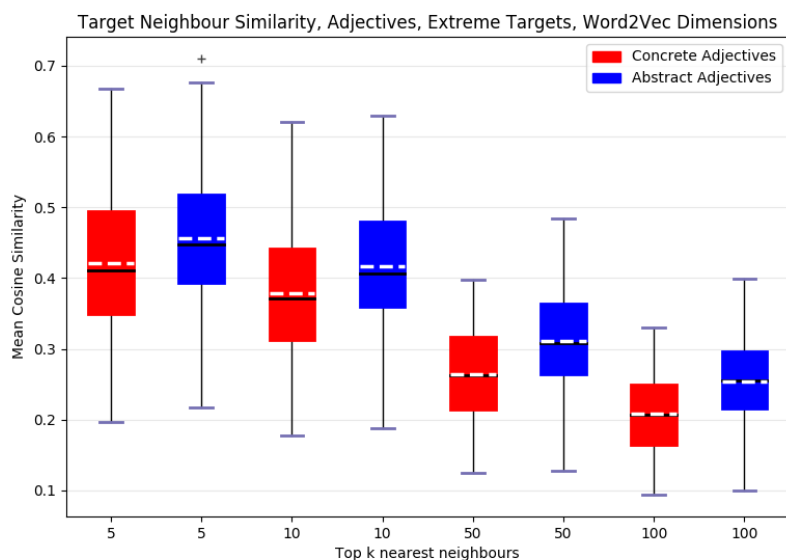


Figure 53: "Target-Neighbour Similarity" of the 200 concrete and abstract target adjectives calculated using the pre-trained word2vec model.

**Adjectives** The results regarding the "Target-Neighbour Similarity" calculated for the abstract and concrete adjective target lists are shown in figure 53. In accordance with all previous results as well as the expectations, increasing the number of nearest neighbours used to calculate the cosine similarity decreases the corresponding mean values, i.e. concrete means decrease from 0.42 to 0.21 and abstract means decrease from 0.45 to 0.25 when increasing the number of nearest neighbours from 5 to 100.

Looking at results for concrete and abstract adjectives for the same k, similar to previous results abstract adjectives have higher similarity scores than the corresponding concrete verb targetset. Again, this does not support the hypotheses that concrete words appear in distinct, smaller contexts whereas abstract words appear in broader, more general contexts.

## Singular Value Decomposition

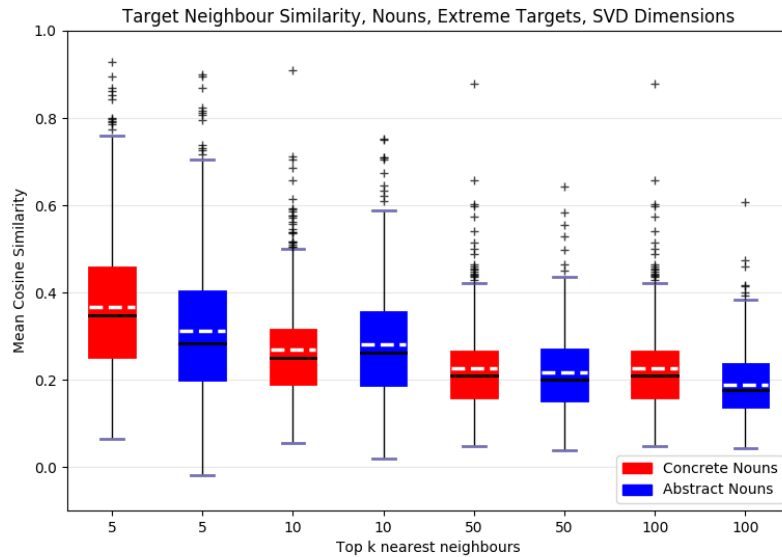


Figure 54: "Target-Neighbour Similarity" of the 500 concrete and abstract target nouns calculated using the SVD dimensionality reduction technique.

**Nouns** The "Target-Neighbour Similarity" for the 500 concrete and abstract target nouns using the dimension reduced model (via SVD) are represented in figure 54. Again, similarity values decrease for an increasing number of nearest neighbours. Concrete similarity values decrease from 0.47 for  $k = 5$  to 0.21 for  $k = 100$ . Abstract values decrease from 0.31 to 0.19 for 5 and 100 neighbours respectively. However this effect has to be expected, since, by definition increasing the number of nearest neighbours, the similarity between the target and the included neighbours decreases.

