Akari Izumi, Daniel Smits, Max Mines
{akari_izumi,daniel_smits,max_mines}@brown.edu
CLPS 1850/CSCI 2952I - May, 2020

Ambiguity within Ambiguity: ELMo and the Homonymy-Polysemy Distinction

Introduction:

Word ambiguity is everywhere: many—if not most—of the words we interact with on a daily basis have multiple meanings. Linguists traditionally distinguish between two kinds of ambiguity: homonymy and polysemy. Homonymy refers to a single phonological term that has several semantically *unrelated* meanings, or 'senses.' A paradigmatic example of homonymy is *bank*, which can refer to a financial institution or to a riverside. Polysemy, on the other hand, commonly refers to a single phonological term with several *related* senses. A common example of polysemy is *run*, which has many related but non-identical meanings, such as in "go for a 3 mile run" or "go for a grocery run."

How are these ambiguous terms represented in the mental lexicon? How are these mental representations used to contextually disambiguate word senses? Are homonymy and polysemy two distinct phenomena? Or does ambiguity exist along a spectrum? Though there is some psycholinguistic and computational evidence for a homonymy-polysemy (HP) distinction, experts are yet to reach a consensus as to how we lexically represent and process ambiguous terms in context.

Psycholinguists have found evidence that homonymous words have a single phonological and orthographic representation that is linked to separate lexical entries for each of its meanings. They observed that, at the presentation of a homonym, all of its meanings are initially activated, independently of context, and therefore prime words only immediately after target presentation. The competing senses are disambiguated milliseconds later, as the remaining context is processed (Swinney, 1979; Williams, 1992). This same pattern, however, was not observed during the presentation of polysemous words. In fact, polysemous words were able to prime contextually unrelated words even moments after presentation of the target (Williams, 1992).

In studies which provide evidence for an HP distinction, there are two views on how polysemous words are represented in the brain. According to the "single lexical entry" view, all polysemous senses are derived from a single lexical entry. There may be no need to separately store each polysemous sense, since there are common ways to derive polysemous senses from a single core meaning using context and familiar patterns of extension (Nunberg, 1979). Common polysemous words can have more than 20 distinct senses, which leads some researchers to argue that they could not all be represented separately (Caramazza and Gober, 1976). MEG and EEG studies have also shown activations in the brain that are consistent with such a theory (Pylkkanen, Llinas and Murphy, 2006; MacGregor, Bouwsema and Klepousniotou, 2015). However, this "single lexical entry" view may fail to account for edge cases, whose senses cannot be directly derived from a single core representation and must be represented separately.

Countering the core representation view is the "separate-lexical entry" theory, in which the different senses of a polysemous word each have their own separate entries in the lexicon, much like how homonymous words are represented. Klein and Murphy (2001) observed an inhibition effect in using two different senses of a word, meaning there had to have been separate entries because it is not possible to prime one sense of a word and not the other if they were represented under a single entry. Other studies have found that the different senses are too different to derive from a core meaning and therefore that it does not make sense to represent them in a single entry (Langacker, 1987; Tuggy, 1993).

While these distinctions are heavily disputed in psycholinguistic circles, there is less extensive computational research into the HP distinction. In fact, many researchers in computational linguistics use the words “homonymy” and “polysemy” interchangeably. The closest related work is the effort to computationally model lexical ambiguity. Developing computational systems to determine the intended sense of a word in a given context, known as word-sense disambiguation (WSD), is an ongoing area of research. Early computational work includes Senseval, a program created in 1997 to evaluate the strengths and weaknesses of WSD algorithms. Since then, Senseval has evolved to incorporate WordNet and has seen success in both supervised and unsupervised systems to reliably solve WSD tasks (Edmonds and Cotton, 2001; Mihalcea, Chlovski and Kilgarriff, 2004). Other traditional studies have adapted WordNet to a knowledge-based, WSD-specific database to improve its performance (Dhungana et al., 2015).

