

Analysis of Lexical Ambiguity in Vector Space Models

Marta Vázquez Abuín

Centro Singular de Investigación en Tecnoloxías Intelixentes (CITIUS)
Universidade de Santiago de Compostela
15782, Santiago de Compostela, Galicia/Spain

Abstract

The aim of this PhD is to analyze the capabilities of current vector models to deal with lexical ambiguity, particularly in the context of polysemy and homonymy, with a special focus on how models based on Transformer architectures represent these semantic phenomena. To achieve the main objective we have set the following specific objectives: (i) to perform a comprehensive analysis of the state-of-the-art, (ii) to compile a training and evaluation dataset according to the Word in Context format, and (iii) to extend it with data from other languages, with the potential to create a cross-lingual dataset including false friends and equivalent forms. In addition, (iv) to evaluate the computational models and (v, vi) to conduct experiments to compare the results obtained above with human judgments. To do so, we will compile and create datasets to perform an evaluation in terms of lexical ambiguity. We will assess state-of-the-art models to see if they are able to identify the different meanings of ambiguous words in context. Finally, the human judgments on the same dataset will be evaluated and compared with the computational models that have been analyzed. This analysis will be carried out mainly for the Galician language, but also for Portuguese and Spanish according to the evolution of the research.

Keywords

lexical semantics, distributional semantics, Word Sense Disambiguation, Galician

1. Introduction and Motivation

The development of deep learning and transformer-based models in recent years has been a huge improvement for the field of Natural Language Processing (NLP) [1, 2]. Nevertheless, the identification and processing of lexical ambiguity remains a significant challenge for linguistic technologies, particularly in the context of automatic and unsupervised approaches [3, 4, 5]. One strategy for evaluating the ability of computational models to address lexical ambiguity is through Word Sense Disambiguation (WSD) tasks [6]. The objective of these tasks is to discern the precise meaning of a word within diverse contextual settings, analyze its behavior, and identify potential solutions [7, 8, 9].

In the computational modelling of lexical semantics, there are two main approaches, both inspired by theoretical proposals [10]: (a) symbolic modelling, where each element has a specific meaning in a network of semantic relations, and (b) continuous approaches, where lexical forms are represented in a vector space. In the first case, we can highlight the case of Princeton WordNet (PWN) [11], which is one of the most frequently used lexical resources in the field of natural language processing. WordNet is a database of semantic relations that groups words into synonym sets (or synsets) and links them to other words according to their shared meaning [11]. These synsets act as nodes within the semantic networks, with the relations between them represented as edges [11, 12]. Furthermore, it provides glosses and examples that help users to illustrate word usages and clarify their meanings in context [12]. In continuous approaches, we can include both approaches based on the distributional hypothesis [13, 14] and new models trained with deep learning architectures [15]. In this context, we can differentiate between two categories: static models such as Word2Vec [16], fastText [17] and GloVe [18], where each word is represented by a single vector with its meaning in a specific context; and models based on language models, which present each occurrence of a word in a given context with a different vector that can potentially disambiguate the meaning of the word in context (BERT [1], ELMo [15] or GPT-3 [19]).

Doctoral Symposium on Natural Language Processing, 26 September 2024, Valladolid, Spain

✉ martavazquez.abuin@usc.gal (M. Vázquez Abuín)

ORCID [0000-0002-2134-9493](https://orcid.org/0000-0002-2134-9493) (M. Vázquez Abuín)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Lexical ambiguity phenomena, such as polysemy or homonymy, are pervasive across natural languages and present a challenge to computational models as a single word form may have different meanings depending on the context [3, 4, 5]. In the context of this thesis, it is appropriate to describe these two phenomena. Polysemy involves a single lexeme with several related meanings depending on the context (e.g., the Galician *coche* ‘car’ referring to a four-wheeled vehicle or a train carriage). Homonymy, on the other hand, involves a single lexical form consisting of different lexemes with multiple independent meanings (e.g. the Galician *canto* as one of the divisions of an epic poem ‘canto’ or as the gap between two walls ‘edge’) [20].

