

Concreteness vs. Abstractness: A Selectional Preference Perspective

Tarun Tater¹, Diego Frassinelli², Sabine Schulte im Walde¹

¹Institute for Natural Language Processing, University of Stuttgart,

²Department of Linguistics, University of Konstanz

{tarun.tater, schulte}@ims.uni-stuttgart.de
diego.frassinelli@uni-konstanz.de

Abstract

Concrete words refer to concepts that are strongly experienced through human senses (*banana, chair, salt*, etc.), whereas abstract concepts are less perceptually salient (*idea, glory, justice*, etc.). A clear definition of abstractness is crucial for the understanding of human cognitive processes and for the development of natural language applications such as figurative language detection. In this study, we investigate selectional preferences as a criterion to distinguish between concrete and abstract concepts and words: we hypothesise that abstract and concrete verbs and nouns differ regarding the semantic classes of their arguments. Our study uses a collection of 5,438 nouns and 1,275 verbs to exploit selectional preferences as a salient characteristic in classifying English abstract vs. concrete words, and in predicting their concreteness scores. We achieve an f1-score of 0.84 for nouns and 0.71 for verbs in classification, and Spearman’s ρ correlation of 0.86 for nouns and 0.59 for verbs.

1 Introduction

Concepts can be viewed in accordance with how humans perceive them. Those that are easily perceptible with any of the five senses are referred to as concrete concepts, whereas those that cannot be seen, heard, touched, smelled, or tasted as abstract concepts (Brysbaert et al., 2014). Examples of concrete concepts are *axe, cup, salt*, and *elephant*, whereas examples of abstract concepts are *belief, spirituality*, and *intuition*. Based on an analysis of noun concepts from the University of South Florida dataset (Nelson et al., 2004) and their occurrence in the British National Corpus (Leech et al., 1994), abstract words tend to be much more common in everyday usage (Hill et al., 2014).

The distinction between concrete and abstract concepts is quite important in linguistics, psycholinguistics, as well as computational linguistics. Furthermore, studies have shown that concreteness

measures are useful in a number of applications, such as lexicography (Kwong, 2011), document comprehensibility (Tanaka et al., 2013), and figurative language detection (Turney et al., 2011; Köper and Schulte im Walde, 2016; Aedmaa et al., 2018; Piccirilli and Schulte im Walde, 2022).

Theories of cognition contend that concrete and abstract words should co-occur most frequently with concrete words because concrete information connects the actual use of both concrete and abstract words to their mental representation (Barsalou, 1999; Pecher et al., 2011). However, previous corpus-based empirical studies do not show the same pattern. Bhaskar et al. (2017), Frassinelli et al. (2017), and Naumann et al. (2018) found that concrete words tend to co-occur with other concrete words, whereas abstract words tend to co-occur with other abstract words. Zooming into more specific co-occurrence conditions, Frassinelli and Schulte im Walde (2019) however demonstrated a more diverse empirical picture: they investigated interaction patterns of abstract and concrete English nouns and verbs in subcategorisation relations, and found that specific combinations indicated specific types of literal vs. figurative language usage, e.g., strongly associated abstract verbs subcategorising concrete direct objects often exhibited metonymy (e.g., *recommend a book*), while concrete verbs in the same relationship more often indicated literal language use (e.g., *write a book*).

In this study, we focus on selectional preferences as a way to investigate the inconsistencies between cognitive theories and empirical results reported above. Selectional preferences indicate the tendency that predicates impose semantic restrictions on the realisations of their complements, i.e., co-occurrence in a syntactic predicate-argument relationship (Resnik, 1993; Brockmann and Lapata, 2003; Erk et al., 2010; Schulte im Walde, 2010). For example, see sentences (1)–(3) with the verb *eat*, which requires an edible entity as direct object.

- (1) Amy is eating *chocolate*.
- (2) *Chris is eating *justice*.
- (3) Joe had to eat *dirt* for his earlier statement.

While the example in (1) is perfectly plausible, the example in (2) is not, because *justice* violates the selectional preferences of the governing predicate *eat*. Similarly, in (3) we see a violation that can only be resolved as a metaphorical reading.

