WordNet-based data augmentation for hybrid WSD models

Arkadiusz Janz and Marek Maziarz

Wroclaw University of Science and Technology, Poland
{arkadiusz.janz, marek.maziarz}@pwr.edu.pl

Wrocław
University
of Science
and Technology

Outline

- Introduction
- Related work
- Hybrid models
- Results
- Conclusions

Introduction: WSD and its limitations

- **Our task**: We aim at improving multi-lingual **Word Sense Disambiguation**
- **Methods**: Deep neural networks (DNNs) with cross-lingual language models
- **Rationale**:
  - Recent years: **a great improvement** has been achieved **with the use of DNNs**.
  - **The lack of large-scale sense annotated corpora** required by modern neural models for low-resourced languages: **still an open problem**.
  - The **large number of categories** is a serious limitation, because of the bottleneck of sense annotation sparseness.
  - **Constructing a large sense annotated corpus is a very laborious task**, so this problem affects NLP for most world languages.
  - The models trained on large WSD resources (i.e. SemCors and wordnet-based corpora) have to cope with a **huge number of senses that rarely occurr in texts**.

Introduction: The light at the end of the tunnel

Solutions:

1. **The usage of knowledge bases** facilitates WSD algorithm through propagating information within a semantic network (**hybrid models**).
2. **The use of pre-trained language agnostic models** allows to train on existing WSD resources and apply it to a new language context and partially solves data scarcity issue.
   - The problem: a limited capacity of deep neural networks and *negative transfer* phenomenon.

Introduction: Our contribution

1. **Main aim:** Enhance existing hybrid WSD models with data augmentation technique
    - The models utilize both the relational structure and text utterances.
    - **Question#1**: how to transfer relational structure from language to language?
    - **Question#2**: does the Polish sense inventory help in multilingual word sense disambiguation?
2. **Things done**: A slight yet effective modification of the **EWISER**[11] model
    - Data augmentation by **transfering internal wordnet structure**
    - Data augmentation by **transfering sense usage examples and glosses**
    - Extended the evaluation of EWISER model to more languages using **XL-WSD**[65] framework.

Related work

- Supervised models
  - Cross-lingual models based on multilingual transformers (MULAN [7], SensEmBERT[74], . . .)
    - structural properties of lexico-semantic networks used to be ignored in neural architectures
  - Hybrid supervised models enhanced with wordnet data and structure (**EWISER** [12], CONSEC [8], . . .)
    - utilising textual descriptions of senses together with their structural properties

Related work

- Benchmarks
  - ”Standard” monolingual framework for English language [71]
  - Multilingual benchmark from Semeval competitions (English, Spanish, French, German and Italian)
  - XL-WSD [65]: a multilingual benchmark built on wordnet-based inventories (18 languages)
    - built on the basis of Open Multilingual WordNet data and BabelNet resources
    - a platform to evaluate zero-shot WSD methods and crosslingual transfer

Hybrid Approaches: Models

EWISER:

- a supervised hybrid architecture utilising sense annotated corpora and knowledge base structure simultaneously,
- a transformer architecture with additional sense discrimination layer and structured logit mechanism injecting structural information into model during training,
- a baseline architecture for us.

The KEY idea:

- utilise existing wordnet links between senses to reinforce training procedure and incorporate logit scores of neighboring senses into scoring function of word's candidate meanings

Hybrid Approaches: Models

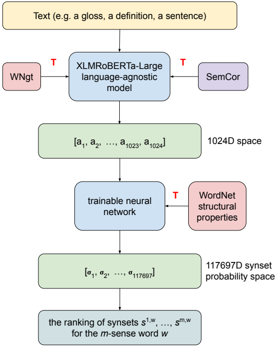

Figure: The architecture of EWISER model [11] from our perspective.

$$\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \dots & 0 \\ \dots & \dots & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

A

Figure: **Wordnet structural properties** as adjacency matrix. We aim at extending the matrix using non-English wordnets.

