How well do distributed representations convey contextual lexical semantics: a Thesis Proposal

Zhu Liu
Tsinghua University
Computational Linguistics, School of Humanities
liuzhu22@mails.tsinghua.edu

Abstract

Modern neural networks (NNs) trained on large-scale raw sentence data build distributed representations, compressing individual words into dense, continuous, high-dimensional vectors. These representations are specifically designed to capture the meaning of word occurrence within context. In this thesis, we aim to investigate how well distributed representations from NNs encode lexical meaning. Initially, we identify four linguistic dimensions—homonymy, polysemy, semantic roles, and multifunctionality—based on the relatedness and similarity of meanings influenced by context. Subsequently, we intend to assess these dimensions by gathering or constructing datasets, utilizing various language models, and employing linguistic analysis tools.

1 Introduction

A word, functioning as a linguistic signifier, exhibits a complex mapping to its corresponding meanings or concepts, known as the signified (De Saussure, 2004). Words can manifest as homonymous, polysemous, or context-specific including nuances like vagueness and grammatical functions (Geeraerts, 2017), with varying degrees of similarity between their senses. Human cognition, equipped with a "mind model," effortlessly grasps these variations, comprehending both individual components and the overall sentence with remarkable fluidity, often at a subconscious level, even when the internal representation of meaning remains implicit. In our study, we shift focus to computational language models, particularly emphasizing their ability to discern context-sensitive meaning. Can these models adeptly capture the intricate nuances of word meaning within their respective contexts?

Thesis Proposal In addressing this inquiry, we initially identify various linguistic dimensions pertaining to contextual lexical meaning (refer to Section 2). Subsequently, we gather datasets and devise experiments to assess each dimension (refer to Section 3). Finally, we deliberate on potential methodological challenges and draw conclusions within this thesis (refer to Section 4).

2 Identification of Linguistic Dimensions

Our thesis delves into the various interpretations of lexical semantics within a given context. Based on the similarity between different senses, we categorize them into four distinct categories: homonymy, polysemy, semantic roles, and multifunctionality, as depicted in Table 1.

- Homonymy: Words with different meanings but identical pronunciation, such as *bat* meaning both the animal and the sporting equipment.
- Polysemy: A single word having multiple related but slightly distinct senses[1], for instance, *face* referring to the visage of a person, a clock, or a building.
- Semantic Roles: In a typical Subject-Verb-Object (SVO) construction, the Subject (S) denotes an agent, the Verb (V) a transitive action, and the Object (O) a patient. However, varying contexts can alter the degree of agentivity, transitivity, and patientivity (Hopper and Thompson, 1980). This dimension targets at each usage of an occurrence, even within the same sense.
- Multifunctionality: Grammatical morphemes (or grams), including function words and affixes, exhibit considerable multifunctionality in their semantics (Haspelmath, 2003). In most cases, the functions are so fine-grained that each usage in a context owns a unique function.

[1] We refer to each enumerate item listed under a word as “sense”.

Aspects	Level	Related	Lexical items	Multilingual	Unit
Homonymy	Word	✗	✓	✗	content words
Polysemy	Sense	✓	✓	✗	content words
Semantic Roles	Usage	✓	✗	✓	content words from SVO
Multifunctionality	Usage	✓	✗	✓	function words and affix

Table 1: Different linguistic dimensions of multiple meanings for a word. We show their meaning level, relatedness, categorization as a lexical item, universality across languages, and the specific unit they target.

We emphasize that homonymy and polysemy are typically formalized as tasks within word sense disambiguation (WSD) and primarily target content words. These aspects are extensively evaluated across various WSD benchmarks (Raganato et al., 2017; Pilehvar and Camacho-Collados, 2019). However, semantic roles and multifunctionality have received less attention and lack multilingual benchmarks and datasets for evaluation.

3 Evaluation

We initially employ various types of language models to extract representations, followed by an evaluation of these representations for four linguistic dimensions.

3.1 Models and Representations (Done Work)

Drawing on the Distributional Hypothesis (Harris, 1954), language models utilize neural networks to derive continuous vectors from large-scale corpora. Models such as Word2Vec (Mikolov et al., 2013) and Glove (Pennington et al., 2014) generate static word representations, which do not differentiate between different senses in varying contexts. Conversely, transformer-based models (Kenton and Toutanova, 2019) acquire hierarchical and contextual representations. We examine the configurations of lexical semantics derived from two types of models: BERT-like bidirectional models and GPT-like generative models.