Comparing the results across concrete and abstract target nouns, contrary to the three originally used dimensionality models, however, in accordance with the results for the word2vec model, concrete nouns score visibly higher similarity scores than the abstract nouns, indicating that, indeed, dimensionality reduction could not only preserve but even strengthen concrete and abstract

characteristics encoded in a vector space.

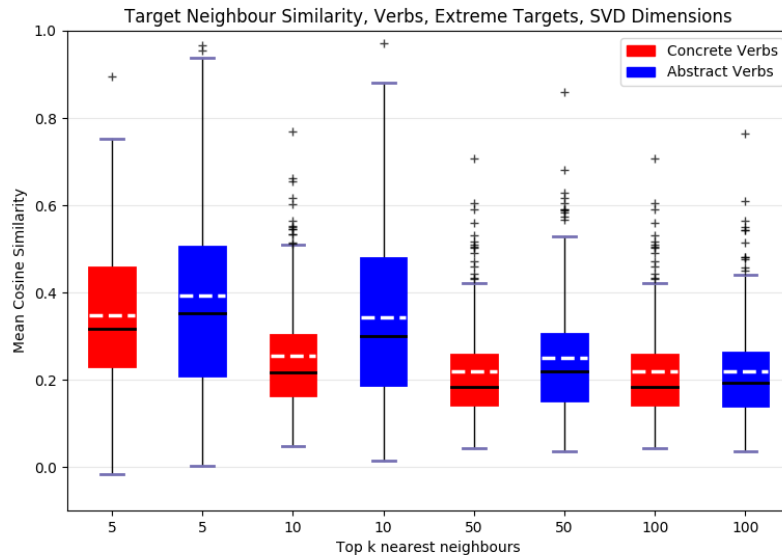


Figure 55: "Target-Neighbour Similarity" of the 200 concrete and abstract target verbs calculated using the SVD dimensionality reduction technique.

**Verbs** The same results as described above but for the 400 different target verbs are shown in figure 55. In accordance with the definition of "Target-Neighbour Similarity", an increase of nearest neighbours considered leads to a decrease in mean similarity values. For the concrete verbs, mean values change from 0.33 for the 5 nearest neighbours to 0.22 for the 100 nearest. Over the same span, abstract scores decrease from 0.40 to 0.22. Similar to the three originally used dimensionality models, as well as the word2vec pre-trained model, abstract verbs show a higher similarity to their nearest neighbours in comparison to their concrete counterparts, not supporting the underlying hypothesis (see section 1).

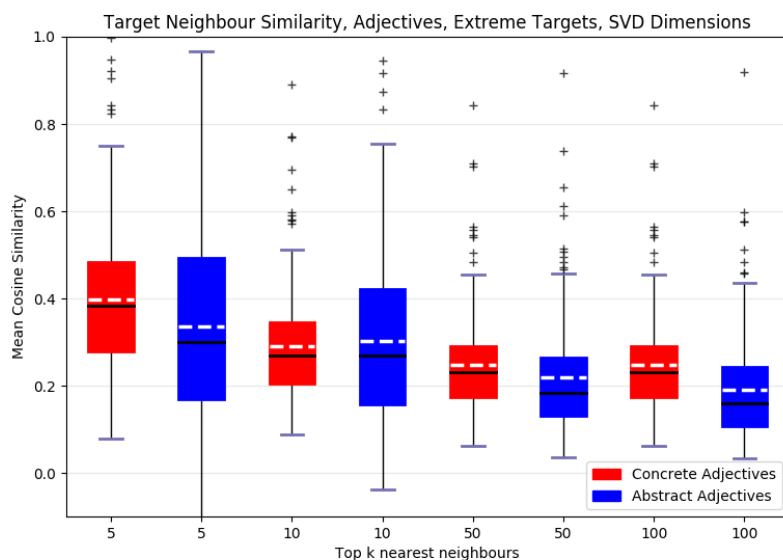


Figure 56: "Target-Neighbour Similarity" of the 200 concrete and abstract target adjectives calculated using the SVD dimensionality reduction technique.

**Adjectives** Concluding the results section, the final results for the "Target-Neighbour Similarity" for the abstract and concrete adjective target sets using the SVD dimensionality reduced model are represented in figure 56. Once again, the increase of nearest neighbours considered leads to an expected decrease in similarity, as has been argued previously. Mean concrete values switch from 0.4 for 5 neighbours to 0.23 for 100 neighbours. Abstract values change from 0.31 to 0.19 for the 5 and 100 nearest neighbours respectively.

