Universität Stuttgart

Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart
Pfaffenwaldring 5B
D-70569 Stuttgart

Bachelor thesis

# Automatic discrimination between abstract and concrete word classes

Momo Takamatsu

Study program: B.Sc. Machinelle Sprachverarbeitung

Examiner: Apl. Prof. Dr. Sabine Schulte im Walde
Superviser: Apl. Prof. Dr. Sabine Schulte im Walde

Commenced: 01.09.2023
Completed: 28.02.2024

**Erklärung (Statement of Authorship)**

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst habe und dabei keine andere als die angegebene Literatur verwendet habe. Alle Zitate und sinngemäßen Entlehnungen sind als solche unter genauer Angabe der Quelle gekennzeichnet. Die eingereichte Arbeit ist weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen. Sie ist weder vollständig noch in Teilen bereits veröffentlicht. Die beigefügte elektronische Version stimmt mit dem Druckexemplar überein.[1]

Stuttgart, 31.01.2024 ______________________________

(Momo Takamatsu)

[1] This thesis is the result of my own independent work, and any material from work of others which is used either verbatim or indirectly in the text is credited to the author including details about the exact source in the text. This work has not been part of any other previous examination, neither completely nor in parts. It has neither completely nor partially been published before. The submitted electronic version is identical to this print version.

# Contents

# 1 Introduction and Related Works

## 1.1 Introduction

Abstraction in contrast to concreteness has been an interesting topic in the fields of Computational Linguistics and Psycholinguistics. One of the core differences between abstract and concrete concepts is perceptibility (Connell and Lynott, 2012). For instance, *umbrella* can be perceived with visual and tangible senses and is therefore categorized as a concrete concept. Conversely, *afterlife* is abstract since, unlike *umbrella*, it cannot be directly experienced by senses such as tangible, auditory or visual senses. Furthermore, the Context Availability Theory (Schwanenflugel and Shoben, 1983) suggests that meaning arises from appropriate context. When processing abstract or concrete words, without context reading times and lexical decision times for abstract words were longer and more challenging, whereas with appropriate context neither reading times nor lexical decision times differed between abstract and concrete words, emphasizing the importance of context for abstract concepts. From a cognitive point of view, the brain processes concrete concepts more easily than abstract ones for tasks such as single-word recognition (Strain et al., 1995). This phenomenon is seen in the language development of children as concrete words are learnt first (Bolognesi et al., 2020). In Computational Linguistics, abstraction is especially popular for text modelling in computational semantics. Examining the nature of abstraction in words can help for tasks like detection of figurative language such as metaphors. Lakoff and Johnson (1980) explain that metaphor is a method to transfer knowledge from concrete to abstract domains. Through the information of abstraction, language that is used metaphorically, conveying meaning beyond literal interpretations, can be identified. To gain more insight into abstract and concrete concepts, several studies dealt with the analysis of concreteness by examining the context based on English texts, particularly the concreteness on word level, applying different measures such as entropy (further explained in Section 1.2) on nouns, adjectives and verbs (Naumann et al., 2018) or only nouns and verbs (Frassinelli et al., 2017).

In this study, the generalization of previously mentioned existing measures (Naumann et al., 2018) is examined by applying them first on single words (nouns, adjectives and verbs) such as *miracle* and *water*, and then on semantic WordNet classes (noun classes). Single words represent meanings, while word classes categorize words based on semantic relations (e.g. synonyms). For example, the single word *water* represents a specific type of liquid, and the word class *liquid* contains words that are semantically related such as *drink, medium, water, beverage, liquor* etc. First, the distribution of the concreteness of context words is analyzed to examine whether the context of abstract/concrete concepts tends to be abstract/concrete. The concreteness is measured by concreteness ratings that are based on human judgment scaling from 1.0 (very abstract) to 5.0 (very concrete). As briefly mentioned earlier, the more words can be experienced through sensory perceptions such as visual and tangible senses, the more concrete they are. Abstract words, on the other hand, need words to be described. Secondly, the analysis of the most commonly co-occurring context words is performed with cosine similarity. Additionally, similarities and differences between the single words and the word classes are investigated by comparing the results for both single words and word classes, as well as measuring the distances between the context distribution of word classes and of the set of single words through cosine similarity.

In the following, first related works are introduced. Then, in Section 2 the research questions are specified. Afterward, in Section 3 materials and the aforementioned two methods are described. Finally, the results for both sets of single words and word classes are discussed in Section 4, and concluded by summarizing the main findings of this work in Section 5.

## 1.2 Related Works

The closest work to this thesis is Naumann et al. (2018) which introduces three measures to assess semantic variations in the contexts of concrete and abstract across a large set of English nouns, adjectives, and verbs. The first measure analyzes concrete and abstract co-occurrences, while the second measure investigates the semantic diversity of the context through cosine similarity, and lastly the entropy of concrete and abstract words is computed. This thesis implemented the first two measures for English nouns, adjectives and verbs. In contrast to Naumann et al. (2018) where local mutual information is utilized to examine the semantic diversity of the context, the second measure was implemented using the co-occurrence frequency to calculate the cosine similarity. Since all three measures yielded similar results when applied on target sets with window sizes 2, 10 and 20 suggesting that the window size is not very significant, for this thesis, only one context size, 20, was selected. Overall, Naumann et al. (2018)'s results revealed that concrete words primarily tend to co-occur with a concrete context, while for abstract words, abstract adjectives and verbs were seen with abstract context words. Additionally, lower entropy values for concrete context indicated that the context is more predictable for concrete targets compared to abstract targets.

Another quantitative investigation of similarities and differences between concrete and abstract words was conducted by Frassinelli et al. (2017). After assessing the concreteness in distributionally similar words, the concreteness of the context was analyzed at type level and at token level - focusing solely on nouns, as well as only employing a single context window size. Similar to Naumann et al. (2018), Frassinelli et al. (2017) examine the semantic diversity for distributionally similar words using local mutual information and cosine similarity. Frassinelli et al. (2017) reported that concrete words co-occurred with rather concrete words, whereas abstract words co-occurred with an abstract to mid-ranged (concreteness) context.

In a different study carried out by Schulte im Walde and Frassinelli (2021), various measures to examine semantic abstraction were provided. More precisely, frequency and word entropy were used to distinguish concrete and abstract words. Moreover, neighborhood density variants were applied to explore the diversity regarding the context of concrete and abstract words in terms of the abstract–concrete dichotomy. Inspired by Schulte im Walde and Frassinelli (2021), among the four presented variants, the target-context diversity variant was examined in this thesis. Like the aforementioned works, Schulte im Walde and Frassinelli (2021) utilized local mutual information and the cosine similarity. The results show that all measures are reliable achieving a precision of higher than 0.7. Moreover, Schulte im Walde and Frassinelli (2021) reveal that general words tend to be more frequently used and have lower entropy values signaling higher predictability than specific words. The findings align with the previously mentioned works that concrete words tend to co-occur with a concrete context with more variable concreteness scores, while abstract words tend to co-occur with abstract words with very low concreteness scores.

All three studies contribute insights into words with respect to abstraction (or concreteness) using different approaches to analyze the relationships between abstract and concrete words in different linguistic contexts.

# 2 Research Questions

The main research questions are *how does the context behave around abstract and concrete concepts? And how well is the generalization from word to class level?*

To investigate the context around abstract and concrete concepts, the following two measures referred from Naumann et al. (2018) were used:

(1) The concreteness of context words for abstract/concrete target words (distribution)

(2) The similarity of the context words and their target words (diversity, density)

As mentioned earlier (Section1.1), both measures were firstly applied on abstract and concrete target words. Secondly, instead of the context of target words, the context with respect to the entities within word classes was examined. For example, for each word class such as *liquid* and *cactus*, the context of the words *alcohol, ammonia, antifreeze, beverage, distillate, distillation, drink, elixir, ink, instillation, water, medium, potable, spill, tuberculin*, and *liquor* (= entities within the class *liquid*) and *cholla, mezcal, nopal, peyote, saguaro* (= entities within the class *cactus*) are analyzed, respectively.

Expectations are that for abstract words, the distribution of the context is more abstract and vice versa for concrete words. Similarly, since abstract words co-occur in more diverse context and often need to be described by other words (Brysbaert et al., 2014), a more diverse context is expected for abstract words compared to concrete words.

For the word classes, following research questions are formulated:

1. Do the word classes show similar context behavior as the target words (generalization of measures)?
   1a. Is the size of the word class relevant?
   1b. Is the choice of the word class influencing?

For instance, is the distribution of the concreteness of the context of concrete target words such as *water* similar to the concreteness distributions of the word classes *liquid* and *cactus* (concrete word classes)? Or do *liquid* and *cactus* show different results since both word classes contain a different amount of words (16 and 5, respectively) and categorize words of different topics?

I hypothesize that depending on the size and the choice, the word classes should produce similar/dissimilar results to the extremes. Possibly, smaller word classes tend to be more sparse, and therefore not too reliable.

# 3 Methods and Materials

In this section, the used methods and materials are explained. Firstly, the origin of the data and its formats are explained, as well as shown through examples. Next, the procedure of measuring the concreteness of the context words of the corresponding target words is described, followed by an explanation of the calculation of the cosine similarity between the five strongest context words of each target word. At last, the application of the previous two measures on word classes is demonstrated. For implementation, the Python programming language was used.

## 3.1 Data

In this study, three datasets were employed. The first dataset, the Brysbaert collection (Brysbaert et al., 2014), contains a wide range of information about words including concreteness ratings and a variety of part-of-speech (POS) tags. The dataset for the target words is derived from the same collection and includes concreteness ratings, POS-tags,and co-occurrence frequencies. The last dataset enlists all context words associated with the target words. For each target word, the POS-tag of the target word, the context words and their respective POS-tags, as well as the co-occurrence frequency is given. For the word classes, WordNet - an online large lexical database of English that categorizes words based on semantic relations - is used.
In the subsequent sections, after presenting information and descriptions of the Brysbaert et al. (2014) collection in Section 3.1.1, in Section 3.1.2 the dataset for the target words is explained, followed by the description of the dateset of the context words in Section 3.1.3. Lastly, WordNet is introduced in Section 3.1.4.

