Mapping Distributional Semantics to Formal Concept Lattice-based Property Norms

Dandan Li and Douglas Summers-Stay

June 10, 2019

Mapping Distributional Semantics to Formal Concept Lattice-based Property Norms

Dandan Li, Douglas Summers-Stay
U.S. Army Research Laboratory
happydandan2016@gmail.com, douglas.a.summers-stay.civ@mail.mil

Abstract

Distributional models characterize the meaning of a word by its observed contexts. They have shown great success in many natural language processing tasks, however they are unable to differentiate clearly between different semantic relations. In cognitive psychology, a word is represented by its relations with properties. In this work, we propose that the mathematical structure of formal concept lattice (FCL) can be attached to property-based concepts in the property norm space to model the conceptual hierarchies. The k-nearest neighbors (KNN) method is then used to build a mapping from a distributional semantic space onto a FCL-based property space automatically for predicting property norms of unknown concepts. We evaluate our method on word embeddings learned with different types of contexts and demonstrate the potential of learning large-scale property-based concept representations from a modest-sized human-annotated perceptual data.

1 Introduction

Semantic word representation plays important roles in a broad range of natural language processing (NLP) tasks, such as query expansion, machine translation, information extraction and question answering. Previous work addressing the problem can be roughly classified into two categories: (1) distributional semantic models learned with different types of contexts from large collections of text using variants of neural networks (Mikolov et al. 2013a; Mikolov et al. 2013b; Levy and Goldberg 2014); (2) property-based representation in terms of constituent properties generated by participants in property norming studies (McRae et al. 2005; Devereux et al. 2014), or extracted from manually-curated knowledge bases,such as FreeBase and Wikidata (Gupta et al. 2015).

Distributional semantic models characterize the meaning of a word through the contexts in which it appears. Word2vec (W2V) is one of the most popular word embedding methods that learn word vectors from raw text by setting different sizes of context window. Levy and Goldberg (2014) propose the dependency-based word embeddings (DEP), which generalize the skip-gram model with negative sampling, and can deal with syntactic contexts that are derived from automatically produced dependency parse-trees. All these models depend in some way or another on the distributional hypothesis which states that words that occur in similar contexts tend to have similar meanings (Harris 1954; Firth 1957). The empirical evidence shows that distributional models can do a good job in capturing word similarities. However, the basis vectors of distributional models tend to be uninterpretable, unlike property-based representations where each bit encodes the presence or absence of a particular property for that concept. For example, distributional models can tell us that *motorcycle* is similar to *motorbike* and *mudguard* with different similarity scores, but it is difficult to differentiate how *motorcycle* is related to *motorbike* from how it is related to *mudguard* based on these models. This is one of the main drawbacks of distributional models (Murphy 2002).

There is a wide consensus in cognitive psychology that the meaning of a concept is a complex assembly of properties that characterize how they are related to the concept. For example, the concept *ambulance* can be represented by properties like `has_a_siren`, `has_flashing_lights`, `is_fast`, and `used_by_hospitals`. There are several ways to obtain a comprehensive set of concept properties. The widely used property norm dataset is from (McRae et al. 2005), which consists of 541 concepts and 2526 properties. However, it is expensive and time-consuming to produce property norms by human annotations. Moreover, it is still unrealistic to extract accurate properties from a large-scale text corpus (Devereux et al. 2009). This raises the question of how we predict property norms for new concepts.

We detail our main contributions as follows. (1) In this paper, we propose that the mathematical structure of formal concept lattice (FCL) can be attached to property-based concepts in the property norm space. (2) We predict properties for new concepts by learning a mapping from a distributional semantic space to FCL-based property space. (3) We evaluate our method on word embeddings learned with different types of contexts and show the effectiveness of our method.