Our study investigates whether selectional preferences represent a semantic criterion to establish empirical differences between the two semantic classes of abstract vs. concrete words. We thus suggest a more in-depth look into co-occurrence conditions in comparison to previous work that explored either window-based or purely syntactic co-occurrence. In this vein, we present two data-driven experiments focusing on (i) selectional preferences of English verbs regarding their subcategorisation of subjects and direct objects, and (ii) inverse selectional preferences of English nouns being subcategorised as subjects and direct objects. We use selectional preferences as features (a) in a binary classification task, to distinguish between more abstract vs. more concrete nouns/verbs, and (b) in a regression analysis, to predict the concreteness ratings of nouns and verbs.

2 Related Work

Frassinelli et al. (2017) quantitatively investigates differences between abstract and concrete words by analysing the abstractness of their respective context words. They showcase that concrete words tend to co-occur with other concrete words whereas abstract words co-occur with abstract words. Naumann et al. (2018) and Frassinelli and Schulte im Walde (2019) analyse the interactions of nouns and verbs in verb-noun subcategorisation by looking at types of syntactic relations between nouns and verbs (see above).

Another strand of research has exploited multi-modal approaches to infer concreteness. Hill et al. (2014) investigates which aspects of concreteness can be learned using multi-modal models and which are the most salient linguistic features contributing to it. Bhaskar et al. (2017) combine visual properties extracted from images and distributional representations built from textual data to distinguish between abstract and concrete words.

3 Dataset

For our research, we utilise the concreteness ratings for approximately 40,000 English words from Brysbaert et al. (2014) (henceforth, *Brysbaert norms*). The ratings were collected via crowdsourcing on Amazon Mechanical Turk. Each word was presented to at least 25 participants who were asked to rate the word on a scale from 1 – 5 where 1 indicates clearly abstract and 5 indicates clearly concrete concepts. The scores were then averaged across participants to obtain a mean concreteness rating for each word. The ratings were collected out-of-context and without providing any information about part-of-speech (POS). In a post-processing step, part-of-speech tags and frequencies were added to the target words, based on the SUBTLEX-US corpus (Brysbaert et al., 2012).

Following Schulte im Walde and Frassinelli (2022), we extracted and added frequency information based on the English web corpus ENCOW16AX¹ (Schäfer and Bildhauer, 2012; Schäfer, 2015), as well as the most frequent POS tag associated with each target word. In our final dataset, we only included targets where the POS provided in the original collection corresponded to the POS extracted from the ENCOW16AX corpus, the corpus that we use in our experiments. We also removed words for which their predominant POS tag does not represent at least 95% of all POS tags of the target, to reduce ambiguity, and all words with a frequency below 10,000, to remove infrequent words. After filtering, the resulting collection includes 5,438 noun targets and 1,275 verb targets.

4 Methods and Experiments

In the following, we present our two experiments exploiting selectional preferences to distinguish between degrees of abstractness. The selectional preference features for our verb and noun targets are induced from the ENCOW16AX corpus mentioned above, which contains 20 billion sentences and is syntactically parsed. We focus on two word-class interactions regarding our verb and noun targets.

- *Verb-Noun Interaction*: The verbs interplay with nouns in two ways: verb-object interaction and subject-verb interaction. We investigate these two scenarios in the following way:

¹<https://www.webcorpora.org/encow/>

- A root verb having a direct object (dobj) as a syntactic child. For example: *Filip baked a cake*. Here, the noun *cake* is a direct object argument of the verb *bake*.
 - A root verb with a syntactic child as a nominal subject (nsubj). For example: *The student is sleeping*. Here, the verb *sleep* takes the noun *student* as subject.
- *Noun-Verb Interaction*: We consider the inverse selectional preferences from the point of view of nouns (Erk et al., 2010), again as two sub-cases.
 - A nominal subject (nsubj) which is a singular noun (NN) whose syntactic parent is a root verb.
 - A direct object (dobj) which is a NN whose syntactic parent is a root verb.

We now discuss how selectional preference features for these two cases were computed and used.