For English, and other languages like Italian and German, there exists a comprehensive array of resources for the training and evaluation of the resolution of lexical ambiguity in linguistic models. Some examples include the different WordNets and Word in Context datasets, such as WiC [21], XL-WiC [22] or ConSec [23]). In contrast, the availability of these resources is very limited for under-resourced languages such as Galician. Furthermore, the representation of Galician in comparison with the other languages analyzed is often very small, as evidenced by the XL-WSD dataset [24].

The objective of this PhD is to examine the strategies followed by languages models in the context of lexical ambiguity in Galician¹.

The remainder of this work is organized as follows: Section 2 is dedicated to an examination of the diverse tasks and challenges associated with lexical ambiguity and word sense disambiguation. Sections 3 and 4 present the objectives, hypotheses, and proposed research methodology. Finally, section 5 presents the preliminary results of the research.

2. Related work

Lexical ambiguity is very important in human language understanding [5] and one of the most common problems in natural language technologies because it causes misinterpretation of natural language. As a result, identifying and resolving ambiguity is essential for improving the efficiency and reliability of these technologies [6, 25].

We talk about disambiguation to describe the process of resolving ambiguity errors where WSD is one of the most studied aspects by researchers [6]. The main objective of the study of this aspect is to determine the word sense in a particular context, such as associating words in context with the most appropriate entry in a predefined sense inventory [7].

One of the most used linguistic resources for lexical disambiguation is WordNet [11]. This lexical database has become the primary repository for senses in NLP and it is the source of the majority of the datasets and evaluation frameworks for WSD [7]. WordNet was initially developed specifically for the English language. However, over time, other languages began to develop their own lexical resources based on it, for example, Galnet [26] (The Galician WordNet [27]) that is part of the Multilingual Central Repository [28]. The creation of such resources represents a significant challenge in the majority of languages, especially for languages with limited linguistic resources, where the investment of time and human effort is often a significant obstacle. While new automatic and unsupervised methods have been developed for this purpose, it is also necessary to ensure that they are trained with the quality and quantity of data to obtain adequate results. For this reason, the most popular option has been to extend them with PWN using different methods adapted to their developments and needs, looking for automatic and unsupervised approaches that minimize human and economic effort.

However, there is still a notable disparity between English and other languages in terms of NLP tools and improvements, as well as the identification and treatment of lexical ambiguity. The gap is particularly significant when attempting to conduct research and work with under-resourced languages, such as Galician.

To assess the efficacy of automatic lexical disambiguation, WordNet has been employed to construct datasets such as WiC [21] to evaluate vector models in a particular context and its extended version

¹During the course of the thesis, we will also be able to carry out comparative analyses between other languages such as Portuguese and Spanish.

XL-WiC [22]. Each instance in the dataset will comprise a target word and two contexts, each with a specific meaning of the target word. The objective is to identify whether the two contexts of the word have the same meaning or not. However, Galician is not included in this analysis due to the limited availability of examples [22].

Regarding the models that will be evaluated, there are the static models, inspired by the distributional hypothesis, where each vector represents all the meanings of a word [14, 16] and the new models based on the deep neural networks [9, 29] that have been a revolution in the field [2]. These language models can model the meaning of words in context since different vectors are generated for each word in each sentence. Among the most used ones for these tasks, we can mention BERT [1], XLM-RoBERTa [30], DeBERTa [31], ELMo[15] or GPT[19].

Concerning the behavior of the different models regarding lexical ambiguity, we can observe that contextualized models [9] are effective in identifying homonyms and distinguishing between meanings [29, 32, 33]. However, the accuracy of the results declines when the contexts are similar due to the fine-grained distinctions [8]. In Loureiro *et al* [8] we can see that Transformer models, specifically BERT, have high results and capture sense distinctions, even with few examples, grouping polysemous words according to sense.

Concerning the human interpretation of lexical ambiguity, some studies have demonstrated that large models are capable of performing in a manner that is similar to human preference [34, 35], but in Trott *et al* [36] they found that the human behavior cannot be explained purely by exposure to language statistics as in the models assessed.