2 Related Work

Several previous works addressed property inference from distributional data. Baroni and Lenci (2008) explore the capability of producing property-based descriptions of concepts from computational models which are derived from word co-occurrence statistics. Strudel (Baroni et al. 2010) is an unsupervised algorithm to extract a structural and comprehensive set of concept descriptions directly from an English corpus and then represent concepts by weighted properties. Herbelot and Vecchi (2015) present an approach to automatically map a standard distributional semantic space onto a set-theoretic model. Făgărăşan, Vecchi, and Clark (2015) explore the possibility of generalizing property-based representation to a large scale dataset. The method they used was based on partial least square regression (PLSR). Dernoncourt (2016) introduces three methods: random vectors, mode and nearest neighbor to build the mapping between the two spaces. Erk (2016) proposes a probabilistic mechanism for distributional property inference. Boleda and Herbelot (2017) present an overarching semantic framework called "formal distributional semantics" which combines formal and distributional semantics together. There are also some papers that focus on formal concept lattice-based knowledge representation. Priss (1998) describes WordNet's hierarchical and relational structure in the form of a mathematical lattice. Ikeda and Yamamoto (2017) solves the problem of extending various thesauri by finding synonym sets from a formal concept lattice.

3 A Formal Concept Lattice-based Property Norms

3.1 Mathematical Foundations

Formal concept analysis was first proposed by Rudolf Wille in 1980s (Ganter and Wille 1999; Davey and Priestley 1990; Grätzer 1998). It has two key components: the generation of concepts from data and the identification of the inherent structure of these concepts. The formal definitions of formal concept analysis are given below (Ganter and Wille 1999) .

Definition 1 A **formal context** is a triple (G, M, I), where G and M are sets and I is a binary relation between the two sets. The elements $g \in G$ are called the **objects** of the context and the elements $m \in M$ are called the **attributes** of the context. If the object g has the attribute m, the relationship will be written as $(g, m) \in I$ or gIm.

Definition 2 Let (G, M, I) be a formal context. A **formal concept** of (G, M, I) is a pair (X, Y), where $X \subseteq G$, $Y \subseteq M$, $X = \{g \in G | gIm, \forall m \in Y\}$ and $Y = \{m \in M | gIm, \forall g \in X\}$. We call X and Y the **extent** and the **intent** of the concept (X, Y) respectively. The set of all concepts of the context (G, M, I) is denoted as $\mathcal{C}(G, M, I)$.

Definition 3 Let (G, M, I) be a formal context and let (X_a, Y_a) and (X_b, Y_b) be two formal concepts in $\mathcal{C}(G, M, I)$. We write $(X_a, Y_a) \preceq (X_b, Y_b)$ if and only if $X_a \subseteq X_b$ or equivalently $Y_b \subseteq Y_a$. We call (X_a, Y_a) a **subconcept** of (X_b, Y_b). Alternatively, we call (X_b, Y_b) a **superconcept** of (X_a, Y_a).

The relation $\preceq$ on the formal concepts is a partial ordered relation. A **formal concept lattice** is a partially ordered set of formal concepts in which every pair of formal concepts has a unique least upper bound and a unique greatest lower bound.

3.2 Formal Concept Analysis for Property Norms

Semantic property norms have been used to explore and enhance many aspects of the semantic representation and the processing of concepts in cognitive science. The semantic property norms described in (McRae et al. 2005) is one of the most widely used property norm datasets to date. The dataset is collected from approximately 725 participants for 541 living (alligator) and nonliving (airplane) basic-level concepts. Each named concept corresponds to an English noun, which are normed by 30 participants through a questionnaire. The 541 concepts are annotated by a total of 2526 properties. It is a very sparse dataset in which each concept has an average of 13 properties. Table 1 lists some properties of the concept for *bayonet* in the McRae dataset and their production frequencies, i.e., the number of subjects out of 30 participants that listed a property.

Concept	Some Properties	Production Frequency
bayonet	used_in_wars	8
	is_dangerous	6
	a_long_knife	5
	is_sharp	5

Table 1: Some properties for the concept *bayonet* in McRae and their production frequencies.

Conceptual hierarchies, such as hyponymy/hypernymy and meronymy/holonymy, are very important relationships for many natural language processing applications. We attach the mathematical structure of formal concept lattice to property-based concepts and then discover many interesting abstractions from a modest-sized human-annotated perceptual data based on that structure. We use the following simple example to explain the modeling process.