4.1 Selectional Preference Features

For each of the above four sub-cases, we calculate the (inverse) selectional preference scores for each verb and each noun in three ways:

- (i) *Frequency-based*: number of times a noun represents an argument (subject/direct object, depending on the sub-case) of a particular verb.
- (ii) *Feature normalisation*: min-max normalisation of selectional preference frequencies in (i) by normalising the co-occurrences for a *particular noun across all verbs*.
- (iii) *Row normalisation*: min-max normalization of selectional preference frequencies in (i) by normalising the co-occurrences for a *particular verb across all nouns*.

In this way, we construct three variants of (inverse) selectional preference vectors for all our verb targets across all subject/object nouns, and for all our noun targets as subjects/objects across all sub-categorising verbs (i.e., the reverse syntactic dependency direction). These variants are assessed and compared against each other as well as against co-occurrence irrespective of any syntactic relationship (i.e., "just" co-occurrence within the same sentence context, because previous studies looked at any co-occurring words), for each of the above-mentioned sub-cases, and in two experimental setups.

4.2 Binary Classification

In this first set of experiments, we classify both the 5, 438 nouns and the 1, 275 verbs into abstract vs. concrete words. Since the concreteness ratings range from 1 – 5, we treat words with ratings ≤ 3 as abstract and those with ratings > 3 as concrete. The resulting two classes are henceforth referred to as *Complete* set.

Given that mid-range concreteness scores are generally more difficult in their generation by humans and consequently noisier in their distributional representations (Pollock, 2018; Schulte im Walde and Frassinelli, 2022), we additionally construct the following variants of our target sets.

- We exclude target words that have concreteness scores between 2.5 and 3.5. These words can be difficult to classify because they are neither clearly abstract nor clearly concrete. After excluding these ‘neutral’/‘mid-scale’ words we have 4, 061 nouns (2, 757 concrete and 1, 304 abstract), and 769 verbs (118 concrete and 653 abstract). We call this set the *Extremes* set.
- We exclude target words with a standard deviation > 1.3 because in these cases annotators strongly disagreed. We refer to the set of words excluding these ‘disagreed’ words as *Agreed* set, containing 3, 456 nouns and 766 verbs.

The distribution of the *Brysbaert* norms for nouns is skewed heavily towards high scores (concrete) and, on the contrary, for verbs towards low scores (abstract). For example: the most concrete 1, 000 nouns can be found in the interval 4.86 – 5.00 whereas the most abstract 1, 000 nouns range from 1.00 to 1.92. So, instead of considering the extreme 1, 000 abstract and 1, 000 concrete nouns or 500 concrete and 500 abstract verbs, as done in some of our previous studies (Bhaskar et al., 2017; Naumann et al., 2018; Schulte im Walde and Frassinelli, 2022), we investigate how words in different *binned* ranges of concreteness ratings differ. To do this, we binary classify target words that have scores in the range of 1 – 2 against words with scores 2 – 3, 3 – 4, and 4 – 5. In this way, we manage to overcome the skewness in the distributions albeit with a trade-off for class imbalance. The binary classification between words having ratings 1 – 2 vs. 4 – 5 is similar to classifying only the most abstract and concrete words.

Datasets		Train			Test		
		Total	Abstract	Concrete	Total	Abstract	Concrete
Nouns	All	4,350	1,628	2,722	1,088	407	681
	Extremes	3,248	1,043	2,205	813	261	552
	Agreed	2,764	851	1,913	692	213	479
Verbs	All	1,020	774	246	255	194	61
	Extremes	616	522	94	155	131	24
	Agreed	572	463	109	144	116	28

Table 1: Data split 80 : 20 across experiments.

Targets & Selectional Preferences		Accuracy	Precision	Recall	F1-score
Verbs	Subject	0.80	0.75	0.63	0.65
	Direct Object	0.77	0.70	0.72	0.71
	Co-occurrence	0.77	0.78	0.77	0.77
Nouns	Subject (inverse)	0.84	0.83	0.84	0.83
	Direct Object (inverse)	0.85	0.84	0.84	0.84
	Co-occurrence	0.87	0.86	0.87	0.87

Table 2: Evaluation of binary classifications using SVMs with row-normalised features.