For Galician, the number of resources and datasets created for lexical disambiguation tasks and models evaluation is very low [33]. This situation places Galician at a position of disadvantage in relation to other languages in terms of WSD tasks and evaluation. In this context, the present research proposes the design and creation of datasets to evaluate computational models for cases of lexical ambiguity, as well as the behavior in cases of inter-linguistic disambiguation, with special emphasis on false friends. Furthermore, we intend to examine and contrast the results with the human interpretation.

3. Objectives

The objective of our research proposal is to conduct an analysis of the behaviour of current vector models with regard to lexical ambiguity in Galician. In order to achieve the primary objective, the following secondary objectives have been proposed.

1. To review the current state-of-the-art in lexical ambiguity for single words and natural language processing
2. To define and compile training and evaluation datasets in Galician, and eventually in other languages (Portuguese and Spanish) following the Word in Context (WiC) format
3. To compile a cross-lingual dataset in Galician, Portuguese and Spanish with a focus on the analysis of false friends
4. To evaluate the performance of computational models in handling lexical ambiguity using the developed datasets
5. To evaluate human judgments on the same dataset as the computational models analyzed
6. To examine and compare the identification and recognition of the phenomena by humans and models

3.1. Research Questions

In the first stage of the PhD, the following questions were formulated to achieve the established objectives:

- **RQ1:** Is written language sufficient for the assessed models to learn how to resolve lexical ambiguity or, on the contrary, would other types of information (visual, speech, etc.) be

necessary?

H1: Our hypothesis is that it can be sufficient, but under optimal conditions concerning training and computing resources as in Loureiro *et al* [8].

- **RQ2:** Can distributional models resolve lexical ambiguity?

H2: We assume that it would differ depending on the semantic relationship and the assessed model.

- **RQ3:** Which models are better to resolve the ambiguity?

H3: Our hypothesis is that the new models based on deep neural networks perform better than static models [37, 5, 38].

- **RQ4:** Can the analyzed methods demonstrate an enhanced capacity to comprehend and process lexical ambiguity in a manner that transcends the limitations of human understanding?

H4: We assume that humans can perform better regarding lexical ambiguity than actual vector space models [37, 32], but we can have systematic errors [34]. However, in certain models, the representations utilized to achieve this are largely aligned with human intuitions [38].

4. Methodology

The research will be conducted in the following stages, each one with a specific methodology based on the actual state-of-the-art in computational linguistics and natural language process.

The initial stage of the research will entail the creation of a training and evaluation dataset based on WiC [21] for Galician. The dataset will be created, in accordance with the methodology proposed in the original [21], and the other languages of the XL-WiC [22], using Galnet² to obtain the context sentences. Despite the valuable insights it offers, the size of this resource is insufficient for training and evaluating WSD tools, particularly with regard to the number of example sentences [22], so it was necessary to work on expanding the examples and words. In relation to the computational models that will be assessed with this dataset, we will utilize models based on Transformer architectures, in addition to other proposed architectures throughout the study.

In regard to human behavior and understanding in relation to lexical ambiguity, the objective is to analyze how people identify some semantic phenomena and assess the limit at which one lemma changes its meaning to a new one, as well as the extent to which the senses of the same lexeme diverge from one another. These experiments will be conducted following the standard practices in the fields and will be guided in their performance of the tasks by the previously established methodology [39, 32].

Moreover, the intention is to create a cross-lingual dataset that will include other languages, such as Portuguese and Spanish. The details of the dataset have yet to be determined, but it will be structured in accordance with the format previously outlined. Each instance will contain a word (which should have the same or a similar lexical form, taking into account orthographic conventions in both languages) with two contexts: one for Galician and one for the other language (Portuguese or Spanish). The target word has a specific meaning in each context, which may or may not be the same. The selection of comparable lexical forms allows for an examination of the performance of the analyzed models in the context of false friends in closely related languages.

Finally, a quantitative analysis of the results will be conducted through an evaluation employing accuracy and correlation metrics.

5. Preliminary results

In the initial phase of the investigation, it was determined that the number of Galician examples was insufficient for the requisite WSD disambiguation tasks [22]. We decided to design a method [40] to

²<https://ilg.usc.gal/galnet/>

enhance the quantity of Galician examples by translating the English examples associated with a synset containing a Galician word in Galnet relying on the state-of-the-art of neural machine translation systems, specifically NOS-MT-OpenNMT-gl-en [41]. In some cases, however, the synset does not have an associated Galician word. Therefore, bilingual word embeddings were employed as probabilistic dictionaries to search for new words. The following procedure was utilized: the Wikipedia versions of each language were processed with FreeLing [42] for Galician and UDPipe [43] for English. Two monolingual *fastText* models [17] were trained for each language: with the two versions of the corpus: one lemmatized, and another representing each word as a *lemma_POS-tag* pair. Finally, we have mapped the monolingual models to a shared vector space. Then, we have designed and evaluated straightforward heuristics to expand Galnet to check if the new sentence can be added as a new one or not. Following the preliminary experiments, we have increased more than 4,5k synsets and 13k Galnet examples. These are being employed to construct WiC-type developes in Galician.

6. Acknowledgments

This project is carried out within the Research Group in Computational Linguistics (LComp, GI-2201), which is part of the Department of Spanish Language and Literature, Theory of Literature and General Linguistics of the University of Santiago de Compostela. Within the activity 2021-PG012 ‘Consolidación 2021 Modalidade C. Proxectos de excelencia - Exploración do coñecemento semántico en modelos vectoriais: homonimia, sinonimia, polisemia e idiomaticidade’, following the research line ‘Comprensión das linguas naturais: sintaxe e semántica computacionais’ (PIESP0027) in the Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS) with funding from a pre-doctoral grant from the Xunta de Galicia (ED481A-2024-070).

References

- [1] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, in: J. Burstein, C. Doran, T. Solorio (Eds.), 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), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186. URL: <https://aclanthology.org/N19-1423>. doi:10.18653/v1/N19-1423.
- [2] H. Zhang, M. O. Shafiq, Survey of transformers and towards ensemble learning using transformers for natural language processing, *Journal of big Data* 11 (2024) 25.
- [3] R. Navigli, Word sense disambiguation: A survey, *ACM computing surveys (CSUR)* 41 (2009) 1–69.
- [4] D. Loureiro, K. Rezaee, M. T. Pilehvar, J. Camacho-Collados, Analysis and Evaluation of Language Models for Word Sense Disambiguation, *Computational Linguistics* 47 (2021) 387–443. URL: https://doi.org/10.1162/coli_a_00405. doi:10.1162/coli_a_00405.
- [5] A. Liu, Z. Wu, J. Michael, A. Suhr, P. West, A. Koller, S. Swayamdipta, N. A. Smith, Y. Choi, We’re afraid language models aren’t modeling ambiguity, 2023. URL: <https://arxiv.org/abs/2304.14399>.
- [6] A. Yadav, A. Patel, M. Shah, A comprehensive review on resolving ambiguities in natural language processing, *AI Open* 2 (2021) 85–92. URL: <http://dx.doi.org/10.1016/j.aiopen.2021.05.001>. doi:10.1016/j.aiopen.2021.05.001.
- [7] A. Raganato, J. Camacho-Collados, R. Navigli, Word sense disambiguation: A unified evaluation framework and empirical comparison, in: M. Lapata, P. Blunsom, A. Koller (Eds.), Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, Association for Computational Linguistics, Valencia, Spain, 2017, pp. 99–110. URL: <https://aclanthology.org/E17-1010>.
- [8] D. Loureiro, K. Rezaee, M. T. Pilehvar, J. Camacho-Collados, Language models and word sense disambiguation: An overview and analysis, *CoRR abs/2008.11608* (2020). URL: <https://arxiv.org/abs/2008.11608>. arXiv:2008.11608.

- [9] M. Apidianaki, From Word Types to Tokens and Back: A Survey of Approaches to Word Meaning Representation and Interpretation, *Computational Linguistics* 49 (2022) 465–523. URL: https://doi.org/10.1162/coli_a_00474. doi:10.1162/coli_a_00474.
- [10] S. Trott, B. Bergen, Word meaning is both categorical and continuous., *Psychological Review* 130 (2023) 1239–1261. URL: <http://dx.doi.org/10.1037/rev0000420>. doi:10.1037/rev0000420.
- [11] C. Fellbaum, *WordNet: An Electronic Lexical Database*, American Psychological Association (APA), 1998. URL: <http://dx.doi.org/10.7551/mitpress/7287.001.0001>. doi:10.7551/mitpress/7287.001.0001.
- [12] G. A. Miller, C. Fellbaum, Wordnet then and now, *Language Resources and Evaluation* 41 (2007) 209–214. URL: <http://dx.doi.org/10.1007/s10579-007-9044-6>. doi:10.1007/s10579-007-9044-6.
- [13] M. Baroni, A. Lenci, Distributional memory: A general framework for corpus-based semantics, *Computational Linguistics* 36 (2010) 673–721. URL: <https://aclanthology.org/J10-4006>. doi:10.1162/coli_a_00016.
- [14] K. Erk, Vector space models of word meaning and phrase meaning: A survey, *Language and Linguistics Compass* 6 (2012) 635–653. URL: <http://dx.doi.org/10.1002/lnco.362>. doi:10.1002/lnco.362.
- [15] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, in: M. Walker, H. Ji, A. Stent (Eds.), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, Association for Computational Linguistics, New Orleans, Louisiana, 2018, pp. 2227–2237. URL: <https://aclanthology.org/N18-1202>. doi:10.18653/v1/N18-1202.
- [16] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient estimation of word representations in vector space, *Proceedings of Workshop at ICLR* (2013).
- [17] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching word vectors with subword information, *Transactions of the Association for Computational Linguistics* 5 (2017) 135–146. URL: <https://aclanthology.org/Q17-1010>. doi:10.1162/tacl_a_00051.
- [18] J. Pennington, R. Socher, C. D. Manning, Glove: Global vectors for word representation, in: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
- [19] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., *Language models are few-shot learners*, 2020.
- [20] D. A. Cruse, *Cambridge textbooks in linguistics: Lexical semantics*, Cambridge University Press, Cambridge, England, 1986.
- [21] M. T. Pilehvar, J. Camacho-Collados, WiC: the word-in-context dataset for evaluating context-sensitive meaning representations, in: J. Burstein, C. Doran, T. Solorio (Eds.), *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)*, Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 1267–1273. URL: <https://aclanthology.org/N19-1128>. doi:10.18653/v1/N19-1128.
- [22] A. Raganato, T. Pasini, J. Camacho-Collados, M. T. Pilehvar, XL-WiC: A multilingual benchmark for evaluating semantic contextualization, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Association for Computational Linguistics, Online, 2020, pp. 7193–7206. URL: <https://aclanthology.org/2020.emnlp-main.584>. doi:10.18653/v1/2020.emnlp-main.584.
- [23] E. Barba, L. Procopio, R. Navigli, ConSeC: Word sense disambiguation as continuous sense comprehension, in: M.-F. Moens, X. Huang, L. Specia, S. W.-t. Yih (Eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2021, pp. 1492–1503. URL: <https://aclanthology.org/2021.emnlp-main.112>. doi:10.18653/v1/2021.emnlp-main.112.
- [24] T. Pasini, A. Raganato, R. Navigli, XL-WSD: An extra-large and cross-lingual evaluation framework for word sense disambiguation., in: *Proc. of AAAI*, 2021, pp. 13648–13656.