Example 1 Given four property-based concepts, *bayonet*, *grenade*, *spear*, *harpoon*, we can obtain a 4×7 incidence matrix (Table 2) by setting nonzero production frequency to 1.

	is_sharp	is_dangerous	used_in_war	found_on_boats	a_long_knife	has_a_head	explodes
bayonet	1	1	1	0	1	0	0
grenade	0	1	1	0	0	0	1
spear	1	1	0	0	0	1	0
harpoon	1	0	0	1	0	0	0

Table 2: A concept-property matrix

In this example, since no other concept includes the two properties `is_sharp` and `found_on_boats`, $(\{harpoon\}, \{is\_sharp, found\_on\_boats\})$ is a formal concept according to Definition 2. Property `is_sharp` is the only property common to the concepts *bayonet*, *spear* and *harpoon*. Thus, $(\{bayonet, spear, harpoon\}, \{is\_sharp\})$ is a formal concept. Since $\{harpoon\} \subseteq \{bayonet, spear, harpoon\}$, we have $(\{harpoon\}, \{is\_sharp, found\_on\_boats\}) \preceq (\{bayonet, spear, harpoon\}, \{is\_sharp\})$. In a similar way, we can obtain other formal concepts of the formal context and partial order relations among them. A Hasse diagram in Figure 1 is used to visualize the hierarchical structure of the concept lattice, where each node represents a formal concept and each edge denotes the subconcept-superconcept relationship between two formal concepts.

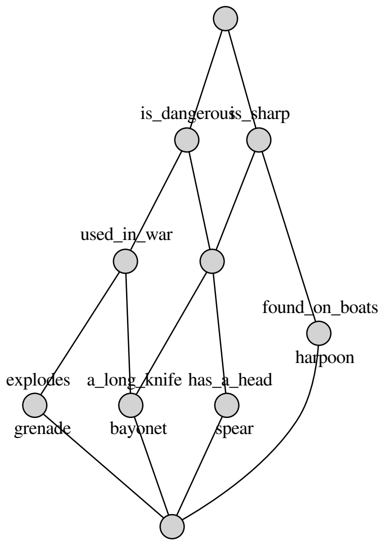

Figure 1: The concept lattice generated from the formal context in Table 2.

4 Experiments and Evaluation

4.1 Data preparation

For the distributional semantic space, we use the 300-dimensional W2V and DEP vectors pretrained on a concatenation of three large, diverse English corpora: (1) English Wikipedia 2015, (2) UMBC web corpus (Han et al. 2013), and (3) English Gigaword newswire corpus (Parker et al. 2011). The concatenated corpus comprises about 10B words and yields a vocabulary of about 500K words after filtering words with frequency lower than 100. The word embeddings W2V1, W2V5, and W2V10 are learned with context window sizes of 1, 5 and 10 respectively. For learning DEP, the corpus is tagged with parts-of-speech and parsed into Stanford dependencies by CoreNLP (Manning et al. 2014). Those vectors are described in (Melamud et al. 2016) and available for downloading [1].

The McRae-based property norm space contains 541 concepts as described in section 3. We delete 3 concepts (*axe*, *armour*, *dunebuggy*) from the McRae dataset because they are not available in the word2vec vocabulary. For the concepts with multiple meanings in McRae, they are disambiguated by providing some cue, for example, *bat_(animal)* and *bat_(baseball)*. In distributional semantic space, their vector representations are also different by computing the average vectors of the concept and its cues respectively. The 538 concepts are split randomly into 400 training data and 138 test data. There are 437 out of 2526 properties which are not seen in the training set. We normalize each property vector by the sum of its production frequencies at training time. Our goal is to learn a mapping from distributional semantic space (400 * 300) to property norm space (400 * 2526), and then predict property norms for 138 new concepts from their distributional semantic representation.

By applying formal concept lattice to these 400 training data, we can generate 3609 formal concepts. We delete two formal concepts with empty intent and empty extent, and use 3607 formal concept as new training data. For each generated formal concept, the input representation of the concept is the average of distributional vectors for words in the extent, and the output representation of the concept is represented by the average of words' property norms for the properties in the intent (those properties not appearing in the intent are set to 0).

[1] http://u.cs.biu.ac.il/~nlp/resources/downloads/embeddings-contexts/

4.2 Task and Baseline

We model the problem of learning the property-based representation of a concept as a multinomial classification problem. Let $\mathcal{X}$ denote the distributional semantic space, and $\mathcal{Y}$ denote the property semantic space. Then we learn a function

$$\Phi : \mathcal{X} \rightarrow \mathcal{Y}, \tag{1}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^m$. Given a test concept $x \in \mathbb{R}^n$, we want to estimate an m-dimensional vector (whose elements sum to 1) which represents the probabilities for m property classes. In this work, we compare two methods: KNN and FCL-based KNN, where for each concept in the test dataset, we choose k most similar concepts from the 400 original training dataset and the 3607 formal concepts-based training dataset respectively. Similarity is based on the cosine similarity of the concept vectors in the distributional semantic space. Then the property vector of the concept in the test dataset is represented by the average of property vectors of these k most similar concepts. We set k to 5, 10, 15, 20 and 25 in our experiments.