Despite these efforts, language models continued being trained and tested using context-invariant word embeddings. These embeddings, such as word2vec (Mikolov et al., 2013) and GLoVe (Pennington et al., 2014) are pre-trained over large corpora with a language modeling task. Once trained, these embeddings are static and perform poorly on word sense disambiguation tasks. A context-insensitive word embedding, on its own, has no way of determining whether a particular instance of a word like *bat* refers to the wooden sports instrument or to the flying mammal.

With the popularization of deep learning came new tools to facilitate word-sense disambiguation. Recent studies have found that recurrent models, such as bidirectional LSTMs, capture useful information for contextual disambiguation in their embeddings and hidden states (Aina, Gulordava and Boleda, 2019). Nevertheless, there is still room for improvement.

This improvement came in the form of deep *contextual* language models, which compute context-dependent word embeddings. ELMo, in particular, was developed to disambiguate polysemy using a bidirectional LSTM (Peters et al. 2018). It models complex features of word behavior and how words interact in context. ELMo was quickly followed by BERT, a transformer-based model trained on a masked-word task (Devlin et al., 2019). These language models pushed the state of the art across several NLP tasks by augmenting static word embeddings with bidirectional contextual information.

While most computational studies have focused on WSD in general, there are a few studies that have recognized the distinction between homonymy and polysemy. As the psychological data suggest, computational linguists have also claimed that there are no general, systematic methods to differentiate between homonymy and polysemy (Klein and Murphy, 2002; Kilgarriff and Palmer, 2000). More recent studies, however, have attempted to model polysemy based on its feature of having systematicity in the different senses a word could have, utilizing WordNet and CoreLex, which provides a finite set of basic ontological features (Utt and Padó, 2011). Even so, this does not account for every instance of polysemy. Researchers in both fields are still exploring ways to systematically differentiate polysemy and homonymy.

The results of this study have implications on how words with multiple meanings can be processed and how different word meanings can be represented and retrieved, as contrasting theories present trade-offs of storage and processing. The issues of disambiguation and lexical representation are hotly disputed, as well as whether this is even a distinction worth making. Nevertheless, one particularly important use for the HP distinction, beyond improving word-sense disambiguation performance, is for machine translation. Polysemous words which have a metonymic sense are particularly challenging to translate correctly, since the referent of the correct *sense* must first be retrieved in the original language in order to correctly find an equivalent sense in the target language (Kamei and Wakao, 1992). The debate over the homonymy-polysemy distinction has important consequences for our understanding of the structure of mental lexicon, and can provide insights to help language instructors and developmental psycholinguists teach children new lexical items and assist them in disambiguation.

In order to help bridge the gaps between psycholinguists and computational linguists, we focus on the lexical semantic representations of contextual language models. We first investigate whether statistically trained contextual language models distinguish between ambiguous and unambiguous words. We then probe as to whether, within the representation of ambiguous words, these contextual language models distinguish between polysemous and homonymous word senses.

Experimental Design:

Many psychological studies arguing for a homonymy-polysemy distinction rely on reaction-time or eye-tracking data. ELMo, however, does not have eyes and does not exhibit meaningfully different reaction times. Language models—unlike humans—allow direct access to semantic word representations. If a language model were to capture a distinction between polysemy and homonymy, we would expect that difference to be detectable within the lexical semantic representations. In the case of ELMo, the semantic representation is instantiated by a contextual output embedding. To investigate the representations of homonymy and polysemy in contextual language models, therefore, we analyze ELMo's embeddings across various senses and contexts. In particular, we operationalize the semantic similarity of two embeddings as their cosine similarity, a function of the angle between two vectors in Euclidean space.[1]

Context-invariant embeddings historically performed poorly on word-sense disambiguation tasks. Contextual language models were designed with the purpose of improving sense-disambiguation, and have shown success in solving WSD tasks with contextual embeddings. Consequently, we chose to evaluate ELMo in this study because its left and right contexts are distinctly represented in the output embedding, and because it requires no additional pre-training or fine-tuning. Issues related to the distribution of our data are discussed below.