In our previous work (Liu et al., 2024), we mainly explore how large language models (LLMs) encode lexical semantics. LLMs have achieved remarkable success in general language understanding tasks. However, as a family of generative methods with the objective of next token prediction, the semantic evolution with the depth of these models are not fully explored, unlike their predecessors, such as BERT-like architectures. In the paper, we specifically investigate the bottom-up evolution of lexical semantics for a popular LLM, namely Llama2 (Touvron et al., 2023), by probing its hidden states at the end of each layer using a contextualized word identification task (WiC (Pilehvar and Camacho-Collados, 2019)). Our experiments show that the representations in lower layers encode lexical semantics, while the higher layers, with weaker semantic induction, are responsible for prediction. This is in contrast to models with discriminative objectives, such as mask language modeling, where the higher layers obtain better lexical semantics. The conclusion is further supported by the monotonic increase in performance via the hidden states for the last meaningless symbols, such as punctuation, in the prompting strategy.

3.2 WSD with Uncertainty (Done work)

Word sense disambiguation (WSD), which aims to determine an appropriate sense for a target word given its context, is crucial for natural language understanding. Existing supervised methods treat WSD as a classification task and have achieved remarkable performance. However, they ignore uncertainty estimation (UE) in the real-world setting, where the data is always noisy and out of distribution. Our paper (Liu and Liu, 2023) extensively studies UE on the benchmark designed for WSD. Specifically, we first compare four uncertainty scores for a state-of-the-art WSD model and verify that the conventional predictive probabilities obtained at the final layer of the model are inadequate to quantify uncertainty. Then, we examine the capability of capturing data and model uncertainties by the model with the selected UE score on well-designed test scenarios and discover that the model adequately reflects data uncertainty but underestimates model uncertainty. Furthermore, we explore numerous lexical properties that intrinsically affect data uncertainty and provide a detailed analysis of four critical aspects: the syntactic category, morphology, sense granularity, and semantic relations.

Our work suggests that the representation of meaning can be conceptualized as a probabilistic inference, wherein meaning functions as a random variable conditioned on both the data and the model, rather than as a deterministic variable. This perspective aligns with the inherent qualities of language, including underspecification, vagueness, and context sensitivity (Sennet, 2023).

3.3 Semantic Roles (Ongoing Work)

In the basic SVO structure, the degree of semantic roles for different grammatical constituents (S, V, and O) can vary. This phenomenon is linguistically universal. For instance, *shot at* in *The hunter shot at the bear* exhibits weaker transitivity than *shot* in *The hunter shot the bear.* In Chinese, word order contributes to semantic roles; typically, the first argument is an agent, and the second is a patient. Alternating subject and object can lead to changes in agency and objectivity, although in some cases of alternation, semantic roles remain unchanged. For example, *Zhangsan Da LiSi* (Tom hit Alice) is totally different from *LiSi Da Zhangsan* (Alice hit Tom). But *Shigeren Chi Yidunfan* (Ten people eat a meal) basically has the same meaning as *Yidunfan Chi Shigeren* (A meal provides ten people to eat). In the former case, hit in two contexts has the same transitivity while eat does not. We propose collecting minimal pairs (where only S and O change) representing different semantic role changes and investigating whether language models can capture such nuances. The forms of minimal pairs may vary across different languages because distinct languages may employ different methods, such as morphological or syntactic, to reflect changes in semantic roles. We highlight that the form of minimal pair is specific to a certain language, which may have different ways to change the degree of semantic roles.

This work is related to an ancient NLP task: semantic role labeling (Jurafsky and Martin, 2020) but has several differences. First, we regard the degree of semantic roles as a continuous variable, not a binary choice. Thus, we leverage representations to calculate the similarity between corresponding items. Second, we adopt psycholinguisic diagnostics (Ettinger, 2020) for language models by designing tests in a controlled manner. Third, we consider cross-lingual universality and compare behaviors of different languages.