4.3 Quantitative Evaluation

In the following experiments, we use the properties in McRae as the gold standard that models are compared against. We study the properties of 138 test concepts from McRae, in which each test concept has 13.39 properties in average. For each test concept, we rank the properties from the predicted property vector in term of their values and pick the top 10 properties from the list. Given the top 10 ranked properties generated by different methods from different distributional models, precision and recall are used for evaluation. The definitions of two standard performance measurements at the 10th property in the ranked list are specified as follows.

Precision is the fraction of the predicted properties that are correct, i.e.,

$$P = \frac{\text{number of correct properties}}{\text{total number of properties returned}}.$$

Recall is the fraction of the properties that have been predicted correctly, i.e.,

$$R = \frac{\text{number of correct properties}}{\text{total number of gold standard properties}}.$$

Table 3 report percentage average precisions and recalls across 138 test concepts by each method matched against the McRae gold standard. Four kinds of word embeddings W2V1, W2V5, W2V10 and DEP are used in the experiments. From the average precision and recall, we see that when $k > 5$, FCL-based KNN is clearly better than KNN, and that higher k (5,10,15,20,25) continues to improve the performance of FCL-based KNN but not that of KNN. The better experimental results of FCL-based KNN illuminate the advantage of formal concept lattice, which has the ability to learn and model potential hierarchical relationships among concepts.

4.4 Qualitative Evaluation

Table 4 reports the top 5 predicted properties returned by KNN(k=10) and FCL-KNN (k=25) based on W2V1 for eight test concepts. We also list their top 5 gold-standard properties annotated by participants in McRae. From Table 4, we see that FCL-based KNN can provide more reasonable properties for these eight concepts than KNN. Properties annotated with $*$ in the table are not listed in McRae. Although we call the property norms in McRae as "gold standard", these annotated properties are sometimes not completely true representations of concepts because the annotation depends on the knowledge background or linguistic habits of participants. A property with zero production frequency for a concept in McRae simply means that the property is not elicited from the conceptual knowledge of participants during the questionnaire.

To further evaluate the quality of property inference, we perform nearest neighbor search for the predicted vector of a concept. The predicted vectors are produced by FCL-based KNN (k=25)and KNN(k=10) based on W2V1. The ideal performance is that the predicted vector of a concept should

Parameters	Distributional Models	**Precision**		**Recall**	
		KNN	FCL-KNN	KNN	FCL-KNN
k=5	W2V1	**36.01**	32.68	**27.76**	24.87
	W2V5	34.63	33.11	26.47	25.50
	W2V10	34.85	32.24	26.74	24.74
	DEP	35.79	32.68	27.26	25.20
k=10	W2V1	36.01	38.18	27.89	29.15
	W2V5	35.86	38.26	27.73	29.36
	W2V10	36.37	37.89	27.95	28.75
	DEP	36.66	**38.98**	28.08	**29.82**
k=15	W2V1	36.08	40.79	28.10	31.17
	W2V5	36.37	40.94	28.33	**31.31**
	W2V10	36.30	40.28	28.05	30.73
	DEP	36.01	**41.01**	27.65	31.28
k=20	W2V1	35.21	41.44	27.40	31.72
	W2V5	35.57	41.59	27.61	31.73
	W2V10	35.65	**42.10**	27.44	31.97
	DEP	36.15	42.02	27.87	**32.15**
k=25	W2V1	33.76	41.73	26.15	31.85
	W2V5	35.28	41.52	27.28	31.64
	W2V10	34.56	**42.46**	26.90	**32.31**
	DEP	35.36	41.52	27.39	31.64

Table 3: The average precisions and recalls (%) of KNN and FCL-based KNN methods for different distributional models.

be close to its gold standard vector in McRae (Herbelot and Vecchi 2015). Table 5 shows the Top 5 neighbors of the predicted vectors among the 138 gold standard property vectors for above eight concepts. For the FCL-based KNN method, six out of eight gold standard vectors are the 1-nearest neighbor to their predicted vectors, while four out of eight gold standard vectors are the 1-nearest neighbor to their predicted vectors based on the KNN method.