Just as static word embeddings tend to cluster together in the embedding space based on semantic similarity, so, too, may contextual word embeddings cluster based on sense. If ELMo distinguishes between ambiguous and unambiguous word *senses*, we would expect the cosine similarity between contexts for unambiguous terms to be significantly different (and in this case, greater) than between contexts for different senses of ambiguous terms. Additionally, if ELMo is sensitive to the homonymy-polysemy distinction, we expect the cosine similarity between senses of homonymous words to be significantly different from the between-sense cosine similarity of polysemous words. Given that the senses of homonymous words are supposedly more different

[1] The cosine similarity of vectors A and B is the dot product of A and B divided by the product of their respective magnitudes.

from each other than those of polysemous words, we would expect the cosine similarity of polysemous words to be *greater* (i.e. angle is smaller) than that of homonymous terms. Similarly, if homonymy and polysemy are graded phenomena, existing along a spectrum, we would expect a significant difference, though our current methods cannot investigate whether the distinction is binary or graded. If, alternatively, ELMo is insensitive to any HP distinction, then there should not be a significant difference in the between-sense cosine similarity of polysemous and homonymous words.

We still have much to learn about the properties and types of ambiguity. As researchers identify new dimensions along which polysemy can vary, they find new ways to control for extraneous factors and probe into the representation and disambiguation of words. In fact, some psycholinguistic studies have identified phenomena which are associated with only some types of polysemy but not others (Utt and Padó, 2011). More work must be done in identifying and evaluating these dimensions of ambiguity, which will provide key insights into how and whether homonymy and polysemy are differently processed and represented by the human mind. This present study, however, solely probes ELMo's semantic representations for a distinction between ambiguity and unambiguity or homonymy and polysemy.

Data:

Our data consist of 40 polysemous, 40 homonymous, and 40 unambiguous nouns. These nouns were primarily chosen from canonical examples in prior psycholinguistic literature (Klepousniotou and Baum, 2007; Durkin and Manning, 1989). We augmented the data with additional nouns of each word type in order to create sets of equal size across conditions.

As their name implies, *contextual* language models are sensitive to changes in context. ELMo computes its output embedding for a word as a weighted sum of its static embedding with the encodings of the left and right contexts. Consequently, a word's embedding can vary across contexts, even if used in the *same* sense. In order to control for contextual variation, we compute the within-sense variation of word embeddings across a diversity of contexts. Using unambiguous nouns—which ideally have only one sense regardless of context—avoids confounds caused by subtle polysemy. We use these unambiguous nouns as a baseline for contextual variation, without any variation due to sense.

In order to control for syntactic confounds, our data are restricted to only nouns. Some words can have different senses as a noun or verb (e.g. *rock*: stone/shake), and some verbs can have multiple senses (e.g. *lie*: rest/deceive'). Verbs and nouns occur in different syntactic frames, and most studies of homonymy or polysemy focus on nouns. We therefore focus on noun-noun ambiguity (e.g. *bank*: river/financial) and unambiguous nouns.

Examples of noun-noun homonymy from our data include *cell*: prison/biology, *mass*: church/matter, and *pen*: writing/enclosure. Examples of noun-noun polysemy from our data include *bottle*: object/quantity, *chest*: anatomical/treasure, *glass*: object/material. Examples of unambiguous nouns from our data include *oatmeal*, *guest*, *earring*.

For each term in our dataset, we constructed pairs of "artificial" contexts to control for syntactic and semantic variation. These sentences have a generic actor as the subject, a transitive verb, the target noun as a direct object, and a prepositional phrase. For ambiguous terms, the bidirectional context facilitates disambiguation. For unambiguous terms, the contexts are deliberately contrastive. We hope that this controls for variation caused by the syntactic frame or semantic role. For exceptional nouns which are infelicitous in this frame, we modify the artificial template slightly but ensure that sentence pairs are syntactically and semantically symmetrical.

Despite the benefits of controlling for syntactic, semantic, and contextual variation, our artificial sentences may be unnatural. These examples are not drawn from the same distribution as that on which ELMo was trained. Therefore, we hope to replicate and validate our results from the artificial sentences by compiling three natural contexts for each sense, selected from an online corpus (SketchEngine). Since these natural examples vary in length, style, and context, we take the mean over their embeddings to reduce variation and capture an 'average sense.'

An example of a standardized sentence pair from our data is "the sailor sweeps the *deck* of the ship"/"the gambler wins the *deck* of cards." An example of an 'alternative' to the standardized sentence pair is "the employee cashes his paycheck at the *bank* of america"/"the rower pulls the boat onto the *bank* of the river." An example of a natural sentence pair in our data is "*Seal* populations are like a huge fishery."/"The use of a *seal* by men of wealth and position was common before the Christian era."[2]

Due to the asymmetry between one artificial sentence and three natural sentences, we conduct two analyses. First, we compare each of the word types {Homonymy, Polysemy, and Unambiguous} between context types {Artificial, Natural} by randomly sampling one natural context for each sense. Next, we compare the word-types in the averaged natural contexts for any indication of a distinction between ambiguity and unambiguity or polysemy and homonymy.

Frequency and co-occurrence are very important factors in how a deep language model learns. Previous psycholinguistic studies have noted that some senses of polysemous and homonymous words seem primary or 'dominant,' implying that some ambiguous terms have one core sense with several variants. By operationalizing this dominance as the sense with highest frequency, some studies have accounted for differences in processing to the 'accessibility' of a particular sense. These studies indicate that the dominant sense—which is more frequent and accessible—may be privileged over secondary and tertiary senses (Simpson, 1981; Tabossi, Colombo and Job, 1987). Later work used this hypothesis to distinguish between ambiguous terms which are 'balanced' (equivalent frequency, no dominance) versus those which are 'unbalanced' (more frequent sense is dominant) (Klepousniotou and Baum, 2007). Though we would ideally control for word-sense frequency and consider the role of 'dominant' senses, sense-frequency data is sparse and unreliable. Additionally, the criteria used in previous studies for labeling some words 'balanced' and others 'unbalanced' are opaque. Though language models are sensitive to the distributions on which they are trained, we do not have the resources to balance our data for frequency. We therefore set aside dominance and frequency to focus on the semantic similarity of different senses.

Implementation:

The implementation was executed on a Python (v3.6.9) notebook in Google Colab. All computations were remotely processed on a Google Collab server. The data was first compiled in a Google Sheets document (.csv), then loaded into a Pandas DataFrame (v1.0.3) in the Python notebook.

There were 960 sample sentences in our dataset, coming from eight contexts per word (six natural contexts, two artificial contexts) for each of the 40 terms for each of the three word

[2] The complete terms and sentences dataset can be accessed in a printable format here

types (homonymy, polysemy, unambiguous.) These sentences were preprocessed by converting them to lowercase, stripping all punctuation and additional whitespace using a regular expression, and splitting on spaces to create a list of words. Sample sentences were tokenized using the word_tokenize method from NLTK (v3.2.5) and then lemmatized by the WordNetLemmatizer. This allowed us to search the list of lemmas to identify the index of the target word within the list after being processed by ELMo.

ELMo (v3) was loaded by Tensorflow Hub (v0.7) in Tensorflow 1 (v1.15). The sentence lists were then individually fed to the model using the tokens input mode (signature = "tokens", asdict = True). The embedding of the target word was then extracted as Numpy (v1.18.3) vectors based on the index described above, and stored in a pickled JSON for data analysis.

Analysis:

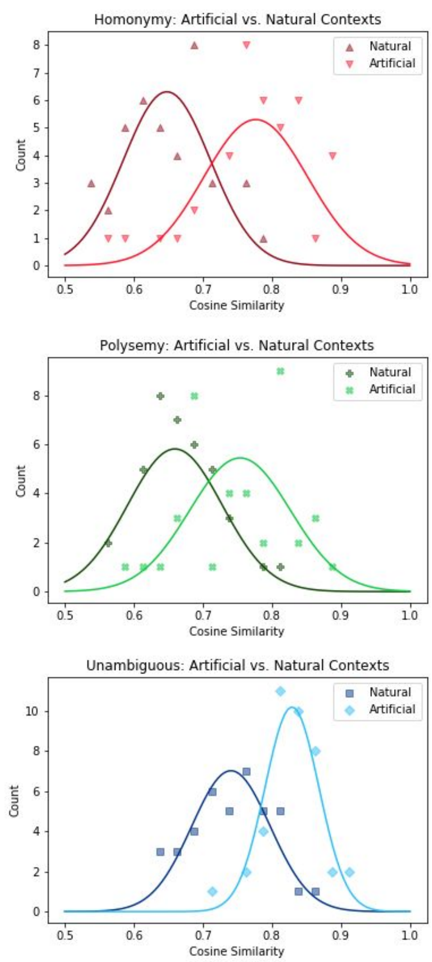
Initial analysis was executed on a local Jupyter Notebook (v 6.0.3) running Python (v3.8.1). Data was managed with Pandas (v1.0.1). Cosine similarities were calculated using a custom Numpy function (v1.18.1). Plots were generated in PyPlot (matplotlib v3.1.3) with Gaussians calculated in SciPy (v1.4.1). Statistical analyses were computed in JASP (v0.12.2.0).

Results:

In our first analysis, a two-way ANOVA yields significant main effects for both context ($F(2, 39) = 36.456$, $p<.001$) and word ($F(1, 39) = 154.804$, $p<.001$) types, but no significant interaction ($F(2, 39) = 2.266$, $p=.106$). As shown in Figures 1.A-C[3], context type has a large effect size, with artificially similar contexts leading to more similar sense-embeddings ($t = -12.442$, Cohen's $d = -1.404$, $p_{tukey} < .001$). As demonstrated by Figures 2A-B (below), post-hoc comparisons find significant difference between ambiguous and unambiguous nouns (homonymy: $t(39) = -7.108$, Cohen's $d = -.886$, $p_{tukey} < .001$; polysemy: $t(39) = -7.652$, Cohen's $d = -1.021$, $p_{tukey} < .001$), with the different senses of ambiguous nouns being less similar than the self-same senses of unambiguous nouns in different contexts. We do not, however, find any such difference between homonymy and polysemy ($t(39) = .543$, $p_{tukey} = .85$).

In our second analysis, a one-way ANOVA again yields a significant difference between word types ($F(2, 39) = 45.891$, $p < .001$). A post-hoc comparison finds that averaging over natural contexts leads to an even larger effect size, with ambiguous nouns being much less similar than unambiguous nouns (homonymy: $t(39) = -8.339$, Cohen's $d = -2.037$, $p_{tukey} < .001$; polysemy: $t(39) = -8.254$, Cohen's $d = -2.036$, $p_{tukey} < .001$). In fact, averaging over contexts erases nearly any difference between homonymy and polysemy ($t(39) = -.084$, Cohen's $d = -.016$, $p_{tukey} = .996$) which—as demonstrated in Figure 3 (below)—makes the two plots overlap almost perfectly.

Figures 1.A-C



[3] In all plots, a Gaussian is computed over the cosine similarities. For transparency, we overlay the histogram counts (bin size = .05, range = [0,1]) from the raw data as a scatter plot.

Figure 2.A

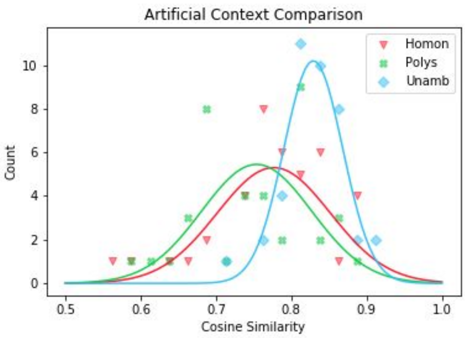


Figure 2.B

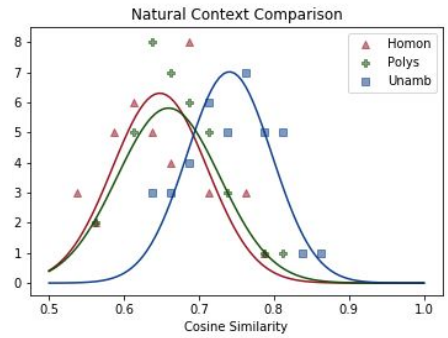
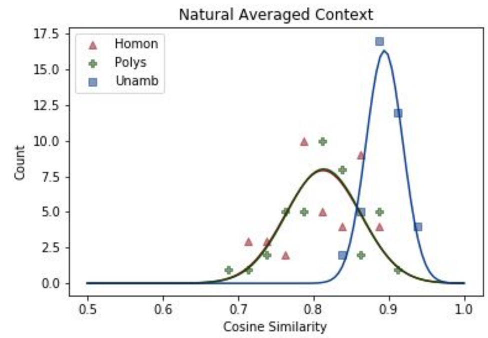


Figure 3[4]



Discussion:

Our analyses lead us to reject the first null hypothesis. This result provides evidence that ELMo distinguishes its representation of ambiguous and unambiguous nouns, likely implicating some latent notion of "sense" beyond just contextual differences. This result aligns with the fact that ELMo was developed in order to improve contextual word-sense disambiguation, and pushed the state-of-the-art for WSD tasks.

This WSD objective, however, does not seem to naturally give rise to any additional distinction between homonymy and polysemy. Our analyses fail to reject the second null hypothesis, as our data provides no evidence that ELMo distinguishes between homonymy and polysemy. Our findings seem to contrast ELMo's lexical representation of ambiguity to those that psycholinguists theorize in explaining language processing in humans. Some readers may weigh this as evidence against the existence of an HP distinction, but we must emphasize that although ELMo—or any other computational model—does not learn an HP distinction 'out of the box', that does not license conclusions about the human mind or about the upper bounds of statistical learning. Studying computational language models cannot serve as substitutes for psycholinguistic research. When language models *succeed* at a task, however, we can learn that certain objectives and architectures are *sufficient* for the performance of said task.

The success of ELMo on WSD tasks, combined with our findings about the distinct representations of ambiguous senses, provide compelling evidence that ambiguity resolution in general can be indirectly acquired through statistical learning. Nevertheless, the question remains open for psycholinguists and computer scientists alike as to how humans might learn the HP distinction. These early results bode ill for any theory that humans use distributional models to learn the distinction between homonymy and polysemy, though symbolic approaches do not yet provide a compelling alternative.

Our study has important limitations. Firstly, this is a study of whether ELMo does—*not* whether it *could*—encode differences between homonymy and polysemy. Secondly, the asymmetry between artificial and natural contexts in our data limits our ability to generalize our results; our comparison of artificial and natural contexts is merely exploratory. Thirdly, with more time and resources, we evaluate the naturalness of our sentences, balance sets for word-sense frequency, and validate whether the contexts allow for decisive disambiguation. Finally, it remains possible that averaging the embeddings or relying on cosine similarity obfuscates important distinctions between embeddings.

[4] The homonymy and polysemy Guassians overlap in this plot, making it hard to distinguish the curves.

Future studies should evaluate the kinds of words and contexts in which ELMo fails to disambiguate senses, as well as whether ELMo's performance on WSD tasks is affected by the polysemy-homonymy distinctions or other features of ambiguity. Such research could broaden the scope of inquiry to include more than just noun-noun ambiguity, and investigate differences in frequency and type of ambiguity, such as the difference between metonymic and metaphoric polysemy. Whether it would be advantageous to train models to make the homonymy-polysemy distinction remains unknown. Perhaps, WSD performance could improve through a more lexically flexible architecture inspired by the psycholinguistic "separate lexical entry" theory. In any case, the fact that state-of-the-art models maintain single lexical entries and disambiguate from context *without* any HP distinction indicate that said distinction may not be computationally necessary.

Conclusion:

The present study shows that contextual language models, such as ELMo, indirectly learn to distinguish between ambiguous and unambiguous nouns through statistical learning. This provides evidence that these models can compute representations of sense which are at least partially independently of context. We do not find evidence for ELMo distinguishing between homonymy and polysemy through the spatial organization of its contextual embeddings, though more research must be done into how deep contextual models lexically represent and process ambiguity.