Contrary to the three previously used dimensionality models, as well as the word2vec model, for the same amount of nearest neighbours, concrete adjectives actually score higher similarity values than the abstract adjectives. While this could be an outlier result, this could also further indicate the usefulness and possibilities provided by dimensionality reduction regarding the topic of abstract and concrete words.

## 4 Conclusion

The main goal of this thesis is to find characteristics of neighbourhood vector spaces both for concrete and abstract words. In order to do so five different vector space models are used. Each of the different vector spaces uses a different number of context dimensions. While three of the models use the traditional approach of word-word matrices, i.e. each dimension can be mapped to exactly one word, the two other models lose this property by performing dimensionality reduction techniques, by either performing a mathematical matrix factorization or using machine learning in the form of neural networks.

Employing all different models on various different similarity measures, like the "Target-Context Similarity" or the "Context-Context Similarity", as well as the entropy, an in depth description and analysis regarding the characteristics of vector spaces for concrete and abstract words is provided.

Most results were in line with previous studies, for example by Naumann et al., strongly suggesting that the main underlying hypothesis (see section 1) that concrete words appear in a smaller, more specialized context while abstract words tend to appear in a broader, distinct context, holds true. Results especially indicated that this is particularly true for concrete and abstract nouns, probably hinting at the possibility that the semantic concept of concrete versus abstract words can best be used and is best described by nouns in comparison to verbs and adjectives.

Regarding the results of the dimensionality reduced vector spaces, the results seem to prove that, while, somewhat expectedly, loosing the connection between context word and context dimension renders similarity measures employing this connection ("Target-Context" and "Context-Context") less usefull or even useless, reducing the dimensions seems to preserve or even enhance the characteristics of the neighbourhood vector spaces regarding concrete and abstract words.

A few results especially with regard to the entropy seem to not support



the underlying hypothesis, possibly hinting at flaws in the hypothesis or the underlying vector space models and measuring methods.

Further work could be done by employing other dimensionality reduction methods in order to investigate the change in characteristics for these vectors spaces compared to the regular ones. Also possibly other measures, like for example the number of zero dimensions and further analysis of the concreteness and abstractness of the context dimensions could be done as in Naumann et al. (2018).

## References

- Kirk Baker. Singular Value Decomposition Tutorial. *The Ohio State University*, 24, 2005.
- Sudipto Banerjee and Anindya Roy. *Linear Algebra and Matrix Analysis for Statistics*. Chapman and Hall/CRC, 2014.
- Lawrence W Barsalou and Katja Wiemer-Hastings. Situating Abstract Concepts. *Grounding cognition: The role of perception and action in memory, language, and thought*, pages 129–163, 2005.
- Sai Abishek Bhaskar, Maximilian Köper, Sabine Schulte Im Walde, and Diego Frassinelli. Exploring Multi-Modal Text+ Image Models to Distinguish between Abstract and Concrete Nouns. In *Proceedings of the IWCS workshop on Foundations of Situated and Multimodal Communication*, 2017.
- Richard L Bishop and Richard J Crittenden. *Geometry of Manifolds*, volume 15. Academic press, 2011.
- Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. Concreteness Ratings for 40 Thousand Generally Known English Word Lemmas. *Behavior Research Methods*, 46(3):904–911, 2014.
- Trevor Cohen and Dominic Widdows. Empirical Distributional Semantics: Methods and Biomedical Applications. *Journal of Biomedical Informatics*, 42(2):390–405, 2009.
- Max Coltheart. The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology Section A*, 33(4):497–505, 1981.
- John R Firth. A Synopsis of Linguistic Theory, 1930-1955. *Studies in Linguistic Analysis*, 1957.
- Stefan L Frank. Uncertainty Reduction as a Measure of Cognitive Load in Sentence Comprehension. *Topics in Cognitive Science*, 5(3):475–494, 2013.