5 Conclusion

Distributional semantics and property norms play important roles in many linguistic applications. In this work, we attach the mathematical structure of formal concept lattice to property-based concepts, which can model the potential conceptual hierarchies in the property norm space. The impressive performance has been demonstrated when building a mapping from a distributional semantic space onto a FCL-based property space. In future work, we will do further analysis about the lattice-based property norm space and explore how to generalize property-based representations to a large-scale dataset.

Concept	Method	Top 5 Predicted Properties
jar	McRae	has_a_lid, made_of_glass, used_for_holding_things, a_container, is_breakable,
	FCL-KNN	made_of_plastic, made_of_glass, ∗found_in_kitchens, ∗made_of_metal, used_for_holding_things
	KNN	∗made_of_metal, made_of_plastic, ∗found_in_kitchens, used_for_holding_things, ∗a_utensil
sparrow	McRae	a_baby_bird, beh_flies, has_feathers, beh_lays_eggs, has_wings
	FCL-KNN	a_baby_bird, beh_flies, has_feathers, has_wings, has_a_beak
	KNN	a_baby_bird, beh_flies, has_feathers, has_wings, has_a_beak
spatula	McRae	a_utensil, has_a_handle, made_of_plastic, used_for_cooking, is_flat,
	FCL-KNN	made_of_metal, found_in_kitchens, made_of_plastic, a_utensil, used_for_cooking
	KNN	made_of_metal, found_in_kitchens, ∗a_tool, made_of_plastic, has_a_handle
sofa	McRae	found_in_living_rooms, furniture, is_comfortable, used_by_sitting_on, has_cushions
	FCL-KNN	is_comfortable, is_soft, used_by_sitting_on, furniture, ∗made_of_material
	KNN	is_comfortable, is_soft, ∗made_of_wood, ∗worn_for_warmth, used_for_sleeping
bracelet	McRae	worn_on_wrists, made_of_gold, made_of_silver, a_fashion_accessory, a_jewelry
	FCL-KNN	∗worn_around_neck, made_of_silver, made_of_gold, ∗is_long, ∗different_colours
	KNN	∗worn_for_warmth, ∗clothing, ∗is_long, ∗worn_by_women, ∗different_colours
doll	McRae	has_own_clothes, used_for_playing, a_toy, used_by_girls, has_hair
	FCL-KNN	∗is_soft, ∗is_comfortable, ∗has_4_legs, ∗different_colours, ∗is_dirty
	KNN	∗is_comfortable, ∗worn_for_warmth, ∗worn_at_night, ∗is_warm, ∗clothing
walrus	McRae	an_animal, is_large, beh_swims, lives_in_water, is_fat
	FCL-KNN	an_animal, beh_swims, ∗beh_lays_eggs, hunted_by_people, a_mammal
	KNN	an_animal, ∗has_a_tail, has_teeth, ∗is_green, ∗is_furry
platypus	McRae	an_animal, lives_in_water, a_mammal, beh_swims, has_a_bill
	FCL-KNN	an_animal, ∗is_green, ∗a_reptile, beh_swims, ∗has_legs
	KNN	an_animal, ∗is_green, ∗has_4_legs, ∗beh_eats, ∗has_a_tail

Table 4: Top 5 Properties returned by FCL-KNN and KNN. Properties annotated with ∗ are not listed in McRae.

Concept	Method	Top 5 Neighbors
jar	FCL-KNN	jar, spatula, bucket, tongs, plate
	KNN	spatula, tongs, bucket, grater, pan
sparrow	FCL-KNN	sparrow, raven, finch, buzzard, parakeet
	KNN	sparrow, raven, finch, buzzard, parakeet
spatula	FCL-KNN	spatula, tongs, grater, pan, skillet
	KNN	spatula, tongs, hatchet, grater, bucket
sofa	FCL-KNN	sofa, cushion, jeans, bench, sandals
	KNN	sofa, cushion, socks, bench, cabinet
bracelet	FCL-KNN	bracelet, tie, crown, skirt, cape
	KNN	skirt, socks, cape, tie, jacket
doll	FCL-KNN	cushion, sofa, sheep, bear, rice
	KNN	socks, bracelet, cape, skirt, bench
walrus	FCL-KNN	platypus, walrus, otter, ox, elk
	KNN	walrus, ox, platypus, otter, cougar
platypus	FCL-KNN	platypus, ox, walrus, otter, elk
	KNN	cougar, ox, buffalo, elk, walrus

Table 5: Top 5 Neighbors returned by FCL-KNN and KNN.