- [25] M. Abeysiriwardana, D. Sumanathilaka, A survey on lexical ambiguity detection and word sense disambiguation, 2024. *arXiv:arXiv:2403.16129*.
- [26] M. A. Solla Portela, X. Guinovart, Galnet: o wordnet do galego. aplicacións lexicolóxicas e terminolóxicas, *Revista Galega de Filoloxía* 16 (2015) 169–201. URL: <http://dx.doi.org/10.17979/rgf.2015.16.0.1383>. doi:10.17979/rgf.2015.16.0.1383.
- [27] X. Gómez Guinovart, Galnet: Wordnet 3.0 do galego, *Linguamática* 3 (2011) 61–67. URL: <https://www.linguamatica.com/index.php/linguamatica/article/view/91>.
- [28] A. Gonzalez-Agirre, E. Laparra, G. Rigau, Multilingual central repository version 3.0, in: N. Calzolari, K. Choukri, T. Declerck, M. U. Doğan, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, European Language Resources Association (ELRA), Istanbul, Turkey, 2012, pp. 2525–2529. URL: http://www.lrec-conf.org/proceedings/lrec2012/pdf/293_Paper.pdf.
- [29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, volume 30, Curran Associates, Inc., 2017. URL: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.
- [30] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: D. Jurafsky, J. Chai, N. Schluter, J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Online, 2020, pp. 8440–8451. URL: <https://aclanthology.org/2020.acl-main.747>. doi:10.18653/v1/2020.acl-main.747.
- [31] P. He, X. Liu, J. Gao, W. Chen, Deberta: Decoding-enhanced bert with disentangled attention, *International Conference on Learning Representations* (2021).
- [32] S. Nair, M. Srinivasan, S. Meylan, Contextualized word embeddings encode aspects of human-like word sense knowledge, in: M. Zock, E. Chersoni, A. Lenci, E. Santus (Eds.), *Proceedings of the Workshop on the Cognitive Aspects of the Lexicon*, Association for Computational Linguistics, Online, 2020, pp. 129–141. URL: <https://aclanthology.org/2020.cogalex-1.16>.
- [33] M. Garcia, Exploring the representation of word meanings in context: A case study on homonymy and synonymy, in: C. Zong, F. Xia, W. Li, R. Navigli (Eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Association for Computational Linguistics, Online, 2021, pp. 3625–3640. URL: <https://aclanthology.org/2021.acl-long.281>. doi:10.18653/v1/2021.acl-long.281.
- [34] P. D. Rivière, A. L. Beatty-Martínez, S. Trott, Bidirectional transformer representations of (spanish) ambiguous words in context: A new lexical resource and empirical analysis, 2024. URL: <https://arxiv.org/abs/2406.14678>. *arXiv:2406.14678*.
- [35] G. Kamath, S. Schuster, S. Vajjala, S. Reddy, Scope ambiguities in large language models, *Transactions of the Association for Computational Linguistics* 12 (2024) 738–754.
- [36] S. Trott, C. Jones, T. Chang, J. Michaelov, B. Bergen, Do large language models know what humans know?, *Cognitive Science* 47 (2023) e13309.
- [37] G. Wiedemann, S. Remus, A. Chawla, C. Biemann, Does bert make any sense? interpretable word sense disambiguation with contextualized embeddings, in: *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019): Long Papers*, German Society for Computational Linguistics & Language Technology, Erlangen, Germany, 2019, pp. 161–170.
- [38] W. Liao, Z. Wang, K. Shum, A. B. Chan, J. Hsiao, Do large language models resolve semantic ambiguities in the same way as humans? the case of word segmentation in chinese sentence reading, in: *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 46, 2024.
- [39] J. Haber, M. Poesio, Word sense distance in human similarity judgements and contextualised word embeddings, in: C. Howes, S. Chatzikyriakidis, A. Ek, V. Somashekarappa (Eds.), *Proceedings of the Probability and Meaning Conference (PaM 2020)*, Association for Computational Linguistics,

- Gothenburg, 2020, pp. 128–145. URL: <https://aclanthology.org/2020.pam-1.17>.
- [40] M. Vázquez Abuín, M. Garcia, Wordnet expansion with bilingual word embeddings and neural machine translation, in: 23rd EPIA Conference on Artificial Intelligence, EPIA, Springer, 2024.
- [41] P. Gamallo, D. Bardanca, J. R. Pichel, M. Garcia, S. Rodríguez-Rey, I. de Dios-Flores, Nos_mt-opennmt-en-gl, <https://huggingface.co/proxectonos/NOS-MT-OpenNMT-en-gl>, 2023.
- [42] L. Padró, Analizadores multilingües en freeling, *Linguamática* 3 (2012) 13–20. URL: <https://linguamatica.com/index.php/linguamatica/article/view/115>.
- [43] M. Straka, UDPipe 2.0 prototype at CoNLL 2018 UD shared task, in: D. Zeman, J. Hajič (Eds.), *Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 197–207. URL: <https://aclanthology.org/K18-2020>. doi:10.18653/v1/K18-2020.