### 3.1.1 Brysbaert et al. (2014) collection

The Brysbaert collection includes 39,954 commonly known English words and various information such as their concreteness ratings. The words were collected from the SUBTLEX-US corpus (Brysbaert and New, 2009), the English Lexicon Project

(Balota et al., 2007), the British Lexicon Project (Keuleers et al., 2012), the Corpus of contemporary American English (Davies, 2009), and words seen in other various rating studies, shop catalogs or through general reading. All words were standardized to American English spelling. The ratings stem from 4000 participants that were recruited by Internet crowdsourcing. For each participant, meta data such as age, gender, first language(s), the living place (during the ages 0-7) and the level of education was asked. The educational level varied from (still in) high school to a Doctorate degree. A few raters did not specify their educational background. Originally, the data contained 60,099 English words and additional 2,940 two-word expressions such as *zebra crossing* or *zoom in*. Concreteness annotations by non-native speakers as well as words with missing responses (<25 responses) were removed. Furthermore, words that were not known by at least 85% of the raters were discarded as well. Through this data trimming a subset of 39,954 English words, of which 2,896 are two-word expressions, resulted.

The concreteness ratings ranged from 1 to 5 using whole numbers, where 1 indicated clearly abstract and 5 clearly concrete. Words that were unknown to the rater were indicated with the letter $N$ and documented in the final outcome. The instructions specified measuring the concreteness based on one's experience with tangible, visual, auditory, gustatory and olfactory senses. The more the word can be experienced in real life, the more concrete it is. Whereas, abstract words are language-based, meaning that they require to be described by other words for understanding. The collection was validated through high correlations with existing concreteness rating datasets.

The Brysbaert collection encompasses nine columns in total. As seen in the example in Table 1, starting from the the first column with the word itself, followed by information about whether the word is a single word (= 0) or a two-word expression (= 1), the mean concreteness rating, the standard deviation of the concreteness ratings, the number of raters annotating the word as unknown, the total number of raters for the word, the number of raters in percentage who knew the word, the frequency count from the SUBTLEX-US corpus (Brysbaert and New, 2009) and the POS information collected from the SUBTLEX-US corpus (Brysbaert et al., 2012).

Moreover, in this study a version with an additional post-processing step proposed by Frassinelli et al. (2017) was employed. In the post-processed version, the POS-tags were revised depending on POS information in the ENCOW corpus (explained in more detail in Section 3.1.3). The most frequently occurring POS in the ENCOW corpus was assigned to the word. If the POS-tag differed from Brysbaert et al., or the POS-tag did not represent at least 95% of the dominant POS-tag, the word was discarded. Furthermore, words with a frequency less than 10,000 in the ENCOW corpus were also omitted. The POS-tags include standard POS-tags like adjective, adverb, preposition, noun, verb etc., as well as tags for numbers (= Number), words that cannot be classified (= Unclassified) and words that could not be tagged (= #N/A).

| Word | Bigram | Conc.M | Conc.SD | Unknown | Total | Percent known | SUBTLEX | Dom Pos |
|---|---|---|---|---|---|---|---|---|
| roadsweeper | 0 | 4.85 | 0.37 | 1 | 27 | 0.96 | 0 | 0 |
| sled | 0 | 5.00 | 0.00 | 0 | 28 | 1.00 | 149 | Adjective |
| plunger | 0 | 4.96 | 0.20 | 0 | 26 | 1.00 | 48 | Adjective |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| metaphysically | 0 | 1.31 | 0.62 | 3 | 29 | 0.90 | 1 | Adverb |
| misfortunately | 0 | 1.31 | 0.55 | 3 | 29 | 0.90 | 1 | Adverb |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| theirs | 0 | 2.40 | 1.40 | 0 | 30 | 1.00 | 395 | Pronoun |
| plenty | 0 | 2.39 | 1.50 | 0 | 28 | 1.00 | 3178 | Pronoun |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| at | 0 | 2.07 | 1.31 | 0 | 29 | 1.00 | 164072 | Preposition |
| spite | 0 | 2.07 | 1.21 | 0 | 28 | 1.00 | 388 | Preposition |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| collector | 0 | 4.10 | 1.14 | 0 | 29 | 1.00 | 220 | Noun |
| daylight | 0 | 4.10 | 1.11 | 0 | 29 | 1.00 | 488 | Noun |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| sass | 0 | 2.26 | 1.38 | 2 | 29 | 0.93 | 47 | Name |
| hardy | 0 | 2.22 | 1.01 | 1 | 28 | 0.96 | 188 | Name |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| offend | 0 | 1.68 | 1.03 | 0 | 25 | 1.00 | 266 | Verb |
| outthink | 0 | 1.66 | 1.04 | 1 | 30 | 0.97 | 14 | Verb |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| seven | 0 | 3.72 | 1.62 | 0 | 29 | 1.00 | 5327 | Number |
| thirteen | 0 | 3.69 | 1.51 | 0 | 29 | 1.00 | 268 | Number |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| ding | 0 | 3.65 | 1.41 | 1 | 27 | 0.96 | 307 | Unclassified |
| amigo | 0 | 3.28 | 1.39 | 1 | 30 | 0.97 | 259 | Unclassified |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| Vice President | 1 | 4.31 | 1.23 | 0 | 29 | 1.00 | 0 | #N/A |
| PIN number | 1 | 4.17 | 1.20 | 0 | 29 | 1.00 | 0 | #N/A |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |

Table 1: Brysbaert et al. (2014) collection (concreteness ratings)

### 3.1.2 Extremes (target words)

The dataset for the target words were derived from the aforementioned Brysbaert et al. (2014) collection and represent subsets of the Brysbaert et al. (2014) collection. As shown in Table 3, each subset consists of five columns with the word itself, the concreteness rating (Brysbaert et al., 2014), the POS, the POS-tag and the co-occurrence frequency extracted from a version of the *English COW* corpus, the ENCOW16AX corpus. This corpus, besides the co-occurrence frequency includes additional information such as the context words of the target words, their POS-tags etc., and is introduced in more detail in the next section.

Since the target words are subsets of the Brysbaert et al. (2014) collection, the same restrictions were applied, as well as further restrictions were made to choose the target words. Firstly, not just the POS of the target words was limited to nouns, adjectives and verbs, but also, like the Brysbaert et al. (2014) collection, words with a co-occurrence frequency of <10,000 were omitted. Additionally, only the target words that were tagged with the same POS as the ENCOW, as well as the SUBTLEX corpus (Brysbaert et al., 2012), were kept. These POS-tags needed to represent at least 95% of the dominant POS-tag to keep the word. To avoid unclear results due to mid-ranged concreteness ratings, the focus was put on the extremes of each 500 most abstract and 500 most concrete nouns, 200 most abstract and 200 most concrete adjectives and 200 most abstract and 200 most concrete verbs, totaling of 1,800 target words.

| | Nouns (NN) | Adjectives (ADJ) | Verbs (V) |
|---|---|---|---|
| Abstract | 500 | 200 | 200 |
| Concrete | 500 | 200 | 200 |
| **Total:** | **1000** | **400** | **400** |

Table 2: Total of target words

| Word | Conc.M | POS | POS tag | freq |
|---|---|---|---|---|
| oneness | 1.96 | Noun | NN | 12860 |
| miracle | 1.96 | Noun | NN | 130389 |
| liability | 1.96 | Noun | NN | 227363 |
| ... | ... | ... | ... | ... |
| nozzle | 4.91 | Noun | NN | 18609 |
| bottle | 4.91 | Noun | NN | 332667 |
| yogurt | 4.90 | Noun | NN | 29893 |
| ... | ... | ... | ... | ... |
| absurd | 1.64 | Adjective | ADJ | 87385 |
| uncanny | 1.63 | Adjective | ADJ | 18314 |
| predictable | 1.63 | Adjective | ADJ | 73722 |
| ... | ... | ... | ... | ... |
| cloudy | 4.00 | Adjective | ADJ | 25891 |
| stormy | 3.96 | Adjective | ADJ | 15752 |
| juicy | 3.96 | Adjective | ADJ | 24100 |
| ... | ... | ... | ... | ... |
| overwhelm | 1.42 | Verb | V | 84711 |
| suppose | 1.37 | Verb | V | 886431 |
| idealize | 1.19 | Verb | V | 11579 |
| ... | ... | ... | ... | ... |
| handcuff | 4.79 | Verb | V | 19031 |
| drool | 4.61 | Verb | V | 15598 |
| weep | 4.54 | Verb | V | 58266 |
| ... | ... | ... | ... | ... |

Table 3: Examples of extremes (target words)

### 3.1.3 ENCOW16AX (context words)

The *ENCOW16AX* corpus is a POS-tagged version of the *E(nglish) COW* corpus, an English web corpus (Schäfer and Bildhauer (2012); Schäfer (2015)). It consists, as seen in Table 5, of four columns with target words (including POS-tag) , their corresponding context words (including specification of POS-tags), the co-occurrence frequency and the local mutual information (lmi). In the ENCOW16AX corpus, only words of the POS-tags NN (noun), ADJ (adjective) and V (verb) for both target and context words were kept. Additionally, infrequent words with a co-occurrence frequency of <10,000, as well as the top 50 most frequent words such as *'be', 'have'* and *'time'*, were removed. In this study, the ENCOW16AX corpus lists context words of a (context) window size of 20 and is sorted alphabetically irrespective of the POS-tag.