In the binary experiments we use three different classifiers: Support Vector Machines (SVMs) with *rbf* kernel, Random Forests and Logistic Regression. The binary classification is evaluated using accuracy, precision, recall, and f1-score to address the data skewness between classes. We use an 80:20 data split between train and test set using stratified sampling for our experiments, see Table 1. We also perform a hyper-parameter search optimising the parameters.

4.3 Regression: Predicting Concreteness Ratings

This task pertains to predicting the concreteness ratings from 1 – 5. We use Gradient Boosting to predict the concreteness scores of 5,438 nouns and 1,275 verbs. The predicted concreteness ratings are evaluated using Spearman’s rank-order correlation coefficient ρ against the average human ratings from the *Brysbaert* norms.

5 Results and Discussion

Table 2 reports the accuracy, precision, recall and f1-score results for our binary classifications across subject and direct object selectional preference conditions in comparison to simple co-occurrences. Using SVM with row-normalised features and the regularization parameter $C = 5$ for both the

verb-noun and the noun-verb settings,² the best f1-score results are achieved when relying on co-occurrences (0.87 for noun targets and 0.77 for verb targets), while selectional preference features reached 0.84 for nouns and 0.71 for verbs when relying on selectional preferences for direct objects, and 0.83 for nouns and 0.65 for verbs when relying on selectional preferences for subjects.

Figure 1 shows accuracy scores of the binary classification for nominal subjects (left) and direct objects (right) across our *binned* ranges of concreteness ratings, i.e. classifications between words in the concreteness ranges 1 – 2 vs. 2 – 3, 3 – 4, and 4 – 5. Unsurprisingly, accuracy increases with stronger differences between the ratings of the two classes. We also indicate the results for binary classification of the *Complete* sets (red dotted lines, also see accuracies in Table 2), and results for distinguishing the *Extremes* sets (green lines), which are similar as for distinguishing between bins 1 – 2 and 4 – 5, as expected.

Table 3 shows the results for our regression experiments, which are more difficult because they target the whole range of scores. We report best Spearman’s ρ correlations of 0.865 for predicting noun scores, and 0.596 for predicting verb scores. In these experiments, the best results are reached

²Results obtained with Logistic Regression and Random Forest classification models are comparable.

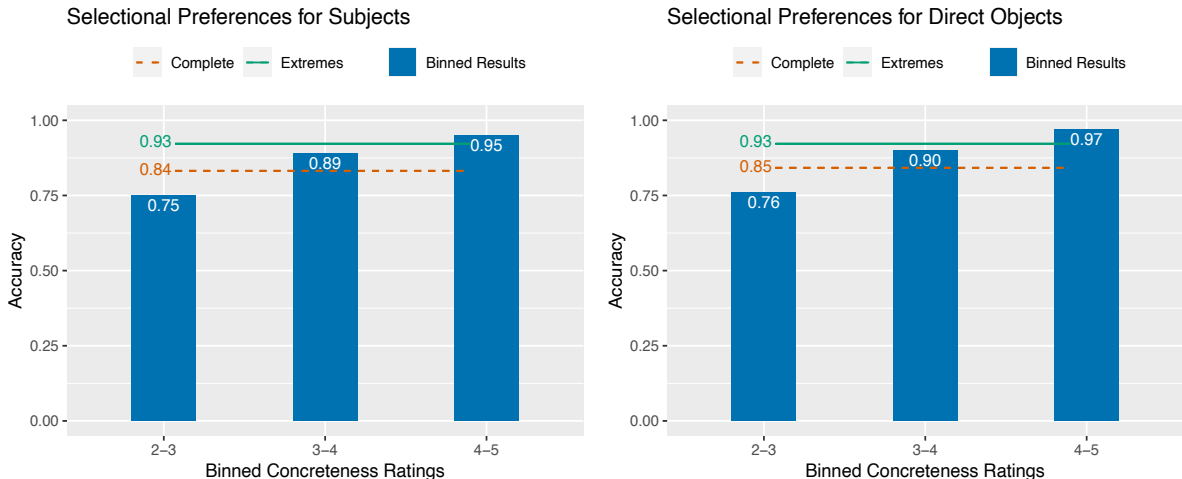


Figure 1: Classifying abstract vs. concrete nouns based on selectional preference features.