References

Marco Baroni and Alessandro Lenci. 2008. Concepts and properties in word spaces. *Italian Journal of Linguistics*, 9:1–36.

Marco Baroni, Brian Murphy, Eduard Barbu, and Massimo Poesio. 2010. Strudel: A corpus-based semantic model based on properties and types. *Cognitive Science*, 34:222–254.

Gemma Boleda and Aurélie Herbelot. 2017. Formal distributional semantics: Introduction to the special issue. *Special Issue of Computational Linguistics on Formal Distributional Semantics*, 42:619–635.

Brian Davey and Hilary Priestley. 1990. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, UK.

Franck Dernoncourt. 2016. Mapping distributional to model-theoretic semantic spaces: a baseline. In *arXiv:1607.02802*.

Barry Devereux, Nicholas Pilkington, Thierry Poibeau, and Anna Korhonen. 2009. Towards unrestricted, large-scale acquisition of feature-based conceptual representations from corpus data. *Research on Language and Computation*, 7:137–170.

Barry J. Devereux, Lorraine K. Tyler, Jeroen Geertzen, and Billi Randall. 2014. The centre for speech, language and the brain (cslb) concept property norms. *Behavior Research Methods*, 46:1119–1127.

Katrin Erk. 2016. What do you know about an alligator when you know the company it keeps? *Semantics & Pragmatics*, 9:1–63.

John Rupert Firth. 1957. A synopsis of linguistic theory 1930–1955. In *Studies in Linguistic Analysis*, pages 1–32, Blackwell, Oxford.

Luana Făgărăşan, Eva Maria Vecchi, and Stephen Clark. 2015. From distributional semantics to feature norms: grounding semantic models in human perceptual data. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 52–57.

Bernhard Ganter and Rudolf Wille. 1999. *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag, Berlin, Germany.

George Grätzer. 1998. *General Lattice Theory*. Birkhauser Verlag, Berlin, Germany.

Abhijeet Gupta, Gemma Boleda, Marco Baroni, and Sebastian Padó. 2015. Distributional vectors encode referential attributes. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 12–21.

Lushan Han, Abhay L. Kashyap, Tim Finin, James Mayfield, and Johnathan Weese. 2013. Umbc ebiquity-core: semantic textual similarity systems. In *Proceedings of the Second Joint Conference on Lexical and Computational Semantics. Association for Computational Linguistics*.

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

Aurélie Herbelot and Eva Maria Vecchi. 2015. Building a shared world: Mapping distributional to model-theoretic semantic spaces. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (Best Paper Award)*, pages 22–32.

Madori Ikeda and Akihiro Yamamoto. 2017. Extending various thesauri by finding synonym sets from a formal concept lattice. *Journal of Natural Language Processing*, 24:323–349.

Omer Levy and Yoav Goldberg. 2014. Dependency-based word embeddings. In *Proceedings of the 52nd Anua Meeting of the Associations for Computational Linguistics (short papers)*, pages 302–308.

Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

Ken McRae, George S. Cree, Mark S. Seidenberg, and Chris McNorgan. 2005. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37:547–559.

Oren Melamud, David McClosky, Siddharth Patwardhan, and Mohit Bansal. 2016. The role of context types and dimensionality in learning word embeddings. In *Proceedings of NAACL-HLT*, pages 1030–1040.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. In *arXiv preprint arXiv: 1301.3781*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*, pages 3111–3119, Nevada, USA.

Gregory Murphy. 2002. *The Big Book of Concepts*. The MIT Press, Cambridge, MA.

Robert Parker, David Graff, Junbo Kong, Ke Chen, and Kazuaki Maeda. 2011. English gigaword fifth edition. In *Linguistic Data Consortium, LDC2011T07*.

Uta E. Priss. 1998. The formalization of wordnet by methods of relational concept analysis. In *WordNet: An Electronic Lexical Database and Some of its Applications*, pages 179–196. MIT Press.