In Table 4 are a few examples of target-context word co-occurrence frequencies, where rows represent target words, while columns represent context words:

| | abandon (V) | abandoned (ADJ) | life (NN) | generation (NN) | think (V) |
|---|---|---|---|---|---|
| abandoned (ADJ) | 289 | 138 | 248 | 0 | 208 |
| generation (NN) | 775 | 0 | 192856 | 62406 | 16995 |
| offend (V) | 83 | 0 | 1621 | 127 | 5665 |
| umbrella (NN) | 0 | 0 | 358 | 0 | 974 |
| turtle (NN) | 0 | 0 | 973 | 51 | 729 |

Table 4: Examples of co-occurrence frequencies

| Target Word (+POS) | Context Word (+POS) | freq | lmi |
|---|---|---|---|
| abandoned:::ADJ | abandon:::V | 289 | 870.67616 |
| abandoned:::ADJ | abandoned:::ADJ | 138 | 688.96722 |
| abandoned:::ADJ | able:::ADJ | 63 | -58.62938 |
| abandoned:::ADJ | abused:::ADJ | 154 | 846.26476 |
| abandoned:::ADJ | access:::NN | 65 | -12.48795 |
| ... | ... | ... | ... |
| helmet:::NN | type:::NN | 798 | 396.77868 |
| helmet:::NN | typical:::ADJ | 90 | 12.56932 |
| helmet:::NN | tyre:::NN | 71 | 48.24947 |
| helmet:::NN | ultimate:::ADJ | 61 | 13.16820 |
| helmet:::NN | uncomfortable:::ADJ | 63 | 59.99858 |
| ... | ... | ... | ... |
| earn:::V | accessible:::ADJ | 106 | -151.64467 |
| earn:::V | accessory:::NN | 106 | -39.41154 |
| earn:::V | accident:::NN | 272 | -246.44250 |
| earn:::V | acclaim:::NN | 1005 | 3311.75766 |
| earn:::V | acclaim:::V | 51 | 69.22158 |
| ... | ... | ... | ... |

Table 5: ENCOW16AX (context words)

### 3.1.4 Word Classes (data)

Word classes were formed based on an online large lexical database of English developed at Princeton University called WordNet (Princeton University). In WordNet, English nouns, adjectives, verbs and adverbs are grouped into sets of synonyms (synsets) like *car* and *automobile* which are linked by conceptual-semantic and lexical relations. Each synset represents a concept. Hence, words with multiple meanings belong to multiple synsets. The relations between the synsets is hierarchical. For instance, for nouns exist a hyponymy-hyperonymy relationship such as *bed* - *furniture* and for verbs a troponymy-entailment relationship like *talk* - *whisper* (troponymy) and *buy* - *pay* (entailment).

In this study, only noun synsets were investigated. The word classes were chosen based on the extremes (target words) from Section 3.1.2. WordNet has a root node *entity* for nouns with the leaves *physical_entity* and *abstract_entity* that is, for this study, interpreted as concrete and abstract, respectively. The hyponymy-hyperonymy information was obtained from previous works (Schulte im Walde and Frassinelli, 2021) where noun and verb synset pairs were extracted from WordNet version 3.0.

The baselines for each chosen class is that it comprises a minimum of five entities (words) that are **not** two-word expressions indicated by '_' like *lemon_peel* (neither hyponyms, nor hyperonyms are two-word expressions). Two-word expressions were therefore removed from the class. Since the Brysbaert et al. (2014) collection includes only American English spelling, words in British English spelling were excluded as well.

The chosen word classes with their hyponyms are listed in the Appendix A.

## 3.2 Concreteness of Context Words

As mentioned in Section 2, to investigate the behavior of the context of abstract and concrete words, the distribution of the contexts of abstract and concrete words based on their concreteness was analyzed by implementing the first measure as follows:

The context words were collected based on the ENCOW16AX corpus (second column, *Context Word (+POS)*) with their concreteness ratings retrieved from the Brysbaert et al. (2014) collection (third column, *Conc.M*). As mentioned in Section 3.1.1, only context words with matching POS-tags between the ENCOW16AX corpus and the Brysbaert et al. (2014) collection were considered. Additionally, context words not existing in the Brysbaert et al. (2014) collection were skipped (not counted) and marked as "not found" in Table 6.

| Context word | POS-tag | co-occ. freqency |
|---|---|---|
| highlight | V | 344 |
| holding | NN | 54 |
| home | NN | 1060 |
| honor | V | 1312 |
| ... | ... | ... |

Table 6: 'Not found' words - example

For each of the six target sets, the concreteness of context words were measured separately. For instance, the context words for the target set *abstract nouns* were measured separately from the context words of the target set *concrete nouns*. To examine the behavior of noun, adjective and verb contexts individually, within each target set, the context words of each target word were grouped based on their POS-tags (NN, ADJ, V). The context words of each POS-tag were then classified by their concreteness ratings into eight concreteness categories, 1 being the class of the lowest concreteness rating and 8 being the class of the highest concreteness rating (Table 7). The concreteness ratings range from 1.0 to 5.0 in 0.5 increments. Each category includes the lower boundary and excludes the upper boundary. For example, the category of concreteness ratings between 1.0 and 1.5 includes 1.0 but excludes

1.5. The next category for concreteness rates from 1.5 to 2.0 again includes 1.5 but excludes 2.0, and so forth.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| 1 - <1.5 | $\geq 1.5 - < 2$ | $\geq 2 - < 2.5$ | $\geq 2.5 - < 3$ | $\geq 3 - < 3.5$ | $\geq 3.5 - < 4$ | $\geq 4 - < 4.5$ | $\geq 4.5 - \leq 5$ |
| concept<br>belief | careful<br>miracle | boast<br>variety | problem<br>return | sale<br>urban | victim<br>wedding | capital<br>solid | telephone<br>student |

Table 7: Concreteness rating categories with examples

For example, given this sample set of three context words *umbrella, pillow*, and *honor* of the POS-tag **NN** with their frequencies (10, 5, 2) and concreteness ratings (4.70, 4.75, 1.30). Based on their concreteness ratings, the noun context words are classified into class 1 and 8 (column *Conc. class*).

| NN context words | Co-occ. freq. | Conc. rating | Conc. class |
|---|---|---|---|
| umbrella | 10 | 4.70 | 8 |
| pillow | 5 | 4.75 | 8 |
| honor | 2 | 1.30 | 1 |

Table 8: Sample set of noun context - example

To investigate the concreteness distribution of the contexts, the concreteness proportions ($proportion(c_i)$) were then calculated based on the co-occurrence frequencies of the context words within the respective concreteness category. These values were used to plot graphs for visualization.

The formula $proportion(c_i)$ is as follows:

(1) $$proportion(c_i) = \frac{1}{\sum frequency_c} * \sum freqency_{c_i}$$

Here, $frequency_c$ is the accumulated co-occurrence frequency of all eight categories and $freqency_{c_i}$ is the co-occurrence frequency of a single class.
As mentioned above, the proportions are computed for each concreteness category within each POS-tag.

Taking the sample set from above (Table 8), the proportions are calculated for each concreteness category as follows (in this case categories 1 and 8):

$$proportion(\mathbf{c_8}) = \frac{1}{10+5+2} * (10+5) \approx \mathbf{0.882}$$

$$proportion(\mathbf{c_1}) = \frac{1}{10+5+2} * 2 \approx \mathbf{0.117}$$

In this sample set, it means, that the proportions of the concrete noun context ($c_8$; *umbrella, pillow*) is larger than the proportions of the abstract noun context ($c_1$; *honor*).

## 3.3 Cosine Similarity of the Strongest Context Words

To examine the density and diversity of the context words, the cosine similarities between the target words and their five strongest context words were calculated (Section 2). By the strongest context words, the words with the highest co-occurrence frequency are meant. For the computation of the cosine similarity between the target words and their five strongest context words, vector representations of both target and context words were used. A vector of a word was represented by the co-occurrence frequencies of its context words. Hence, when computing the cosine similarity between the target word and a context word, the context words of that context word were also needed to construct the vectors for the context words. For example, the target word *miracle* is represented by the frequencies of six context words including the context word *religion* (example below). To compute the cosine similarity between *miracle* and *religion*, the frequencies of the context words of *religion* (= context words of context word) are used to represent *religion* as a vector. The context words of the context word were fetched from the ENCOW16AX corpus as well.

For calculating the cosine similarities, the word vectors have to have the same number of dimensions. To bring the vectors to the same length and number of dimensions, the context words of both words were unified. Context words that did not co-occur with either one of the (target or context) word for the cosine similarity calculation, had a co-occurrence frequency of 0. Below, the aforementioned example of the two word vectors $\overrightarrow{miracle}$ and $\overrightarrow{religion}$ are presented. In general, $\vec{t}$ represents one of the target words from the target word sets (in this example $\overrightarrow{miracle}$) and $\vec{c}$ is one of the context words of $\vec{t}$ with the highest co-occurrence frequency (in this example $\overrightarrow{religion}$), in this case 865.

Here is a list of context words after the unification of the context words of $\vec{t}$ and $\vec{c}$:
Context words $= \vec{t} \cup \vec{c} = \{religion, wisdom, war, crew, yard, training, stupidity, richness\}$

$$\vec{t} = \overrightarrow{miracle} = \begin{pmatrix} religion \\ wisdom \\ training \\ war \\ crew \\ yard \end{pmatrix} = \begin{pmatrix} 865 \\ 238 \\ 86 \\ 352 \\ 74 \\ 55 \end{pmatrix} \overset{\text{extend dimensions}}{\Rightarrow} \begin{pmatrix} religion \\ wisdom \\ training \\ war \\ crew \\ yard \\ stupidity \\ richness \end{pmatrix} = \begin{pmatrix} 865 \\ 238 \\ 86 \\ 352 \\ 74 \\ 55 \\ 0 \\ 0 \end{pmatrix}$$

$$\vec{c} = \overrightarrow{religion} \begin{pmatrix} religion \\ wisdom \\ training \\ stupidity \\ richness \end{pmatrix} = \begin{pmatrix} 82662 \\ 1225 \\ 676 \\ 212 \\ 95 \end{pmatrix} \overset{\text{extend dimensions}}{\Rightarrow} \begin{pmatrix} religion \\ wisdom \\ training \\ war \\ crew \\ yard \\ stupidity \\ richness \end{pmatrix} = \begin{pmatrix} 82662 \\ 1225 \\ 676 \\ 0 \\ 0 \\ 0 \\ 212 \\ 95 \end{pmatrix}$$

As seen in the example, the dimensions of the vectors were extended by adding the missing context words after the unification of all context words ($\vec{t} \cup \vec{c}$). The frequencies for the respective words are the same as before the extension. Thus, the co-occurrence frequency of the context word *wisdom* of $\vec{t}$, for instance, remains the same with 238. The missing context words did not co-occur with the words before the extension and therefore have a co-occurrence frequency of 0. For example, $\vec{t}$ was missing the context words *stupidity* and *richness* and were added with the co-occurrence frequencies 0. Similarly, $\vec{c}$ was missing *war, crew*, and *yard* and were added with the co-occurrence frequencies 0.