- Diego Frassinelli, Daniela Naumann, Jason Utt, and Sabine Schulte Im Walde. Contextual Characteristics of Concrete and Abstract Words. In *IWCS 2017—12th International Conference on Computational Semantics—Short papers*, 2017.
- Zellig S. Harris. Distributional Structure. *WORD*, 10(2-3):146–162, 1954.
- Felix Hill, Anna Korhonen, and Christian Bentz. A Quantitative Empirical Analysis of the Abstract/Concrete Distinction. *Cognitive science*, 38(1): 162–177, 2014.
- Paul Hoffman and Anna M Woollams. Opposing Effects of Semantic Diversity in Lexical and Semantic Relatedness Decisions. *Journal of Experimental Psychology: Human Perception and Performance*, 41(2):385, 2015.
- Paul Hoffman, Matthew A Lambon Ralph, and Timothy T Rogers. Semantic Diversity: A Measure of Semantic Ambiguity Based on Variability in the Contextual Usage of Words. *Behavior Research Methods*, 45(3):718–730, 2013.
- Thomas K Landauer and Susan T Dumais. A Solution to Plato’s Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, 1(2):211–240, 1997.
- Ken McRae and Michael Jones. Semantic Memory. *The Oxford handbook of cognitive psychology*, 206, 2013.
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- Daniela Naumann, Diego Frassinelli, and Sabine Schulte im Walde. Quantitative Semantic Variation in the Contexts of Concrete and Abstract Words. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, pages 76–85, 2018.

- Allan Paivio. *Imagery and Verbal Processes*. Psychology Press, 2013.
- Allan Paivio, Mary Walsh, and Trudy Bons. Concreteness Effects on Memory: When and Why? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5):1196, 1994.
- Eyal Sagi, Stefan Kaufmann, and Brady Clark. Semantic Density Analysis: Comparing Word Meaning Across Time and Phonetic Space. In *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics*, pages 104–111. Association for Computational Linguistics, 2009.
- Roland Schäfer. Processing and Querying Large Web Corpora with the COW14 Architecture. In Piotr Bański, Hanno Biber, Evelyn Breiteneder, Marc Kupietz, Harald Lungen, and Andreas Witt, editors, *Proceedings of Challenges in the Management of Large Corpora 3*, Lancaster, 2015. UCREL, IDS.
- Roland Schäfer and Felix Bildhauer. Building Large Corpora from the Web Using a New Efficient Tool Chain. In *Proceedings of the Eight International Conference on Language Resources and Evaluation*, pages 486–493, Istanbul, Turkey, 2012. European Language Resources Association (ELRA).
- Helmut Schmid. Part-of-Speech Tagging with Neural Networks. In *Proceedings of the 15th conference on Computational linguistics-Volume 1*, pages 172–176. Association for Computational Linguistics, 1994.
- Eugene E Schultz Jr. and Karen E Woodall. A Comparison of Pictorial and Narrative Learning Mediators in Lower- and Middle-Class Children. *The Journal of General Psychology*, 102(1):147–152, 1980.
- Paula J Schwanenflugel. Why are Abstract Concepts Hard to Understand? *The Psychology of Word Meanings*, page 223, 2013.
- Claude Elwood Shannon. A Mathematical Theory of Communication. *Bell system technical journal*, 27(3):379–423, 1948.

Amit Singhal. Modern Information Retrieval: A Brief Overview. *IEEE Data Eng. Bull.*, 24:35–43, 2001.

Gabriella Vigliocco, Stavroula-Thaleia Kousta, Pasquale Anthony Della Rosa, David P Vinson, Marco Tettamanti, Joseph T Devlin, and Stefano F Cappa. The Neural Representation of Abstract Words: The Role of Emotion. *Cerebral Cortex*, 24(7):1767–1777, 2013.

Lin Yui, Roslin Ng, and Hiran Perera-WA. Concrete vs Abstract Words—What do you Recall Better? A Study on Dual Coding Theory. Technical report, PeerJ Preprints, 2017.

## **Erklärung**

Ich versichere, diese Arbeit selbstständig verfasst zu haben. Ich habe keine anderen als die angegebenen Quellen benutzt und alle wörtlich oder sinngemäß aus anderen Werken übernommene Aussagen als solche gekennzeichnet. Weder diese Arbeit noch wesentliche Teile daraus waren bisher Gegenstand eines anderen Prüfungsverfahrens. Ich habe diese Arbeit bisher weder teilweise noch vollständig veröffentlicht. Das elektronische Exemplar stimmt mit allen eingereichten Exemplaren überein.

Datum und Unterschrift:

## **Declaration**

I hereby declare that the work presented in this thesis is entirely my own. I did not use any other sources and references than the listed ones. I have marked all direct or indirect statements from other sources contained therein as quotations. Neither this work nor significant parts of it were part of another examination procedure. I have not published this work in whole or in part before. The electronic copy is consistent with all submitted copies.

Date and Signature: