Analyzing adjectival homonyms and polysemy: Unsupervised methods for enhanced Large Language Model understanding

Enikő Héja and **Noémi Ligeti-Nagy**
HUN-REN Hungarian Research Centre for Linguistics
heja.eniko@hun-ren.nytud.hu
ligeti-nagy.noemi@hun-ren.nytud.hu

Relevant UniDive working groups: WG2, WG3

1 Introduction

1.1 Motivation

Large language models (LLMs) are known for achieving state-of-the-art results on various benchmarks through effective prompting and few-shot learning techniques (Yang et al., 2023; Qin et al., 2023; Laskar et al., 2023). However, the Word-in-Context (WiC) dataset (Pilehvar and Camacho-Collados, 2019) poses unique challenges that set it apart from this general trend: *"WiC is a notable weak spot with few-shot performance equivalent to random chance. We tried a number of different phrasings and formulations for WiC [...], none of which was able to achieve strong performance"* (Brown et al., 2020). Moreover, this finding is strongly correlated with the previous experiments of Véronis, 2003. His study proved that assigning the right word sense to target words in the context of one paragraph text is a great challenge for all PoS categories, even for humans. This challenge becomes more evident, especially when acknowledging the intricate nature of meaning distinctions, which poses difficulties even for expert lexicographers when establishing the micro- and macrostructure of dictionaries (cf. Adamska-Sałaciak (2006)). These observations underpin our past experience, according to which the traditional definitions of polysemy and homonymy[1] do not lend themselves easily to serve as the starting point for data-oriented work. This situation leads us to question the fundamental aspects of the WiC task, more precisely, the traditional conception of homonymy and polysemy.

1.2 The Word-in-Context corpus

The Word-in-Context corpus, which forms part of the SuperGLUE benchmark dataset (Wang et al., 2020), focuses on a specific sense disambiguation task: it has to be decided whether two occurrences of a given target word in two different contexts convey the same meaning or not. A significant concern is the low Inter-Annotator Agreement (IAA) in the WiC task, which points to potential ambiguities in how the task is structured (Artstein and Poesio, 2008). Moreover, the target data is not well established in terms of lexical semantic categories (eg. homonymy, polysemy, regular or irregular polysemy, metonymy and metaphor), which may result in different perceptions of meaning identity.

1.3 Our approach

In the present paper we concentrate on adjectival *homonyms* in the WiC context. We aim to show that instead of conveying unrelated meanings, *homonyms are best to conceive of as semantic idiosyncrasies*, that is, wordforms merging (at least) two random meanings in a unique way. Within the context of distributional semantics, this definition suggests a straightforward operational approach capable of identifying homonyms through unsupervised, data-driven means. This method appears to be language-independent, allowing for a cross-linguistically unified lexicography of homonyms, which are idiosyncratic constructions. We also anticipate that our work will offer valuable insights into lexical semantics and improve LLM performance in tasks related to lexical semantics.

2 Previous work

In the present research, we build upon previous findings that use graph-based models to interpret adjectival senses from monolingual corpora. Héja and Ligeti-Nagy (2022) observed that despite the *meaning conflation deficiency* (cf. Pilehvar, 2019), static word embeddings for adjectives (Mikolov et al., 2013a,b) can consistently and robustly capture certain adjectival semantic structures. This finding was evidenced through focusing on local structures of graph G, which was induced from the word2vec representations of $10,153$ Hungarian adjectives by applying a suitable K cut-off parameter on the pairwise cosine similarities of the adjectives.

[1] "If the different meanings associated with a lexical form can be related to one another conceptually [...], we can be reasonably certain that we are dealing with a polysemous lexeme." Ježek (2016)

Namely:

- Single adjectival cliques[2] are good candidates to represent adjectival meanings.
- Adjectival polysemy is indicated by shared cliques, i.e. an adjective belonging to multiple cliques.
- Connected graph components dissect graph G into neatly characterized semantic domains. $6,417$ adjectives were told apart into $1,807$ categories, such as quantities (eg. *gyűszűnyi* 'thimbleful', *cseppny* 'a drop of', *hajszálnyi* 'hair's breadth'), monastic orders, and country names.

However, an intriguing observation was made regarding country names. Some names, which were prevalent in the corpus, like *lett* ('Latvian', also meaning 'became'), *észt* ('Estonian' and the accusative form of 'wit'), and *ír* ('Irish' and 'writes'), were missing from the otherwise comprehensive list of Héja and Ligeti-Nagy (2022). The closer inspection of the graph showed that these adjectives ended up as isolated nodes in the adjectival graph. Let us recall now that instead of conveying unrelated meanings, homonyms are best to conceive of as *semantic idiosincrasies* (1.3). In other words, in the case of homonymy, a word incidentally denotes two different meanings. If we put that in distributional terms, we find that two randomly related set of contexts appear in the vicinity of the homonym word. That is why the word2vec representations of such words end up as isolates in the graph, i.e. are far from all the other word2vec representations.