The cosine similarity was calculated with the following formula:

$$\text{(2)} \qquad cosine\_similarity(t, c) = \frac{\vec{t} * \vec{c}}{\|\vec{t}\| * \|\vec{c}\|}$$

Here, $\vec{t}$ and $\vec{c}$ represent the vector of the target word and context word, respectively.

In addition, to measure the overall similarities between the target words and their context words, the density of each target-(strongest) context words was determined by averaging the cosine similarities of the five strongest context words. The higher the density, the more related (similar) are target words and their context words.

Taking the example from above, the five strongest context words of *miracle* are *religion, war, wisdom, training*, and *crew* (highest co-occurrence frequencies). The cosine similarities are calculated as described earlier, resulting in the respective five cosine similarities and their density

$$cosine\_similarity(miracle, religion) = 0.70510$$
$$cosine\_similarity(miracle, war) = 0.70396$$
$$cosine\_similarity(miracle, wisdom) = 0.70264$$
$$cosine\_similarity(miracle, training) = 0.67150$$
$$cosine\_similarity(miracle, crew) = 0.64604$$

$$density = \frac{0.70510+0.70396+0.70264+0.67150+0.64604}{5} = \mathbf{0.68585}$$

The density shows that *miracle* and its five strongest context words are overall not extremely similar but related to an extend.

## 3.4 Word Classes

In this section, the two measures, mentioned in Section 2, were applied on manually chosen noun word classes (Section 3.1.4) to examine the differences and similarities between abstract and concrete word classes and the extremes. For measure (1), an extra step of comparing the distribution of the overall proportions of the extremes and the distribution of each word class was added. In measure (2), next to the average cosine similarities of the top five strongest context words (density), the standard deviation was calculated as well.

Both measures were applied on 20 word classes (hypernyms) in total - ten abstract and concrete word classes, respectively - that consist of at least five words (hyponyms). For abstract and concrete targets, the hyponyms of the respective word classes were examined.

### 3.4.1 Concreteness of Context for Word Classes

The context word and concreteness rating information was again retrieved from the ENCOW16AX corpus (Section 3.1.3) and Brysbaert et al. (2014) collection. The overall procedure is consistent with Section 3.2. Additionally, to compare the similarity in the distribution of the context word concreteness, the distance between the context distribution of the extremes and the word classes was computed. For example, the distance between the context distribution of the concrete noun target set and the word class *liquid*. For this, the proportions of the distribution of the abstract and concrete extremes were formed into a 24-dimensional vector, respectively. Similarly, the proportions of the distribution for each word class was vectorized. The dimensions emerge from the aforementioned eight concreteness categories for each of the three POS-tags noun, adjective and verb ($3 \times 8 = 24$). For the computation, the cosine similarity (2) was used. In total, for each word class, two distances - the distance to the distribution of the concrete noun extremes ($\overrightarrow{concrete}$) and equally, the distance to the distribution of the abstract noun extremes ($\overrightarrow{abstract}$) - were calculated. In the case of the word class *liquid*, for instance, the $cosine\_similarity(\overrightarrow{liquid}, \overrightarrow{concrete})$ and $cosine\_similarity(\overrightarrow{liquid}, \overrightarrow{abstract})$ were calculated.

### 3.4.2 Cosine Similarity of Strongest Context for Word Classes

The process of calculating the cosine similarity remains the same as in Section 3.3. Besides the cosine similarity and the densities, the standard deviation of each word class was computed based on all densities within a class. The standard deviation was calculated by calling the stdev() function from the statistics module in Python.

For the standard deviation $s$ the following formula was applied:

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} \tag{3}$$

$x_i$ signifies the individual density for a hyponym within a word class, $\bar{x}^2$ is the mean density of a word class, and $N$ represents the number of densities in a word class.

For example, the word class *luck* contains 5 hyponyms - *failure, fluke, misfortune, mishap, providence*. For each hyponym, the densities were calculated as described in Section 3.3, resulting into 5 density values ($= N$) for the word class *luck* (listed below).

Densities of *luck*:

$x_i = \{0.68726, 0.58944, 0.72317, 0.68318, 0.62678\}$

Mean density of *luck*:

$\bar{x} = \frac{0.68726+0.58944+0.72317+0.68318+0.62678}{5} \approx 0.66197$

Consequently, the standard deviation of *luck* is calculated as follows:

$$s = \sqrt{\frac{(0.68726-0.66197)^2+(0.58944-0.66197)^2+(0.72317-0.66197)^2+(0.68318-0.66197)^2+(0.62678-0.66197)^2}{5-1}} =$$

$$= \sqrt{\frac{0.0071595}{4}} \approx \mathbf{0.04231}$$

The standard deviation of *luck* is small, meaning that all densities are close to the mean density, indicating less disperse and consistent densities in the word class *luck*.

# 4 Results

In this section, the results are discussed. For each target set, bar plots were generated to depict the distribution of the concreteness of the context words. For the cosine similarity of the strongest context words, a couple of computation results of a few target words were sampled, followed by the results of both measures with respect to word classes are presented. For illustration purposes, a few example graphs are displayed for word classes. Further graphs are in the Appendix A.

## 4.1 Results for Concreteness of Context Words

Figure 1a, Figure 1c, and Figure 1e depict the distribution of the context words for the abstract target sets. Each figure shows a bar plot with the concreteness categories (in total 8) on the X-axis and the proportions on the Y-axis. The proportions for the noun, adjectives and verb contexts are indicated by blue, green and red bars, respectively.

For the **abstract noun targets**, the distribution of the adjective and verb context words tend to be more abstract with extremely few very concrete adjectives and verbs. The peak for adjectives ranges around 2.0 and 3.0, while the peak for verbs ranges from 1.5 to 2.5 being slightly more abstract. In contrast, the noun context seems to be more equally distributed except for the lower proportions for very abstract nouns. Compared to the adjectives and verbs, the nouns context comprises a fair amount of very concrete nouns.

The **abstract adjective targets** have a similar distribution to the abstract noun targets. The only difference is that the proportions of the noun context for abstract adjectives is minimally larger than for the abstract noun targets. Since the difference is quite small, it is not very significant.

The **abstract verb targets** are comparable to the abstract adjective target set.

The concreteness distribution of the context words of the concrete target sets are illustrated in Figure 1b, Figure 1d, and Figure 1f.

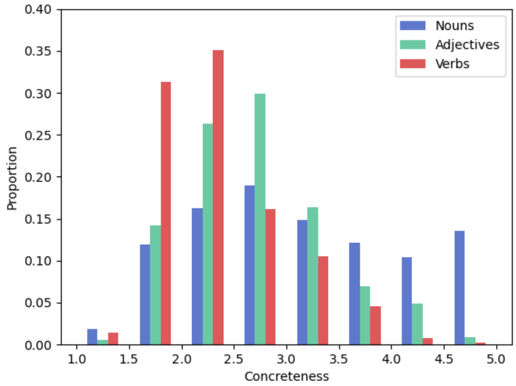
The proportions of noun context words of the **concrete noun target set** increases from abstract to concrete with a minor drop (almost plateau) between 3.5 and 4.5 and a large proportion of very concrete context words, indicated by the spiking blue bar. The adjective and verb context is similar to the adjective and verb context of abstract noun targets with the same respective peak ranges at 2.0 - 3.0 and 1.5 - 2.5. A minor increase for the verb context between 3.0 and 3.5 is observed, suggesting that unlike the abstract targets, concrete targets co-occur also with a considerable number of verbs with a mid-ranged concreteness.

The **concrete adjective and verb target sets** follow an almost identical pattern as the concrete noun targets with no major differences.

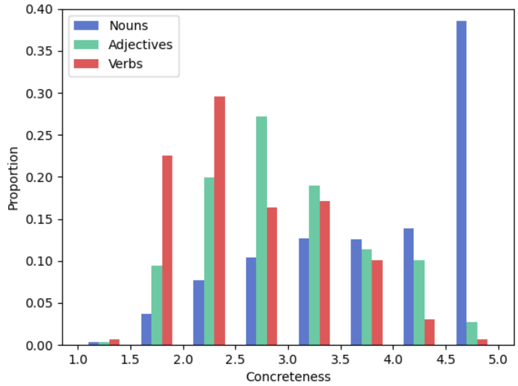
Overall, the general context distribution pattern is similar between abstract and concrete target words with low- to midranged concretenesses for adjective and verb context words and an equally distributed noun context except for the large portion of extremely concrete noun context words for concrete targets. All bar plots are more or less identical to the results of Naumann et al. (2018) with similar interpretations except for the noun context of abstract noun targets. While Naumann et al. (2018) interprets that the noun context shows the maximum peak at low concreteness scores, in this study, the noun context of abstract targets is interpreted as rather equally distributed.

Since the distribution within concrete/abstract target sets show similar patterns irrespective of the POS-tags, analyzing one part-of-speech class may suffice.

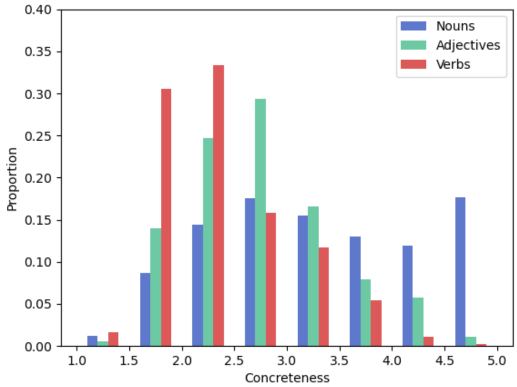
.