3.4 Multifunctionality (Ongoing Work)

Function words and affixes exhibit a broader range of nuanced semantics or functions compared to content words, often not exhaustively listed in dictionaries. For instance, repetitive grams (such as "and" and "again") in various languages demonstrate over 20 functions (Zhang, 2017). Linguists employ Semantic Map Models (SMM) (Haspelmath, 2003) to visually represent these functions in conceptual/semantic space, interconnected by lines to form a network. In this network, functions with greater similarity are positioned closer together on the map. SMM is grounded in cross-linguistic comparison, guided by the "semantic connectivity hypothesis," which posits that functions expressed by a language-specific category should occupy contiguous areas on the semantic map. Our approach involves utilizing representations from language models to measure the similarity between different occurrences of the target word. Subsequently, we design a graph algorithm to construct the semantic map, adhering to the connectivity principle. We intend to assess the quality of the automatic graph against the human-annotated one using designated metrics.

4 Conclusion and Challenges

Distributed representations encode rich lexical semantics, encompassing not only the word itself but also its contextual associations. Our thesis aims to assess the extent to which vectorized intermediate representations capture word meaning. We explore two conventional meaning relations: homonymy and polysemy, as well as two more nuanced relations from the perspective of linguistic typology: Semantic Roles and Multifunctionality. These aspects are investigated across distinct model architectures, utilizing common benchmarks or constructing well-designed datasets and linguistic tools (e.g., Semantic Map Model). The performance on these tasks serves as an indicator of the quality of the representation concerning each aspect.

However, this probing methodology raises several concerns and challenges. First is the “evaluation dilemma.” That is to say, can we confidently assert that the model fails to capture semantics entirely when it performs poorly on a specific task? The results may be influenced by various factors, such as suboptimal strategies for extracting representations. In other words, we cannot definitively conclude that representations are incapable of reflecting lexical semantics without isolating other influencing factors, which theoretically presents a vast search space. The second issue pertains to dataset bias. Contextual meaning is inherently more subjective than static, context-free meaning. Consequently, human annotators may exhibit personal preferences and disagreement on certain judgments. Such uncertainty can impact metric design, the reliability of gold labels, and other aspects. A comprehensive evaluation should therefore consider and address uncertainty.

Modern neural networks derive their power from the "scaling law", where increases in data, model size, and computational resources lead to improved performance. However, this advantage comes with a drawback: opacity. It raises the question, does the model truly understand meaning? Our research aims to enhance the interpretability of contemporary language models, aiming to bridge the divide between the computer science and linguistics communities. By doing so, we hope to foster a deeper understanding of how these models process and represent linguistic information.

References

Ferdinand De Saussure. 2004. Course in general linguistics. *Literary theory: An anthology*, 2:59–71.

Allyson Ettinger. 2020. What bert is not: Lessons from a new suite of psycholinguistic diagnostics for language models. *Transactions of the Association for Computational Linguistics*, 8:34–48.

Dirk Geeraerts. 2017. Lexical semantics.

Zellig S Harris. 1954. Distributional structure. *Word*, 10(2-3):146–162.

Martin Haspelmath. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In *The new psychology of language*, pages 211–242. Psychology Press.

Paul J Hopper and Sandra A Thompson. 1980. Transitivity in grammar and discourse. *language*, pages 251–299.

Daniel Jurafsky and James H Martin. 2020. Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition.

Jacob Devlin Ming-Wei Chang Kenton and Lee Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT*, pages 4171–4186.

Zhu Liu, Cunliang Kong, Ying Liu, and Maosong Sun. 2024. Fantastic semantics and where to find them: Investigating which layers of generative llms reflect lexical semantics. *arXiv preprint arXiv:2403.01509*.

Zhu Liu and Ying Liu. 2023. Ambiguity meets uncertainty: Investigating uncertainty estimation for word sense disambiguation. In *The 61st Annual Meeting Of The Association For Computational Linguistics*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. 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 NAACL-HLT*, pages 1267–1273.

Alessandro Raganato, Jose Camacho-Collados, and Roberto Navigli. 2017. Word sense disambiguation: A unified evaluation framework and empirical comparison. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 99–110.

Adam Sennet. 2023. Ambiguity. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*, Summer 2023 edition. Metaphysics Research Lab, Stanford University.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Ying Zhang. 2017. Semantic map approach to universals of conceptual correlations: a study on multifunctional repetitive grams. *Lingua Sinica*, 3(1):7.