Consequently, according to our expectations, (adjectival) homonyms can be identified as subset of the isolate nodes in the induced graph G. This method is completely unsupervised and language independent. Moreover, an additional important feature of the proposed technique is its interpretability, which we consider a big advantage over more recent contextualized word representations. Therefore, according to our expectaions it could revolutionize the understanding and categorization of adjectives in context-sensitive language tasks.

3 Preliminary results

Our initial investigations revealed the 30 most frequent isolate adjectives can be classified into four main categories:

- Homonymy$_1$: Adjectives with unusual, multiple PoS categories (e.g., *egész* 'whole', 'entire', 'complete', 'total', 'all'; *igaz* 'truthful', 'right', 'true', 'genuine', 'valid', 'OK', etc.).
- Homonymy$_2$: Part-of-speech changers (e.g. *eső* 'falling' vs. 'rain'; *lett* 'Latvian' vs. 'became'; *szilárd* 'solid' vs. a male name);
- Homonymy$_3$: Adjectival homonymy (e.g., *rendes* 'decent' vs. 'usual').
- Monosemic adjectives: derived from postpositions with the derivational suffix $-i$ (*nélküli* 'without', *iránti* 'forward').

This implies that our hypothesis is correct when examining sufficiently frequent adjectives. Based on these findings, we plan to nearly automate the creation of an adjectival WiC corpus containing monosemic and homonymous adjectives, appearing in sentence pairs with the same or different meanings. The corpus construction will include the following steps:

1. Creating the adjectival graph G based on adjectival embeddings.
2. Determining which frequently occurring adjectives form isolated nodes.
3. Identifying nominal contexts of these isolated nodes.
4. Demarcating adjectival meanings through the clustering of nominal contexts.
5. Identifying nouns that distinctly trigger the same meanings within a group and different meanings across groups.
6. Extracting relevant sentence pairs from the corpus for binary classification tasks.
7. Validating the corpus by human annotators.

4 Conclusion

The present research focuses on a novel, data-driven method that aims to grasp semantic idiosyncrasies: homonyms. The proposed unsupervised algorithm has the potential to be easily applicable to other languages as well. Moreover, the resulting WiC benchmark corpus is expected to give a more precise picture on the performance of large language models on the corresponding task. Our approach promises to offer novel insights into lexical semantics as well, which in turn, could greatly assist lexicographers, linguists, and the NLP community at large by bridging the gap between theoretical lexical semantics and practical NLP applications, paving the way for more interpretable language understanding systems.

[2] Maximally connected subgraphs are referred to as cliques.

References

Arleta Adamska-Sałaciak. 2006. *Meaning and the Bilingual Dictionary: The Case of English and Polish*. Peter Lang, Frankfurt.

Ron Artstein and Massimo Poesio. 2008. Survey article: Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4):555–596.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *CoRR*, abs/2005.14165.

Enikő Héja and Noémi Ligeti-Nagy. 2022. A clique-based graphical approach to detect interpretable adjectival senses in Hungarian. In *Proceedings of TextGraphs-16: Graph-based Methods for Natural Language Processing*, pages 35–43, Gyeongju, Republic of Korea. Association for Computational Linguistics.

Elisabetta Ježek. 2016. *The Lexicon: An Introduction*. Oxford Textbooks in Linguistics. Oxford University Press, Oxford.

Md Tahmid Rahman Laskar, M Saiful Bari, Mizanur Rahman, Md Amran Hossen Bhuiyan, Shafiq Joty, and Jimmy Huang. 2023. A systematic study and comprehensive evaluation of ChatGPT on benchmark datasets. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 431–469, Toronto, Canada. Association for Computational Linguistics.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient Estimation of Word Representations in Vector Space.

Tomas Mikolov, Ilya Sutskever, Kai Chen, G.s Corrado, and Jeffrey Dean. 2013b. Distributed Representations of Words and Phrases and their Compositionality. *Advances in Neural Information Processing Systems*, 26.

Mohammad Taher Pilehvar. 2019. On the importance of distinguishing word meaning representations: A case study on reverse dictionary mapping. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2151–2156, Minneapolis, Minnesota. Association for Computational Linguistics.

Mohammad Taher Pilehvar and Jose Camacho-Collados. 2019. WiC: the Word-in-Context Dataset for Evaluating Context-Sensitive Meaning Representations. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1267–1273, Minneapolis, Minnesota. Association for Computational Linguistics.

Chengwei Qin, Aston Zhang, Zhuosheng Zhang, Jiaao Chen, Michihiro Yasunaga, and Diyi Yang. 2023. Is chatgpt a general-purpose natural language processing task solver?

Jean Véronis. 2003. Sense tagging: does it make sense? In *Corpus Linguistics by the Lune: a festschrift for Geoffrey Leech*, Frankfurt. Peter Lang.

Alex Wang, Yada Pruksachatkun, Nikita Nangia, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2020. SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems.

Zhun Yang, Adam Ishay, and Joohyung Lee. 2023. Coupling large language models with logic programming for robust and general reasoning from text. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5186–5219, Toronto, Canada. Association for Computational Linguistics.