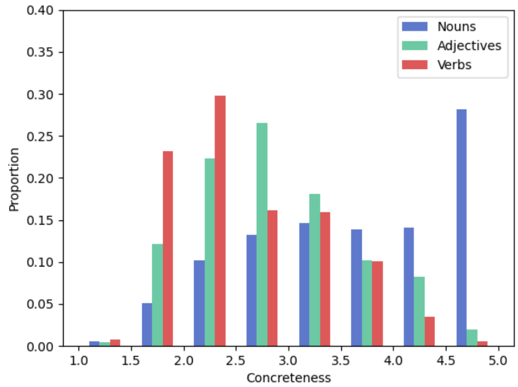

(a) Concreteness of abstract noun targets



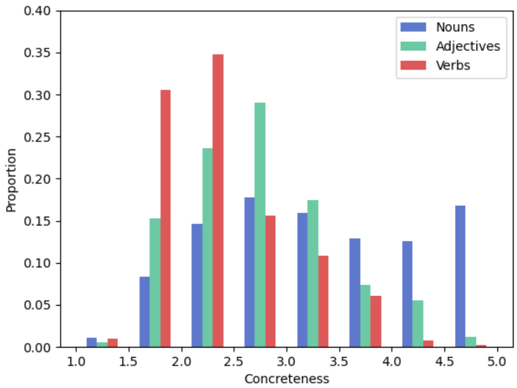

(b) Concreteness of concrete noun targets



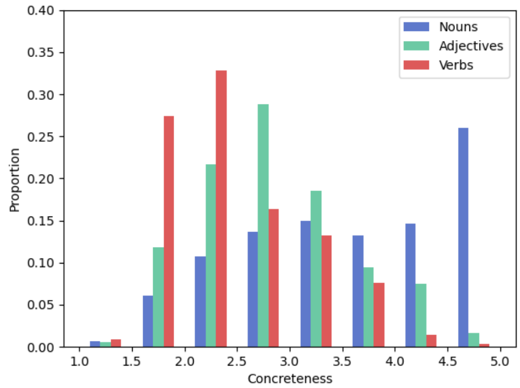

(c) Concreteness of abstract adjective targets




(d) Concreteness of concrete adjective targets




(e) Concreteness of abstract verb targets




(f) Concreteness of concrete verb targets

Figure 1: Concreteness of abstract and concrete targets

## 4.2 Results for the Cosine Similarity of the Strongest Context Words

As mentioned in Section 3.3, the context words with the highest co-occurring frequencies were considered as strongest context words. For some target words, the strongest context word was the target word itself. If this case occurred, the top 5 strongest context words that are unequal to the target word were counted.

Table 9, 10, and 11 each show four random abstract and concrete noun, adjective and verb target words and their five strongest context words with the cosine similarities. The context words of **noun targets** vary from specific words like *blanket* to general words, especially verbs like *make, get, use, take* etc. Furthermore, it is noticeable that the abstract targets tend to co-occur with comparatively more general words such as *life, man* and *world*, than the concrete target words (e.g *duvet*). Another observation is that concrete nouns tend to co-occur with a mixture of very related words like *toothbrush (target) - toothpaste (context)* and general words like *use (context)*.

For **adjective targets**, there is a difference between abstract and concrete adjectives. Abstract adjective targets are seen with general words like *ability* and *behavior*, while concrete adjective targets co-occur with physical objects such as *chicken, rice, shower* etc.

**Verb targets**, irrespective of abstract or concrete, appear with almost only commonly used words like *say, think, make, take* etc. In case of concrete verbs, a few related words like *unlock (target) - door (context)* are observed.

In general, the cosine similarities range from roughly 0.26 up to 0.92, suggesting that the strongest context words are not necessarily similar to their target words, and can be diverse. The average similarity of all five words (density), ranging from approximately 0.30 to 0.86, support this suggestion as well. For instance, the target words *toothbrush* and *cognitive* with densities of 0.55101 and 0.60883, show the co-occurrence with a diverse context that can be a mixture of specific (higher cosine similarity to target words) and general words (lower cosine similarity to target words).

| **Target set** | **Target words** | **Context words** | **Cosine Similarity** |
|---|---|---|---|
| NN abstract | generation | new | 0.65840 |
| | | young | 0.64704 |
| | | come | 0.61541 |
| | | make | 0.59250 |
| | | next | 0.56679 |
| | destiny | life | 0.73517 |
| | | world | 0.72997 |
| | | man | 0.72323 |
| | | take | 0.64663 |
| | | make | 0.62388 |
| NN concrete | toothbrush | toothpaste | 0.77317 |
| | | tooth | 0.56899 |
| | | use | 0.53515 |
| | | get | 0.47814 |
| | | electric | 0.39959 |
| | pillow | blanket | 0.73797 |
| | | sleep | 0.64223 |
| | | duvet | 0.63301 |
| | | get | 0.55501 |
| | | use | 0.52158 |

Table 9: Cosine similarity of the strongest context words for nouns

| Target set | Target words | Context words | Cosine Similarity |
|---|---|---|---|
| ADJ abstract | cognitive | ability | 0.66421 |
| | | therapy | 0.63701 |
| | | science | 0.60625 |
| | | development | 0.58836 |
| | | use | 0.54833 |
| | unethical | behavior | 0.67553 |
| | | say | 0.61387 |
| | | make | 0.61368 |
| | | practice | 0.61190 |
| | | illegal | 0.59869 |
| ADJ concrete | fried | chicken | 0.66421 |
| | | rice | 0.63701 |
| | | eat | 0.51795 |
| | | egg | 0.50083 |
| | | food | 0.48881 |
| | spotless | shower | 0.68628 |
| | | room | 0.43567 |
| | | hotel | 0.42170 |
| | | keep | 0.36057 |
| | | staff | 0.32983 |

Table 10: Cosine similarity of the strongest context words for adjectives

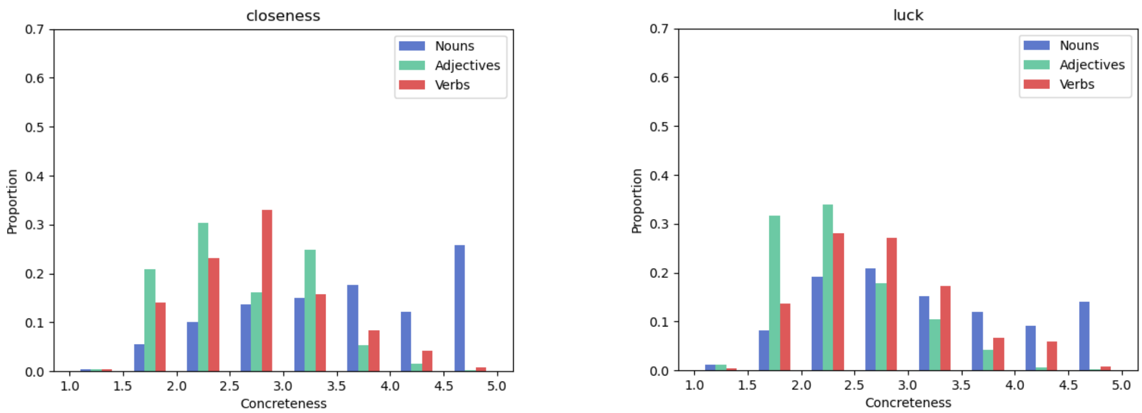
| Target set | Target words | Context words | Cosine Similarity |
|---|---|---|---|
| V abstract | offend | say | 0.81692 |
| | | think | 0.80839 |
| | | anyone | 0.78860 |
| | | get | 0.76174 |
| | | make | 0.75259 |
| | consider | take | 0.88003 |
| | | make | 0.87532 |
| | | need | 0.87185 |
| | | use | 0.84219 |
| | | say | 0.84212 |
| V concrete | unlock | lock | 0.87750 |
| | | door | 0.68235 |
| | | use | 0.61588 |
| | | new | 0.58322 |
| | | get | 0.58256 |
| | applaud | effort | 0.75149 |
| | | audience | 0.74889 |
| | | take | 0.74842 |
| | | say | 0.74704 |
| | | make | 0.74554 |

Table 11: Cosine similarity of the strongest context words for verbs

## 4.3 Results for Word Classes

### 4.3.1 Concreteness of Context for Word Classes

Almost all **abstract word classes** show a similar distribution pattern. As seen in Figure 3, the adjective and verb context climaxes in the lower- to mid-ranged concreteness, roughly between 1.5 - 2.5 and 2.5 - 3.5, respectively. While the proportion for highly concrete verbs is very low but existent, there is an unnoticeable or non-existent proportion for the highly concrete adjectives. For the noun context, it depends on the word class. For instance, the word class *closeness* in Figure 2a, seems to have a linear increase after 1.5, with a plunge at 4.0 - 4.5 in proportions, indicating a rather concrete noun context. On the contrary, Figure 2b, the word class *luck*, depicts a relatively equally distributed noun context except for the small proportion of extremely abstract nouns. Overall, each POS-tag shows a minimal proportion for extremely abstract context words. Majority of the word classes chosen in this study follow the distribution pattern of *luck*. Despite both word classes consisting of the same number of hyponyms (each 7), the distribution differed significantly for the noun context. This, as well as the similar graphs of other word classes (attached in the Appendix A) encompassing up to 43 hyponyms suggest that the size of word classes is irrelevant for abstract word classes.




(a) Abstract word class - *closeness*

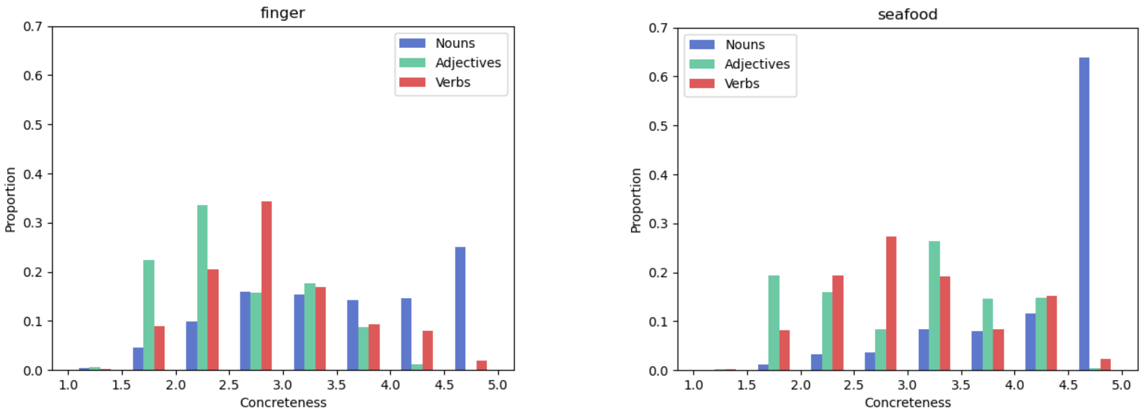
(b) Abstract word class - *luck*

Figure 2: Concreteness of abstract word class (examples)

Depending on the POS-tag, the **concrete word classes** present different patterns. The verb context seems to be the most consistent for all word classes, peaking at 2.5 - 3.0. Most of the word classes, like in Figure 3a, encompass the highest proportions for the adjective context between 1.5 and 2.5. A handful of the word classes such as *seafood* in Figure 3b, show a mid- to higher-ranged concreteness (roughly 3.0 - 3.5, sometimes until 4.0) for adjectives. The noun context has one commonality of a large proportion of extremely concrete nouns. For instance, although the blue bar at 4.5 - 5.0 for the word class *finger* is half the size of the blue bar of *seafood*, for both word classes, it is the highest proportion of noun context. The rest of the noun context is either distributed in the mid-ranged concreteness (e.g. *finger*, Figure 3a), or gradually increases by concreteness (e.g. *seafood*, Figure 3b).

Unlike the abstract word classes, the concrete classes *finger* and *seafood* varying in word class size (5 and 11, respectively), exhibit different distribution patterns, suggesting that the size is relevant. Nevertheless, when observing the sizes of all ten concrete word classes, several word classes with similar size (∼15) as *seafood* are more similar to the pattern of *finger*. Moreover, word classes with more than double the size (∼25) of *seafood* display almost identical graphs as *seafood*. Based on these observations, assumptions for concrete word classes could be that the choice of word classes affects more than the size.

The distances of **abstract word classes** to the concreteness distribution of the abstract and concrete noun target sets were above 0.9 for all word classes demonstrating that the distribution is highly similar. Despite of a few exceptions, the abstract word classes were more similar to the abstract noun targets than the concrete noun targets. For example, in case of the word class *closeness* the distance to the concrete noun targets was slightly closer with a cosine similarity of 0.96473 compared to 0.94669 to the abstract noun targets, while *luck* has a distance of 0.90142 to the concrete, and 0.99635 to the abstract noun targets. Conversely, *luck* revealed a distance of 0.90142 to the concrete noun targets and 0.99635 to the abstract noun targets.




(a) Concrete word class - *finger*

(b) Concrete word class - *seafood*

Figure 3: Concreteness of concrete word class (examples)

In opposition to that, the distances of the **concrete word classes** ranged between 0.59 and 0.99 with a considerable number of distances in the 0.7 range, suggesting that the similarity of the distribution varies with the concrete word class. All concrete word classes were more similar to the concrete noun extremes than the abstract noun extremes. For instance, the distribution similarity of the word class *finger* is 0.97378 to the concrete and 0.95803 to the abstract noun targets. *Seafood* reveals a higher discrepancy to abstract noun targets with 0.69422 compared to 0.91779 to concrete noun targets.

### 4.3.2 Cosine Similarity of Strongest Context Words for Word Classes

Several hyponyms of the word classes had less than five context words. To keep the densities representative, only the densities of hyponyms with at least three context words were considered for the standard deviation.

Table 12 illustrates examples of each abstract and concrete word classes with two hyponyms and their three to five strongest context words. The cosine similarities span a wide range from very low to high similarities (0.02 to 0.85). Depending on the hyponyms, the similarities of the context words are either in a closer or wider range. For example, in the word class *idea*, the hyponym *overestimation* co-occurs with context words with cosine similarities of 0.60 and 0.09, whereas the strongest context words of the hyponym *plan* range from 0.75 to 0.85. Regardless of concrete or abstract, the densities vary from 0.02 to 0.80, showing a diverse context. Majority of the word classes, both abstract and concrete, have a standard deviation of 0.12, signifying a small disparity. Very few word classes have a higher standard deviation due to hyponyms with many very disperse densities such as *drink* - 0.68676 and *pome* - 0.05513. On the other hand, couple of word classes, more concrete than abstract word classes, exhibit consistent densities indicated by a low standard deviation of ~0.05.

| **Word class (hypernyms)** | **Hyponyms** | **Context words** | **Cosine Similarity** |
|---|---|---|---|
| idea (abstract) | overestimation | underestimation | 0.60298 |
| | | lead | 0.14433 |
| | | result | 0.11352 |
| | | effect | 0.09864 |
| | | study | 0.09761 |
| | plan | new | 0.84879 |
| | | include | 0.80608 |
| | | business | 0.80603 |
| | | make | 0.77333 |
| | | say | 0.74986 |
| liquid (concrete) | drink | food | 0.84655 |
| | | soft | 0.66318 |
| | | take | 0.64948 |
| | | go | 0.63934 |
| | | get | 0.63523 |
| | pome | apple | 0.11653 |
| | | fruit | 0.02676 |
| | | stone | 0.02210 |

Table 12: Cosine similarity of the strongest context words for word classes

# 5 Discussion

The aim of this study was to investigate the behavior of context words with respect to abstract and concrete concepts. After analyzing the context of abstract and concrete target words (extremes) for English nouns, adjectives and verbs, abstract and concrete noun word classes were examined, both using the same measures.

For the target word sets, qualitative and quantitative assessments conclude that abstract as well as concrete targets tend to co-occur with rather abstract adjective and verb contexts, whereas noun contexts for abstract targets are mostly equally distributed over all concreteness categories, and for concrete targets, the number of very concrete noun contexts is quite high, suggesting a rather concrete noun context. Overall, while noun contexts indicate a difference between abstract and concrete, adjective and verb contexts are similar irrespective of abstract and concrete target words. Furthermore, the wide range of cosine similarities between target words and their five strongest context words hint that abstract words mostly co-occur with rather general and abstract words like *life*, whereas concrete targets are seen with a mixture of few very related concrete context words such as *toothpaste* and general verbs (e.g. *make, take* etc.). This and the overall densities suggest a high diversity for both concrete and abstract targets of all three POS-tags.

Word classes show varying results for the qualitative assessment of the concreteness of the context words depending on the word class. While the verb context pattern was consistent with the extremes over all word classes, the adjectives and nouns differed depending on the word class. Especially, interesting observations were made for the noun context. For abstract word classes, distinct noun context patterns were seen. While several abstract word classes followed the noun context pattern of the abstract extremes, a few were more similar to the concrete extremes. Akin to this, the noun context of concrete word classes depict either a similar or dissimilar distribution to the abstract or the concrete extremes with the exception of very concrete nouns always being similar to the concrete extremes, constituting the highest noun context proportion. A quantitative evaluation of the distances of the distribution of the word classes to the distribution of the abstract and concrete extremes supports the prior findings. The cosine similarities for abstract word classes were higher to concrete/abstract extremes depending on the word class, whilst the distance from the concrete word classes to the concrete extremes was always closer than to the abstract extremes.

Similarly to the extremes, the diversity of context words exist in word classes as well. The standard deviations also demonstrate moderate disparity.

To answer the previously mentioned research questions (Section 2), the word classes mirror the results of the extremes depending on the word class, suggesting that the choice of the word class is important. The size seems irrelevant for both abstract and concrete word classes since the results differed even with word classes of the same size. However, it is noteworthy that too small or too concrete word classes tend to have sparse data. Consequently, for reliable and informative results, larger word classes are more preferable.

# References

David A. Balota, Melvin J. Yap, Michael J. Cortese, Keith A. Hutchison, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson, and Rebecca Treiman. The English Lexicon Project. *Behavior Research Methods*, 39 (3):445–459, 2007.

Marianna Bolognesi, Christian Burgers, and Tommaso Caselli. On abstraction: Decoupling conceptual concreteness and categorical specificity. *Cognitive Processing*, 21(3):365–381, 2020. doi: 10.1007/s10339-020-00965-9.

Marc Brysbaert and Boris New. Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, 41 (4):977–90, 2009.

Marc Brysbaert, Boris New, and Emmanuel Keuleers. Adding Part-of-Speech Information to the SUBTLEX-US Word Frequencies. *Behavior Research Methods*, 44: 991–997, 2012.

Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. Concreteness Ratings for 40 Thousand generally known English Word Lemmas. *Behavior Research Methods*, 64:904–911, 2014.

L. Connell and D. Lynott. Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, 125:452–465, 2012.

Mark Davies. The 385+ million word corpus of contemporary American English: Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14(2):159–190, 2009.

Diego Frassinelli, Daniela Naumann, Jason Utt, and Sabine Schulte im Walde. Contextual Characteristics of Concrete and Abstract Words. In *Proceedings of the 12th International Conference on Computational Semantics (IWCS) — Short papers*, 2017.

Emmanuel Keuleers, Patrick Lacey, Kathleen Rastle, and Marc Brysbaert. The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, 44:287–304, 2012.

George Lakoff and Mark Johnson. *Metaphors We Live By*. University of Chicago Press, 1980.

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*, page 76–85, New Orleans, Louisiana, 2018. Association for Computational Linguistics.

Princeton University. Wordnet. https://wordnet.princeton.edu/. Accessed: January 26, 2024.

Roland Schäfer. Processing and Querying Large Web Corpora with the COW14 Architecture. In *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora*, pages 28–34, Mannheim, Germany, 2015.

Roland Schäfer and Frederik Bildhauer. Building Large Corpora from the Web Using a New Efficient Tool Chain. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 486–493, Istanbul, Turkey, 2012.

Sabine Schulte im Walde and Diego Frassinelli. Distributional Measures of Semantic Abstraction. *Frontiers in Artificial Intelligence*, 4, 2021. doi: 10.3389/frai.2021.796756. URL https://doi.org/10.3389/frai.2021.796756.

Paula J Schwanenflugel and Edward J Shoben. Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(1):82, 1983.

E. Strain, K. Patterson, and M. S. Seidenberg. Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(5):1140–1154, 1995. doi: 10.1037/0278-7393.21.5.1140.

# A Appendix

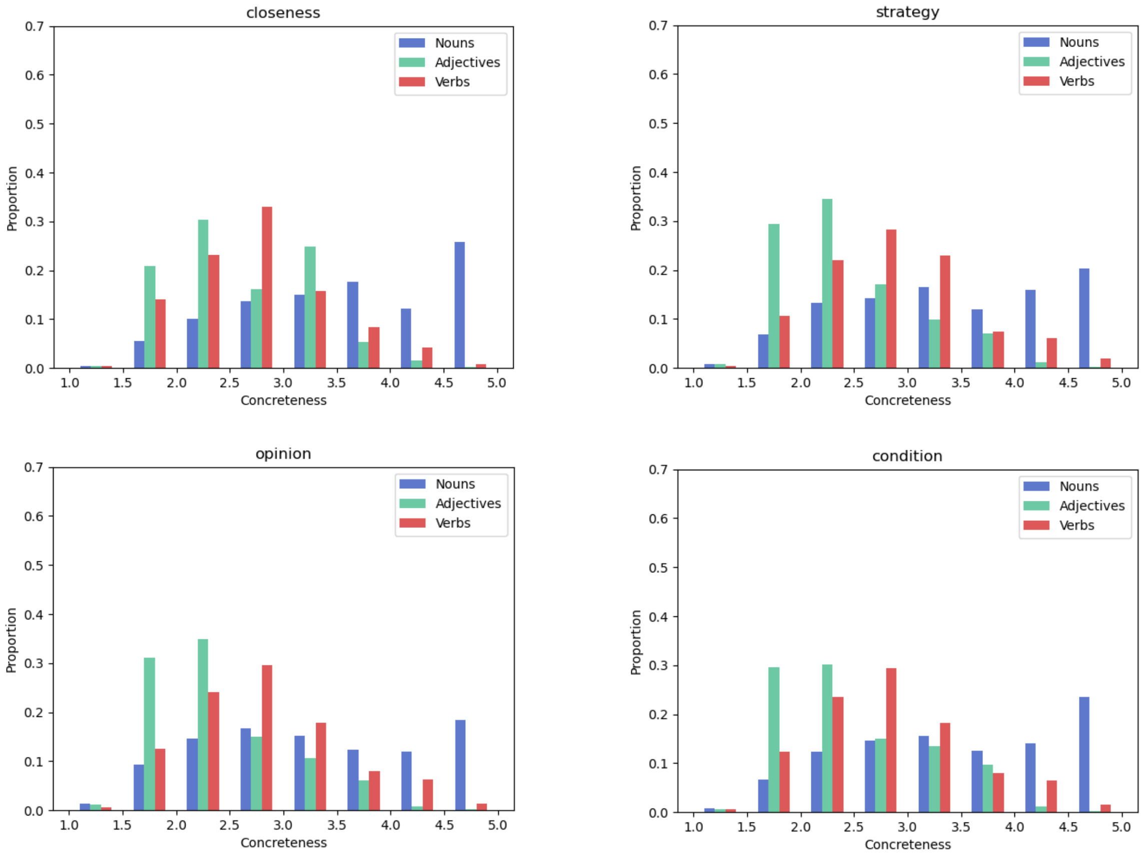



Figure 4: Concreteness of **abstract** word classes

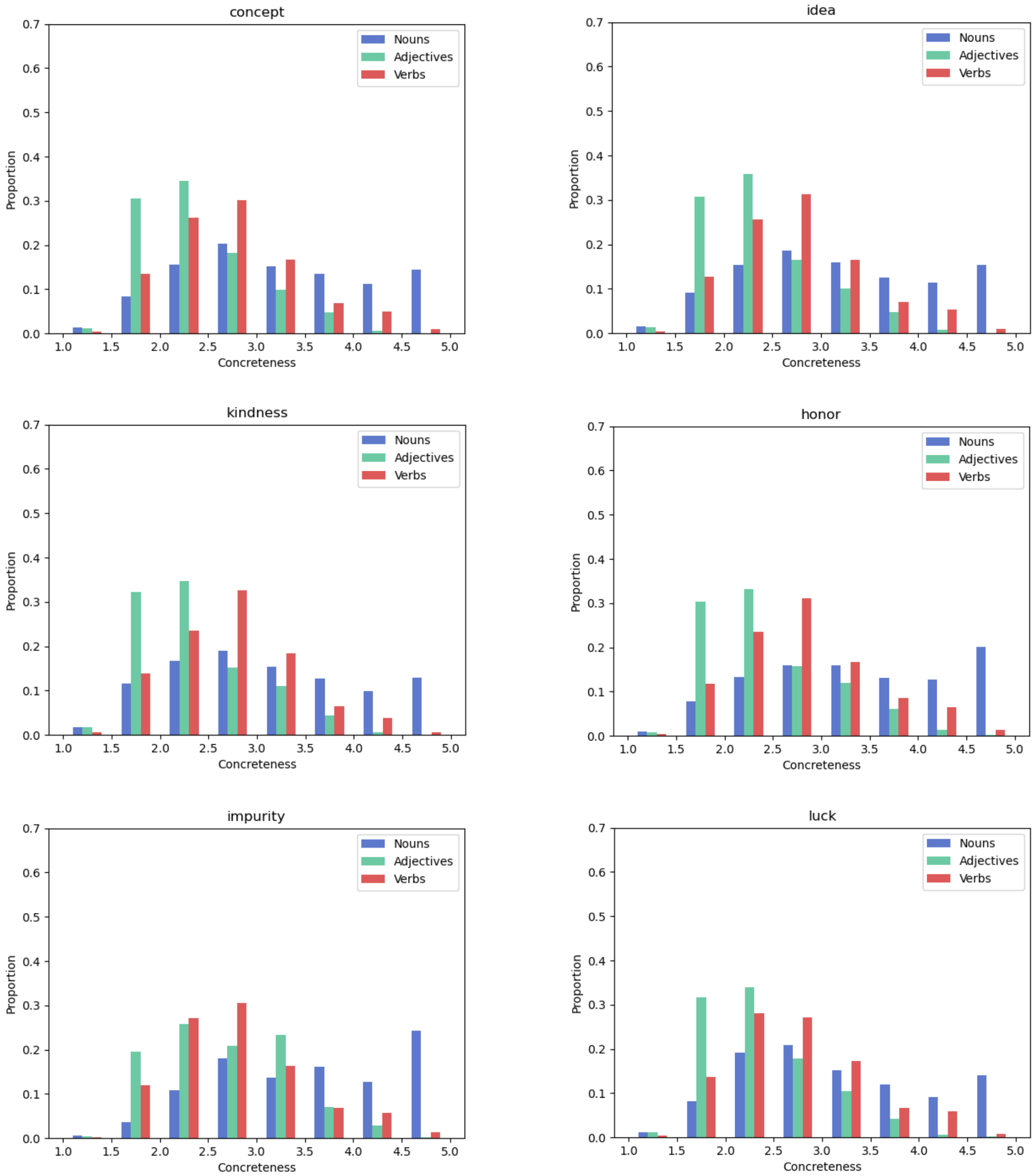



Figure 4: Concreteness of **abstract** word classes (Continued)