Bibliography

Aina, L., Gulordava, K., & Boleda, G. (2019). Putting Words in Context: LSTM Language Models and Lexical Ambiguity. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. doi: 10.18653/v1/p19-1324

Caramazza, A., & Grober, E. (1976). Polysemy and the structure of the subjective lexicon. In C. Rameh (Ed.), *Georgetown University roundtable on language and linguistics*. Washington, DC: Georgetown University Press.

Chelba, C., Mikolov, T., Schuster, M., Ge, Q., Brants, T., & Koehn, P. (2014). One billion word benchmark for measuring progress in statistical language modeling. arXiv preprint, 1312.3005, 2013. arxiv.org/abs/1312.3005.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pretraining of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

Dhungana, U. R., Shakya, S., Baral, K., & Sharma, B. (2015). Word Sense Disambiguation using WSD specific WordNet of polysemy words. *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)*, 148–152. doi: 10.1109/icosc.2015.7050794

Durkin, K., & Manning, J. (1989). Polysemy and the subjective lexicon: Semantic relatedness and the salience of intraword senses. *Journal of Psycholinguistic Research*, *18*(6), 577–612. doi: 10.1007/bf01067161

Edmonds, P. & Cotton, S. (2001). Senseval-2: Overview. In Proceedings of the 2nd International Workshop on Evaluating Word Sense Disambiguation Systems (Senseval-2, Toulouse, France). 1–6.

Kamei, S., & Wakao, T. (1992). Metonymy. *Proceedings of the 30th Annual Meeting on Association for Computational Linguistics* -. doi:10.3115/981967.982015

Kilgarriff, A., Palmer, M. Introduction to the Special Issue on SENSEVAL. *Computers and the Humanities* **34,** 1–13 (2000). https://doi.org/10.1023/A:1002619001915

Klein, D. E., & Murphy, G. L. (2001). The Representation of Polysemous Words. *Journal of Memory and Language*, *45*(2), 259–282. doi: 10.1006/jmla.2001.2779

Klein, D. E., & Murphy, G. L. (2002). Paper has been my ruin: conceptual relations of polysemous senses. *Journal of Memory and Language*, *47*(4), 548–570. doi:

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics, 20*(1), 1-24. doi:10.1016/j.jneuroling.2006.02.00110.1016/s0749-596x(02)00020-7

Langacker, R. W. (1987). *Foundations of cognitive grammar* [Stanford Univ. Press].

Macgregor, L. J., Bouwsema, J., & Klepousniotou, E. (2015). Sustained meaning activation for polysemous but not homonymous words: Evidence from EEG. *Neuropsychologia*, *68*, 126–138. doi: 10.1016/j.neuropsychologia.2015.01.008

Mihalcea, R., Chklovski, T. & Kilgarriff, A. (2004). The Senseval-3 English lexical sample task. *Proceedings of SENSEVAL-3*. 25-28.

Nunberg, G. (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy*, *3*(2), 143–184. doi: 10.1007/bf00126509

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. doi:10.3115/v1/d14-1162

Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K,, & Zettlemoyer, L. (2018). Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long Papers), pp. 2227–2237. Association for Computational Linguistics, 2018. URL http://aclweb.org/anthology/N18-1202.

Pylkkänen, L., Llinás, R., & Murphy, G. L. (2006). The Representation of Polysemy: MEG Evidence. *Journal of Cognitive Neuroscience*, *18*(1), 97–109. doi: 10.1162/089892906775250003

Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior*, *20*(1), 120–136. doi: 10.1016/s0022-5371(81)90356-x

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 645–659. doi: 10.1016/s0022-5371(79)90355-4

Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research*, *49*(2-3), 161–167. doi: 10.1007/bf00308682

Tuggy, D. (1993). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics*, *4*(3), 273–290. doi: 10.1515/cogl.1993.4.3.273

Williams, J. N. (1992). Processing polysemous words in context: Evidence for interrelated meanings. *Journal of Psycholinguistic Research*, *21*(3), 193–218. doi: 10.1007/bf01068072

Utt, J. & Padó, S. (2011). Ontology-based distinction between polysemy and homonymy. In *Proceedings of the 9th International Conference on Computational Semantics*, 265-274.