Targets & Selectional Preferences		Freq-Based	Feature Norm.	Row Norm.
Verbs	Subject	0.479	0.479	0.520
	Direct Object	0.548	0.552	0.596
	Co-occurrence	0.424	0.462	0.473
Nouns	Subject (inverse)	0.795	0.795	0.809
	Direct Object (inverse)	0.851	0.858	0.865
	Co-occurrence	0.822	0.822	0.861

Table 3: Spearman’s ρ correlations across regression experiments.

when using direct object selectional preferences, outperforming both subject selectional preference features and co-occurrences in all conditions, with various difference strengths for feature-based and normalisation variants. Between feature-based and normalisation variants we do not observe strong differences. The reported best results relying on direct object selectional preferences are obtained with the following hyper-parameters for verb targets: 200 trees, with a depth of 3 and learning rate of 0.05, and for noun targets: 200 trees, with a depth of 7 and learning rate of 0.05.

Across binary and regression experiments and experiment settings, the obtained results are better for noun targets than for verb targets, which is in line with our previous work (Schulte im Walde and Frassinelli, 2022). On the one hand, we hypothesise that this is due to the smaller number of data points and higher data skewness for verbs in comparison to nouns, as depicted in the data split in Table 1; on the other hand, we assume that verbs are semantically more difficult to distinguish regard-

ing any meaning aspects, because they are more ambiguous (which is presumably also reflected in their concreteness ratings).

Comparing selectional preference features relying on subjects vs. direct objects, we consistently observe that selectional preferences across direct objects provide more salient features for distinguishing between abstract and concrete nouns and verbs than subjects do.

In comparison to previous work, our Spearman’s ρ correlations for predicted noun ratings (0.865) and direct objects selectional preference features are comparable to Bhaskar et al. (2017), which shows a Spearman’s ρ correlation of 0.86 for 9, 241 nouns and 0.78 for the extreme 2, 000 nouns. However, their best-performing models utilise both textual embeddings as well as image embedding. Our results are able to achieve similar performance on our 4, 538 nouns with only textual selectional preference features.

6 Conclusion

In this study, we explored the use of selectional preferences as a linguistically more specific semantic criterion than purely sentential co-occurrences, when establishing empirical differences between the two semantic classes of abstract vs. concrete English verbs and nouns. Within a set of binary classification experiments varying selectional preference features, normalisations, classifiers, and more or less extreme differences in concreteness scores of the words in the classes, simple co-occurrence generally outperformed the semantically more fine-grained selectional preferences; in contrast, selectional preferences for direct objects improved over subject preferences and co-occurrences when used in the more fine-grained concreteness predictions of regression models. So overall, the more fine-grained semantic features are helpful in the more fine-grained perception-based semantic distinctions, and the core information in these combinations are verb-object semantic subcategorisations.

Acknowledgements

This research was supported by the DFG Research Grant SCHU 2580/4-1 *Multimodal Dimensions and Computational Applications of Abstractness*.