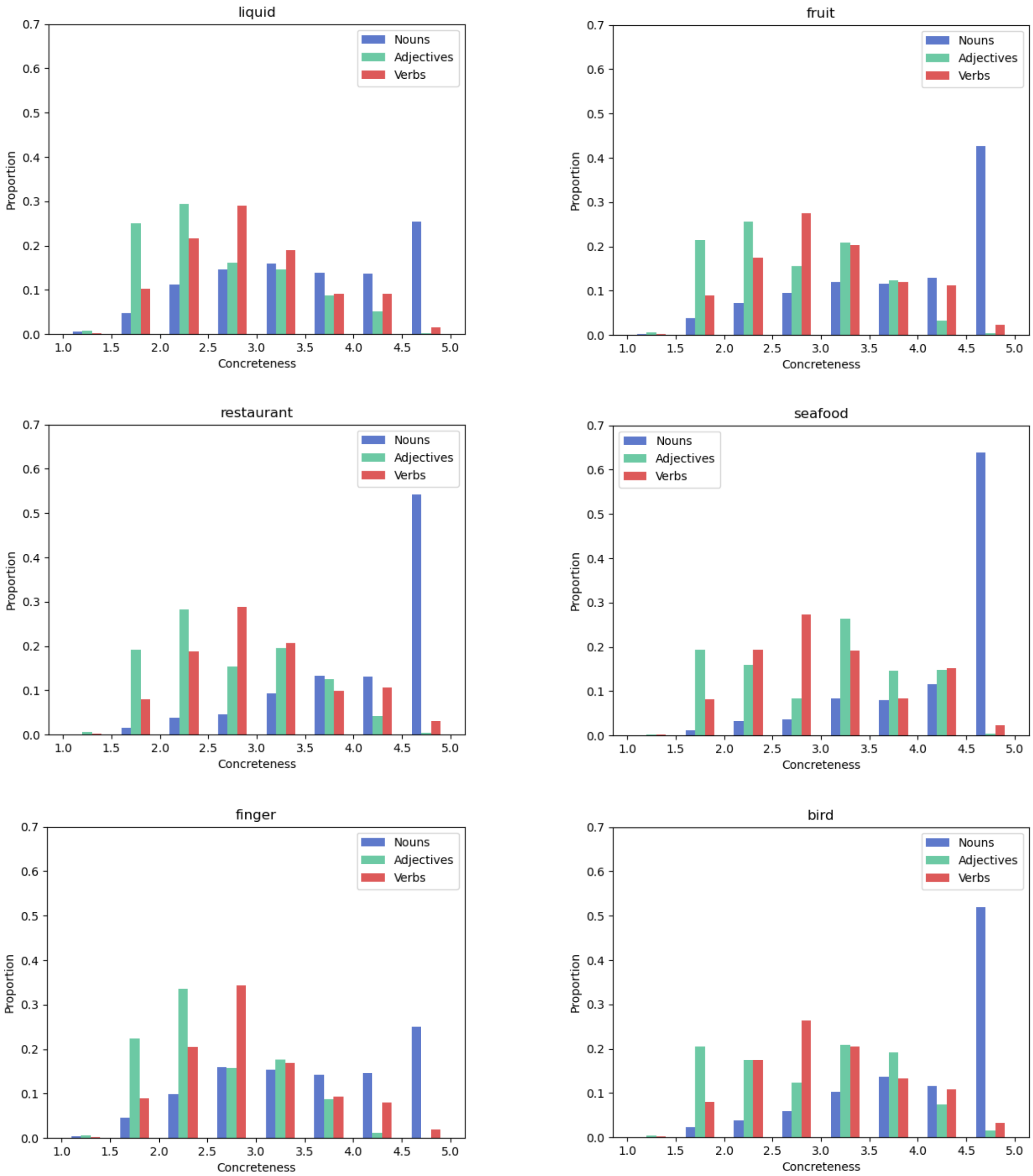



Figure 5: Concreteness of **concrete** word classes

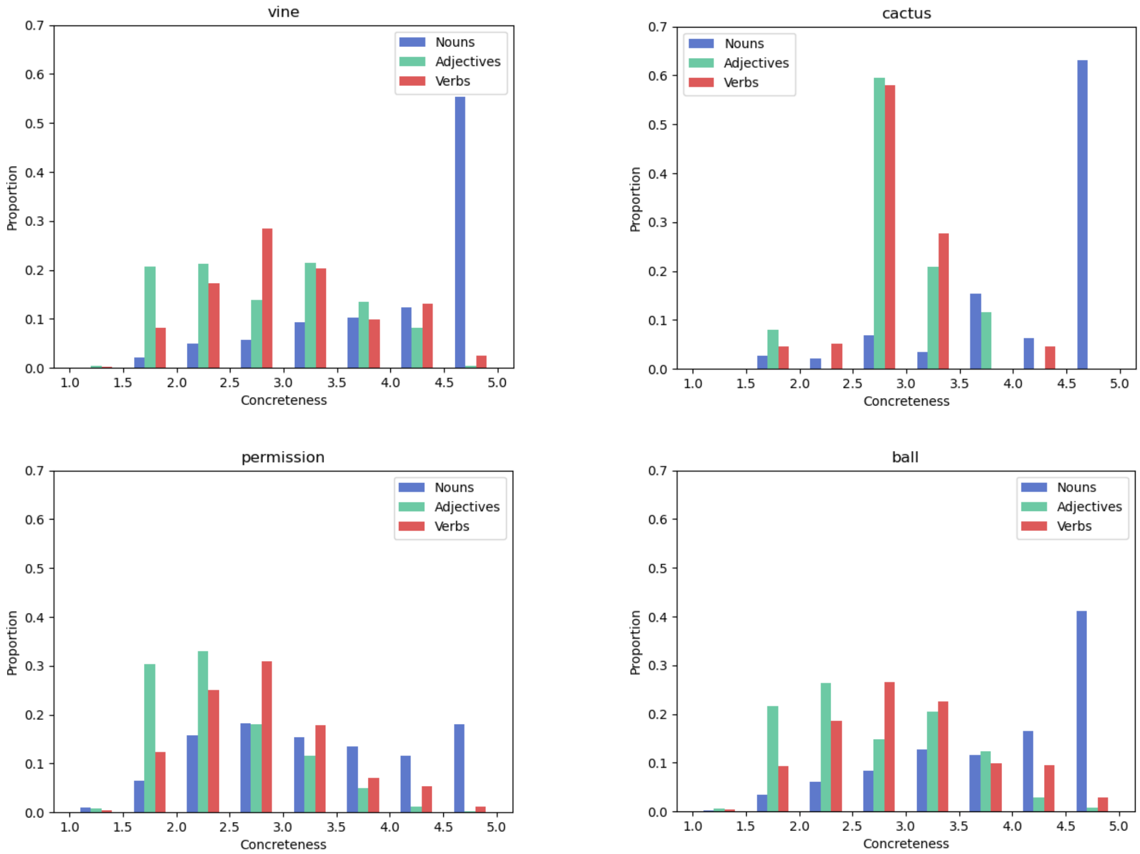



Figure 5: Concreteness of **concrete** word classes (Continued)

| Word class | Hyponyms |
|---|---|
| *closeness* | adjacency, contiguity, littleness, pettiness, proximity, smallness, togetherness |
| *strategy* | bubble, contrivance, counterterrorism, dodge, game, playbook, plot, stratagem, wheeze |
| *opinion* | conjecture, dictum, effect, guess, hunch, hypothesis, idea, intuition, judgment, mind, pole, politics, position, preconception, presence, side, speculation, supposition, surmise, suspicion |
| *condition* | destiny, fate, fortune, introversion, invagination, lot, luck, ordinary, orphanage, polyploidy, portion, stratification, transsexualism |
| *concept* | abstract, abstraction, attribute, category, conceptualization, dimension, division, fact, hypothesis, law, possibility, property, quantity, regulation, rule, section, whole |
| *idea* | belief, burden, concept, conception, construct, credit, feeling, figment, generality, generalization, guess, guesstimate, guesswork, guestimate, ideal, idealization, impression, inspiration, keynote, kink, meaning, misconception, motif, notion, opinion, overestimate, overestimation, plan, preoccupation, program, reaction, shot, statement, substance, suggestion, theme, theorem, underestimate, underestimation, variation, whim, whimsy |
| *kindness* | benefaction, benevolence, consideration, endearment, favor, forgiveness, generosity, generousness, pardon, thoughtfulness |
| *honor* | cachet, celebrity, citation, commendation, crown, decoration, degree, esteem, fame, glorification, glory, letter, medal, medallion, mention, oscar, palm, pennant, regard, reputation, respect, ribbon, seal, trophy |
| *impurity* | admixture, adulteration, alloy, contamination, debasement, pollution, taint |
| *luck* | failure, fluke, mischance, misfortune, mishap, providence, tossup |



Table 13: **Abstract** word classes

| **Word class** | **Hyponyms** |
|---|---|
| *liquid* | alcohol, ammonia, antifreeze, beverage, distillate, distillation, drink, elixir, ink, instillation, liquor, medium, potable, spill, tuberculin, water |
| *fruit* | achene, acorn, berry, capitulum, drupe, ear, gourd, hip, olive, pod, pome, rosehip, seed, spike |
| *restaurant* | bistro, brasserie, busboy, cafe, cafeteria, canteen, diner, grill, lunchroom, rotisserie, steakhouse, teahouse, tearoom, teashop |
| *seafood* | milt, octopus, periwinkle, prawn, roe, shellfish, shrimp, squid, whelk, whitefish, winkle |
| *finger* | forefinger, index, pinkie, pinky, thumb |
| *bird* | archaeopteryx, cock, hen, nester, parrot, passerine, raptor, ratite, trogon, twitterer, wildfowl |
| *vine* | bindweed, bittersweet, bougainvillea, briar, brier, clematis, climber, dodder, gourd, grape, grapevine, groundnut, hop, hoya, ivy, kiwi, kudzu, liana, potato, sarsaparilla, soma, squash, waxwork, wisteria, yam |
| *cactus* | cholla, mezcal, nopal, peyote, saguaro |
| *permission* | allowance, authority, authorization, clearance, consent, dismissal, dispensation, leave, pass, passport, sanction, toleration |
| *ball* | baseball, basketball, bolus, bowl, clew, clot, cobblers, cotillion, fireball, football, gob, handball, jack, marble, mothball, pellet, prom, promenade, racquetball, snowball, softball, spherule, volleyball |

Table 14: **Concrete** word classes

# B German Abstract

Abstraktion in Kontrast zu Konkretheit ist ein grundlegendes Thema in der Computerlinguistik und Psycholinguistik, insbesondere bei der Textmodellierung für die computationelle Semantik. Mehrere Studien befassten sich mit der Analyse der Konkretheit auf Wortebene, indem sie verschiedene Maße auf englische Nomen, Adjektive und Verben anwenden (Naumann et al., 2018). In dieser Studie wurden bestehende Maße (Naumann et al., 2018) sowohl auf einzelne Zielwörter (Nomen, Adjektive und Verben), als auch auf semantischen WordNet-Klassen (Nomen) angewendet. Die erste Analyse untersucht die Verteilung von Kontextwörtern in Bezug auf ihre Konkretheit, basierend auf von Menschen bewerteten Konkretheitswerten, die von 1,0 (sehr abstrakt) bis 5,0 (sehr konkret) reichen. Mit Hilfe der Kosinus-Ähnlichkeit werden dann die am häufigsten vorkommenden Kontextwörter untersucht. Die Ergebnisse zeigen, dass abstrakte Wörter tendenziell mit abstrakten Adjektiv- und Verbkontexten auftreten, während Nomenkontexte für abstrakte Wörter eher gleichmäßig über die Konkretheitskategorien verteilt sind. Umgekehrt ist bei konkreten Zielwörtern der Anteil sehr konkreter Nomenkontexte recht hoch, was auf einen überwiegend konkreten Nomenkontext schließen lässt. Außerdem, kookkurieren abstrakte Zielwörter meist mit eher allgemeinen und abstrakten Wörtern (*life*), während konkrete Wörter mit einer Mischung aus ein paar eng verwandten konkreten Kontextwörtern (*toothpaste*) und allgemeinen Verben (*make*, *take*) gesehen werden. Dies und die Gesamtdichten der Kosinus-Ähnlichkeiten deuten auf eine hohe Diversität sowohl für konkrete, als auch für abstrakte Nomen-, Adjektiv- und Verbzielwörtern hin. Der Vergleich zwischen den Wortklassen und den einzelnen Wörtern zeigt je nach Wortklasse entweder unähnliche oder ähnliche Ergebnisse, was darauf hindeutet, dass die Wahl der Wortklasse wichtig ist. Die Größe scheint sowohl für abstrakte, als auch für konkrete Wortklassen irrelevant zu sein, da die Ergebnisse auch bei Wortklassen gleicher Größe unterschiedlich ausfielen. Es ist jedoch zu beachten, dass zu kleine oder zu konkrete Wortklassen tendenziell zu spärlichen Daten führen. Um zuverlässige und aussagekräftige Ergebnisse zu erhalten, sind daher größere Wortklassen vorzuziehen.