Benchmarking: XL-WSD Framework

Language	Type	#Instances
en	SemEval	8 062
bg	WN-based	9 968
ca	WN-based	1 947
da	WN-based	4 400
de	SemEval	862
es	SemEval	1 851
et	WN-based	1 999
eu	WN-based	1 580
fr	SemEval	1 160
gl	WN-based	2 561
hr	WN-based	6 333
hu	WN-based	4 428
nl	WN-based	4 400
sl	WN-based	2 032
zh	WN-based	9 568

Table: Language-specific test sets, their type and size as reported in [65] publication. SemEval datasets usually are easier to disambiguate when compared against WN-based datasets.

Hybrid Approaches: Polish data

- Polish WordNet (plWordNet) is heavily inter-linked with Princeton WordNet [73].
- 200k interlingual relation instances,
- inter-lingual synonymy, hyponymy and hypernymy were the most prominent.

Link type	Count
i-hyponyms	181 029
i-hypernyms	181 032
i-synonyms	93 654
Total	455 715

Table: Number of interlingual connections between plWordNet-3.2 and Princeton WordNet by category.

Hybrid approach: Augmenting PWN structure

Consider two pairs of counterpart synsets from plWN and PWN:

- $s_1^{plWN} \leftrightarrow \text{I-rel}\, s_1^{PWN}$,
- $s_2^{plWN} \leftrightarrow \text{I-rel}\, s_2^{PWN}$,

where "*I*-rel" signifies an inter-lingual relationship. Each time when there exists a short path between the two Polish synsets in plWN, we add a new link to PWN:

- $s_1^{PWN} \leftrightarrow s_2^{PWN}$

We assumed that for synonymous counterparts the distance should not exceed 2, while for homonymous counterparts the maximum path length was set to 1.

Hybrid approach: Augmenting PWN structure

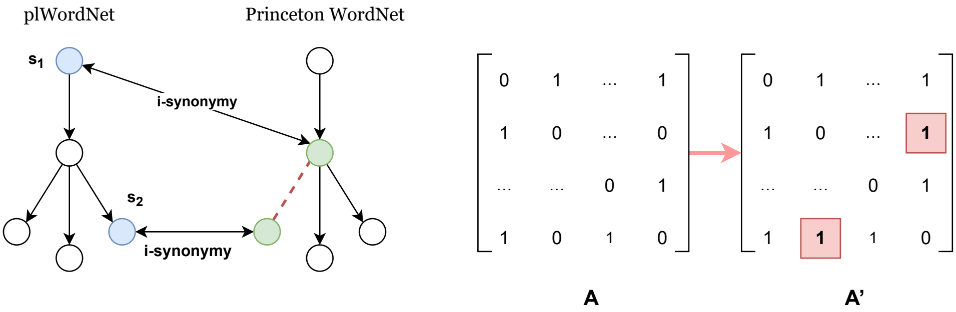

Figure: Structure enhancement procedure: i) find the paths between close synsets that are already mapped onto PWN structure ii) insert a link between senses on PWN side. The result: an updated adjacency matrix for EWISER model.

Hybrid approach: Augmenting PWN structure

Let's talk about separate sets:

1. I^{hyp} of all plWN synsets that have their *I*-hypernyms or *I*-hyponyms on the PWN side and
2. I^{syn} of all plWN synsets that have their *I*-synonyms in PWN.

Hybrid approach: Augmenting PWN structure

1. **I-synonymy case**
   - take the original plWN adjacency matrix $A = \{a_{ij}\}$,
   - produce $S = A^2$ (i.e. the matrix product of 2 copies of A),
     - its elements $\{s_{ij}\}$ are indexed by synset identifiers i, j,
     - they represent the number of random walks of length 2 on the plWN graph [41],
   - calculate $S' = \{s_{ij}\}$, set non-zero elements of the matrix to 1,
   - add $A + (S' - \mathbb{I}) = M = \{m_{ij}\}$
     - we get a matrix with new adjacency links (representing the distance of 2 or less steps in the original graph A),
   - Out of the matrix M we construct the new matrix E with picking up only those synsets that are in the set I^{syn}
     - i.e. $E = \{m_{ij}\}_{i,j \in I^{syn}}$.

Hybrid approach: Augmenting PWN structure

2. **I-hyponymy/I-hypernymy case**
   - take the original plWN adjacency matrix $A = \{a_{ij}\}$,
   - filter it leaving only the synsets from the set I^{hyp},
   - i.e. $H = \{a_{ij}\}_{i,j \in I^{hyp}}$.

Hybrid approach: Augmenting PWN data

- Taking into account all relationships obtainable from **matrices** H **and** E we finally land with the set of new links to be added to PWN.
- Nearly **146,000 Polish synsets** are described by a gloss and/or by (a) usage example(s). These samples were used to extend EWISER's training data.
- To obtain their textual descriptions we used **interlingual links** from plWordNet 3.2 including interlingual **synonymy, hyponymy and hypernymy**.

Hybrid approach: Augmenting textual data - a note

- In **XL-WSD** the authors **used machine translated PWN glosses and usage examples** and found **no significant improvement** over language-specific models.
- In contrast to their approach, **we used** Polish glosses and **native natural language examples** avoiding translation disadvantages.

Hybrid approach: Bilingual Training vs Negative Transfer

- Negative transfer: ***using source domain data undesirably reduces the learning performance in the target domain*** (limited model capacity).
- We believe multilingual downstream task fine-tuning might be beneficial for tasks such as WSD due to pre-training task (parallel corpora in XLM-R).
- However, for tasks such as POS tagging or NER recognition Negative Transfer issue (also called Negative Interference) was observed [80] in multilingual transfer.
- Our work is **one of the first attempts to investigate Negative Transfer phenomenon in WSD task**.

Hybrid approach: Training and testing

All models were trained on

- Princeton WordNet glosses and usage examples,
- SemCor corpus.

The models were tested on SemEval tasks and test data from XL-WSD (glosses and usage examples from several wordnets).

Hybrid approach: Experimental setting

- **Baseline#1**: a zero-shot architecture proposed in XL-WSD framework with XLMR-Large model.
- **Baseline#1** We re-train the EWISER architecture with XLMR-Large model as the underlying context encoder (comparability).
- **Model#1** Structure-only augmentation for EWISER model.
- **Model#2** Joint augmentation of structure and textual data (bilingual dataset).
- **Hyperparameter tuning**: finetuned on SemEval's 2015 validation dataset as it was proposed in the literature, early stopping, the experiments repeated 5 times.

Results

Language ISO 639-1	Baselines EWISER [e]	CONSEC [c]	XLM-R [x]	EWISER-augmented +PLWN (Es)	+PLWN (Es+Ts)
en+	78,9	**83.4**	76.3	79.9	79.6
bg	74,2	—	72.0	74.7	**75.4**
ca	53,6	—	50.0	54.2	**55.2**
da	82,6	—	80.6	82.8	**83.3**
de+	83,1	**84.2**	83.2	83.1	82.9
es+	77,0	77.4	75.8	77.4	**78.2**
et+	71,1	69.8	66.1	70.9	**71.5**
eu	50,2	—	47.2	50.5	**50.8**
fr+	83,8	84.4	83.9	83.9	**84.7**
gl	**67,7**	—	66.3	66.4	67.4
hr	74,1	—	72.3	74.2	**74.3**
hu	**73,7**	—	67.6	73.6	**73.7**
nl+	63,2	63.3	59.2	63.5	**64.1**
sl	66,6	—	68.4	**68.0**	67.5
zh	56,1	—	51.6	56.3	**56.5**
mean+	76.1	**77.0**	74.1	76.5	76.8
mean	70.3	—	68.0	70.6	**71.0**
median+	77.9	**80.4**	76.1	**78.5** [c] (=)	**78.6** [c] (=)
median	73.6	—	68.4	**73.6** [e] * (↑) [x] ** (↑)	**73.7** [e] ** (↑) [x] *** (↑)

Results

In tests on 15 languages **our technique turned out to be successful in beating the XL-WSD and the initial EWISER model** and comparable to some extent with the CONSEC[1] model.

[1] The evaluation of CONSEC model was limited to the results provided by the authors in [8]. At the time of publication, the training procedure was not fully reproducible and the codebase was incompatible with XL-WSD sense indices.

Results

- We compared average F1 performances using *U*-Mann-Whitney paired test (separately for CONSEC and for XL-WSD with EWISER).
- *p*-value correction for false discovery ratio – Benjamini-Hochberg procedure [10].
- Our two models performed better *on average* than XL-WSD (XLMR-L) and EWISER baseline models (for 15 languages) and not worse than CONSEC model (for 6 languages).

Conclusions & Future Work

- We proved that augmenting English training data sets with glosses and examples from other than English wordnets can lead to the improvement of a multilingual WSD algorithm.
- We plan to investigate new ways of enriching Princeton WordNet structure with relation instances derivable from Polish WordNet network.

Acknowledgements

This research was financed by the National Science Centre, Poland, grant number 2018/29/B/HS2/02919, and supported by the Polish Ministry of Education and Science, Project CLARIN-PL.

Bibliografia I

[1] January 2021.

[2] Eneko Agirre, Oier Lopez De Lacalle, Christiane Fellbaum, Shu-Kai Hsieh, Maurizio Tesconi, Monica Monachini, Piek Vossen, and Roxane Segers. Semeval-2010 task 17: All-words word sense disambiguation on a specific domain. In *Proceedings of the 5th international workshop on semantic evaluation*, pages 75–80, 2010.

[3] Eneko Agirre, Oier Lopez de Lacalle, and Aitor Soroa. Random walks for knowledge-based word sense disambiguation. *Computational Linguistics*, 40(1):57–84, 2014.

[4] Eneko Agirre, Oier López de Lacalle, and Aitor Soroa. The risk of sub-optimal use of open source NLP software: UKB is inadvertently state-of-the-art in knowledge-based WSD. In *Proceedings of Workshop for NLP Open Source Software (NLP-OSS)*, pages 29–33, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[5] Eneko Agirre and Aitor Soroa. Personalizing pagerank for word sense disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '09, pages 33–41, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[6] Satanjeev Banerjee and Ted Pedersen. Extended gloss overlaps as a measure of semantic relatedness. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, IJCAI'03, page 805–810, San Francisco, CA, USA, 2003. Morgan Kaufmann Publishers Inc.

Bibliografia II

[7] Edoardo Barba, Luigi Procopio, Niccolo Campolungo, Tommaso Pasini, and Roberto Navigli. Mulan: Multilingual label propagation for word sense disambiguation. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 3837–3844, 2021.

[8] Edoardo Barba, Luigi Procopio, and Roberto Navigli. Consec: Word sense disambiguation as continuous sense comprehension. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1492–1503, 2021.

[9] Pierpaolo Basile, Annalina Caputo, and Giovanni Semeraro. An enhanced Lesk word sense disambiguation algorithm through a distributional semantic model. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1591–1600, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics.

[10] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

[11] Michele Bevilacqua and Roberto Navigli. Breaking through the 80% glass ceiling: Raising the state of the art in word sense disambiguation by incorporating knowledge graph information. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2854–2864, 2020.

Bibliografia III

[12] Michele Bevilacqua and Roberto Navigli. Breaking through the 80% glass ceiling: Raising the state of the art in word sense disambiguation by incorporating knowledge graph information. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2854–2864, 2020.

[13] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(null):993–1022, March 2003.

[14] Terra Blevins and Luke Zettlemoyer. Moving down the long tail of word sense disambiguation with gloss-informed biencoders. *arXiv preprint arXiv:2005.02590*, 2020.

[15] Francis Bond and Ryan Foster. Linking and extending an open multilingual wordnet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1352–1362, 2013.

[16] Francis Bond and Kyonghee Paik. A survey of wordnets and their licenses. *Small*, 8(4):5, 2012.

[17] Jordan Boyd-Graber, David Blei, and Xiaojin Zhu. A topic model for word sense disambiguation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 1024–1033, Prague, Czech Republic, June 2007. Association for Computational Linguistics.

Bibliografia IV

[18] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, 1998.

[19] Bartosz Broda, Michał Marcińczuk, Marek Maziarz, Adam Radziszewski, and Adam Wardyński. Kpwr: Towards a free corpus of polish. In *Proceedings of LREC'12*, Istanbul, Turkey, 2012. ELRA.

[20] A. M. Butnaru and R. T. Ionescu. Shotgunwsd 2.0: An improved algorithm for global word sense disambiguation. *IEEE Access*, 7:120961–120975, 2019.

[21] Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors. *Proceedings of the Tenth International Conference on Language Resources and Evaluation, LREC 2016*, Portorož, Slovenia, 2016. ELRA, European Language Resources Association (ELRA).

[22] José Camacho-Collados, Mohammad Taher Pilehvar, and Roberto Navigli. Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities. *Artificial Intelligence*, 240:36–64, 2016.

[23] Devendra Singh Chaplot and Ruslan Salakhutdinov. Knowledge-based word sense disambiguation using topic models. In *32nd AAAI Conference on Artificial Intelligence (AAAI-18), New Orleans, USA*.

Bibliografia V

[24] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online, July 2020. Association for Computational Linguistics.

[25] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.

[26] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. 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 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[27] Jiaju Du, Fanchao Qi, and Maosong Sun. Using bert for word sense disambiguation. *arXiv preprint arXiv:1909.08358*, 2019.

[28] Agnieszka Dziob, Maciej Piasecki, and Ewa Rudnicka. plWordNet 4.1 - a linguistically motivated, corpus-based bilingual resource. In *Proceedings of the 10th Global Wordnet Conference*, pages 353–362, Wroclaw, Poland, July 2019. Global Wordnet Association.

Bibliografia VI

[29] Agnieszka Dziob, Maciej Piasecki, and Ewa K. Rudnicka. plWordNet 4.1 – a linguistically motivated, corpus-based bilingual resource. In *Proceedings of the Tenth Global Wordnet Conference : July 23-27, 2019, Wrocław (Poland)*, pages 353–362, 2019.

[30] Kawin Ethayarajh. How contextual are contextualized word representations? comparing the geometry of bert, elmo, and gpt-2 embeddings. *arXiv preprint arXiv:1909.00512*, 2019.

[31] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press, May 1998.

[32] John R Gilbert, Steve Reinhardt, and Viral B Shah. High-performance graph algorithms from parallel sparse matrices. In *International Workshop on Applied Parallel Computing*, pages 260–269. Springer, 2006.

[33] Yoan Gutiérrez, Sonia Vázquez, and Andrés Montoyo. A semantic framework for textual data enrichment. *Expert Systems with Applications*, 57:248 – 269, 2016.

[34] Yoan Gutiérrez, Sonia Vázquez, and Andrés Montoyo. Spreading semantic information by word sense disambiguation. *Knowledge-Based Systems*, 132:47 – 61, 2017.

[35] Elżbieta Hajnicz. Lexico-semantic annotation of *składnica* treebank by means of PLWN lexical units. In Orav et al. [62], pages 23–31.

Bibliografia VII

[36] Elżbieta Hajnicz, Anna Andrzejczuk, and Tomasz Bartosiak. Semantic layer of the valence dictionary of Polish *Walenty*. In Calzolari et al. [21], pages 2625–2632.

[37] Luyao Huang, Chi Sun, Xipeng Qiu, and Xuanjing Huang. GlossBERT: BERT for word sense disambiguation with gloss knowledge. *arXiv preprint arXiv:1908.07245*, 2019.

[38] Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli. Embeddings for word sense disambiguation: An evaluation study. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 897–907, 2016.

[39] Arkadiusz Janz and Maciej Piasecki. A weakly supervised word sense disambiguation for polish using rich lexical resources. *Poznan Studies in Contemporary Linguistics*, 55(2):339 – 365, 2019.

[40] Arkadiusz Janz and Maciej Piasecki. Word sense disambiguation based on constrained random walks in linked semantic networks. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 516–525, Varna, Bulgaria, September 2019. INCOMA Ltd.

[41] D. J. Kranda. The square of adjacency matrices, 2011.

[42] Sawan Kumar, Sharmistha Jat, Karan Saxena, and Partha Talukdar. Zero-shot word sense disambiguation using sense definition embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5670–5681, 2019.

Bibliografia VIII

[43] Sunjae Kwon, Dongsuk Oh, and Youngjoong Ko. Word sense disambiguation based on context selection using knowledge-based word similarity. *Information Processing & Management*, 58(4):102551, 2021.

[44] Dariusz Kłeczek. Polbert: Attacking polish nlp tasks with transformers. In Maciej Ogrodniczuk and Łukasz Kobyliński, editors, *Proceedings of the PolEval 2020 Workshop*. Institute of Computer Science, Polish Academy of Sciences, 2020.

[45] Helen Langone, Benjamin R. Haskell, and George A. Miller. Annotating WordNet. In *Proceedings of the Workshop Frontiers in Corpus Annotation at HLT-NAACL 2004*, pages 63–69, Boston, Massachusetts, USA, May 2 - May 7 2004. Association for Computational Linguistics.

[46] Quoc Le and Tomas Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*, ICML'14, page II–1188–II–1196. JMLR.org, 2014.

[47] Michael Lesk. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceeding SIGDOC '86 Proceedings of the 5th annual international conference on Systems documentation*, pages 24–26. ACM Press, 1986.

[48] Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. Multilingual denoising pre-training for neural machine translation. *Transactions of the Association for Computational Linguistics*, 8:726–742, 2020.

Bibliografia IX

[49] Fuli Luo, Tianyu Liu, Zexue He, Qiaolin Xia, Zhifang Sui, and Baobao Chang. Leveraging gloss knowledge in neural word sense disambiguation by hierarchical co-attention. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1402–1411, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

[50] Marco Maru, Federico Scozzafava, Federico Martelli, and Roberto Navigli. SyntagNet: Challenging supervised word sense disambiguation with lexical-semantic combinations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3525–3531, 2019.

[51] Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, Stan Szpakowicz, and Paweł Kedzia. plWordNet 3.0 – a comprehensive lexical-semantic resource. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2259–2268, Osaka, Japan, December 2016. The COLING 2016 Organizing Committee.

[52] Marek Maziarz, Maciej Piasecki, and Stanisław Szpakowicz. The chicken-and-egg problem in wordnet design: synonymy, synsets and constitutive relations. *Langauge Resources and Evaluation*, 47(3):769–796, 2013.

[53] Rada Mihalcea and Dan I. Moldovan. extended wordnet: progress report. In *in Proceedings of NAACL Workshop on WordNet and Other Lexical Resources*, pages 95–100, 2001.

Bibliografia X

[54] Rada Mihalcea, Paul Tarau, and Elizabeth Figa. PageRank on semantic networks, with application to Word Sense Disambiguation. In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics.

[55] George A. Miller, Claudia Leacock, Randee Tengi, and Ross T. Bunker. A semantic concordance. In *Proceedings of the Workshop on Human Language Technology*, HLT '93, page 303–308, USA, 1993. Association for Computational Linguistics.

[56] Andrea Moro and Roberto Navigli. SemEval-2015 task 13: Multilingual all-words sense disambiguation and entity linking. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 288–297, Denver, Colorado, June 2015. Association for Computational Linguistics.

[57] Andrea Moro, Alessandro Raganato, and Roberto Navigli. Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics*, 2:231–244, 2014.

[58] Roberto Navigli, David Jurgens, and Daniele Vannella. SemEval-2013 task 12: Multilingual word sense disambiguation. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 222–231, Atlanta, Georgia, USA, June 2013. Association for Computational Linguistics.

[59] Roberto Navigli and Simone Paolo Ponzetto. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250, 2012.

Bibliografia XI

[60] Maciej Ogrodniczuk and Łukasz Kobyliński, editors. *Proceedings of the PolEval 2020 Workshop*, Warsaw, Poland, 2020. Institute of Computer Science, Polish Academy of Sciences.

[61] Dongsuk Oh, Sunjae Kwon, Kyungsun Kim, and Youngjoong Ko. Word sense disambiguation based on word similarity calculation using word vector representation from a knowledge-based graph. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2704–2714, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.

[62] Heili Orav, Christiane Fellbaum, and Piek Vossen, editors. *Proceedings of the 7th International WordNet Conference (GWC 2014)*, Tartu, Estonia, 2014. University of Tartu.

[63] Martha Palmer, Christiane Fellbaum, Scott Cotton, Lauren Delfs, and Hoa Trang Dang. English tasks: All-words and verb lexical sample. In *Proceedings of SENSEVAL-2 Second International Workshop on Evaluating Word Sense Disambiguation Systems*, pages 21–24, Toulouse, France, July 2001. Association for Computational Linguistics.

[64] Tommaso Pasini and Roberto Navigli. Train-O-Matic: Large-scale supervised word sense disambiguation in multiple languages without manual training data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 78–88, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

Bibliografia XII

[65] Tommaso Pasini, Alessandro Raganato, and Roberto Navigli. Xl-wsd: An extra-large and cross-lingual evaluation framework for word sense disambiguation. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press, 2021.

[66] Adam Pease. *Ontology - A Practical Guide*. Articulate Software Press, 2011.

[67] Maciej Piasecki, Paweł Kędzia, and Marlena Orlińska. plWordNet in word sense disambiguation task. In Verginica Barbu Mititelu, Corina Forăscu, Christiane Fellbaum, and Piek Vossen, editors, *Proceedings of the 8th Global Wordnet Conference, Bucharest, 27-30 January 2016*, pages 280–289. Global Wordnet Association, 2016.

[68] Alexander Popov. Neural network models for word sense disambiguation: an overview. *Cybernetics and information technologies*, 18(1):139–151, 2018.

[69] Sameer Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. SemEval-2007 task-17: English lexical sample, SRL and all words. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 87–92, Prague, Czech Republic, June 2007. Association for Computational Linguistics.

[70] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67, 2020.

Bibliografia XIII

[71] Alessandro Raganato, Claudio Delli Bovi, and Roberto Navigli. Neural sequence learning models for word sense disambiguation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1156–1167, 2017.

[72] Alessandro Raganato, Jose Camacho-Collados, and Roberto Navigli. 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, 2017.

[73] Ewa Rudnicka, Marek Maziarz, Maciej Piasecki, and Stan Szpakowicz. A strategy of mapping polish wordnet onto princeton wordnet. In *Proceedings of COLING 2012: Posters*, pages 1039–1048, 2012.

[74] Bianca Scarlini, Tommaso Pasini, and Roberto Navigli. Sensembert: Context-enhanced sense embeddings for multilingual word sense disambiguation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):8758–8765, Apr. 2020.

[75] Federico Scozzafava, Marco Maru, Fabrizio Brignone, Giovanni Torrisi, and Roberto Navigli. Personalized PageRank with syntagmatic information for multilingual word sense disambiguation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 37–46, 2020.

Bibliografia XIV

[76] Benjamin Snyder and Martha Palmer. The English all-words task. In *Proceedings of SENSEVAL-3, the Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pages 41–43, Barcelona, Spain, July 2004. Association for Computational Linguistics.

[77] Rocco Tripodi and Marcello Pelillo. A game-theoretic approach to word sense disambiguation. *Computational Linguistics*, 43(1):31–70, April 2017.

[78] Ming Wang and Yinglin Wang. A synset relation-enhanced framework with a try-again mechanism for word sense disambiguation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6229–6240, Online, November 2020. Association for Computational Linguistics.

[79] Yinglin Wang, Ming Wang, and Hamido Fujita. Word sense disambiguation: A comprehensive knowledge exploitation framework. *Knowledge-Based Systems*, 190:105030, 2020.

[80] Zirui Wang, Zachary C Lipton, and Yulia Tsvetkov. On negative interference in multilingual models: Findings and a meta-learning treatment. *arXiv preprint arXiv:2010.03017*, 2020.

[81] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016.

Bibliografia XV

[82] C. Yang, Y. Wang, and J. D. Owens. Fast sparse matrix and sparse vector multiplication algorithm on the GPU. In *2015 IEEE International Parallel and Distributed Processing Symposium Workshop*, pages 841–847, 2015.