References

- Eleri Aedmaa, Maximilian Köper, and Sabine Schulte im Walde. 2018. Combining Abstractness and Language-specific Theoretical Indicators for Detecting Non-Literal Usage of Estonian Particle Verbs. In *Proceedings of the NAACL 2018 Student Research Workshop*, pages 9–16, New Orleans, LA, USA.
- Lawrence W. Barsalou. 1999. Perceptual Symbol Systems. *Behavioral and Brain Sciences*, 22:577–660.
- Sai Abishek Bhaskar, Maximilian Köper, Sabine Schulte im Walde, and Diego Frassinelli. 2017. 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*, Montpellier, France.
- Carsten Brockmann and Mirella Lapata. 2003. Evaluating and Combining Approaches to Selectional Preference Acquisition. In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, pages 27–34, Budapest, Hungary.
- Marc Brysbaert, Boris New, and Emmanuel Keuleers. 2012. Adding Part-of-Speech Information to the SUBTLEX-US Word Frequencies. *Behavior Research Methods*, 44:991–997.
- Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness Ratings for 40 Thousand generally known English Word Lemmas. *Behavior Research Methods*, 64:904–911.
- Katrin Erk, Sebastian Padó, and Ulrike Padó. 2010. A Flexible, Corpus-Driven Model of Regular and Inverse Selectional Preferences. *Computational Linguistics*, 36(4):723–763.
- Diego Frassinelli, Daniela Naumann, Jason Utt, and Sabine Schulte im Walde. 2017. Contextual Characteristics of Concrete and Abstract Words. In *Proceedings of the 12th International Conference on Computational Semantics*, Montpellier, France.
- Diego Frassinelli and Sabine Schulte im Walde. 2019. Distributional Interaction of Concreteness and Abstractness in Verb–Noun Subcategorisation. In *Proceedings of the 13th International Conference on Computational Semantics*, pages 38–43, Gothenburg, Sweden.
- Felix Hill, Roi Reichart, and Anna Korhonen. 2014. Multi-Modal Models for Concrete and Abstract Concept Meaning. *Transactions of the Association for Computational Linguistics*, 2(1):285–296.
- Maximilian Köper and Sabine Schulte im Walde. 2016. Distinguishing Literal and Non-Literal Usage of German Particle Verbs. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 353–362, San Diego, CA, USA.
- Oi Yee Kwong. 2011. Measuring Concept Concreteness from the Lexicographic Perspective. In *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation*, pages 60–69.
- Geoffrey Leech, Roger Garside, and Michael Bryant. 1994. Claws4: the tagging of the British National Corpus. In *COLING 1994 Volume 1: The 15th International Conference on Computational Linguistics*.
- Daniela Naumann, Diego Frassinelli, and Sabine Schulte im Walde. 2018. Quantitative Semantic Variation in the Contexts of Concrete and Abstract Words. In *Proceedings of the 7th Joint Conference on Lexical and Computational Semantics*, pages 76–85, New Orleans, LA, USA.
- Douglas L. Nelson, Cathy L. McEvoy, and Thomas A. Schreiber. 2004. The University of South Florida Free Association, Rhyme, and Word Fragment Norms. *Behavior Research Methods, Instruments, & Computers*, 36(3):402–407.
- Diane Pecher, Inge Boot, and Saskia Van Dantzig. 2011. Abstract Concepts. Sensory-Motor Grounding, Metaphors, and Beyond. *Psychology of Learning*

and Motivation – Advances in Research and Theory, 54:217–248.

- Prisca Piccirilli and Sabine Schulte im Walde. 2022. Features of Perceived Metaphoricity on the Discourse Level: Abstractness and Emotionality. In *Proceedings of the 13th International Conference on Language Resources and Evaluation*, Marseille, France.
- Lewis Pollock. 2018. Statistical and Methodological Problems with Concreteness and other Semantic Variables: A List Memory Experiment Case Study. *Behavior Research Methods*, 50:1198–1216.
- Philip Resnik. 1993. *Selection and Information: A Class-Based Approach to Lexical Relationships*. Ph.D. thesis, University of Pennsylvania.
- Roland Schäfer. 2015. 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.
- Roland Schäfer and Felix Bildhauer. 2012. 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.
- Sabine Schulte im Walde. 2010. Comparing Computational Approaches to Selectional Preferences: Second-Order Co-Occurrence vs. Latent Semantic Clusters. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 1381–1388, Valletta, Malta.
- Sabine Schulte im Walde and Diego Frassinelli. 2022. Distributional Measures of Abstraction. *Frontiers in Artificial Intelligence: Language and Computation*. Alessandro Lenci and Sebastian Pado (topic editors): "Perspectives for Natural Language Processing between AI, Linguistics and Cognitive Science".
- Shinya Tanaka, Adam Jatowt, Makoto P. Kato, and Katsumi Tanaka. 2013. Estimating Content Concreteness for Finding Comprehensible Documents. In *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, Rome, Italy.
- Peter D. Turney, Yair Neuman, Dan Assaf, and Yohai Cohen. 2011. Literal and Metaphorical Sense Identification through Concrete and Abstract Context. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 680–690, Edinburgh, UK.