Enriching Word Usage Graphs with Cluster Definitions

Mariia Fedorova[1], Andrey Kutuzov[1], Nikolay Arefyev[1], Dominik Schlechtweg[2]
[1]University of Oslo, Norway; [2]University of Stuttgart, Germany
{mariiaf,andreku,nikolare}@ifi.uio.no; dominik.schlechtweg@ims.uni-stuttgart.de

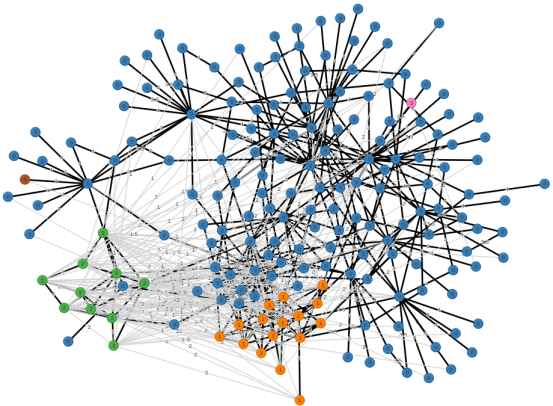
LREC-COLING 2024

May 2024

Contents

What are word usage graphs (WUGs)?

Inferred senses of the English noun ‘record’



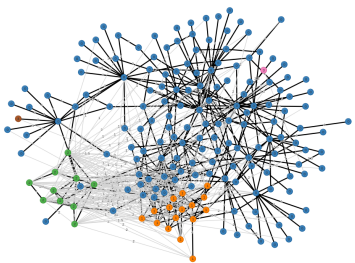
Usage examples of the first largest 3 senses

- 0: *The church records having been consumed when the parsonage house was burned*
- 1: *...a camera aboard the craft has surveyed the sky in the extreme ultraviolet and identified a record 385 bright sources in this wavelength band...*
- 2: *I was making records, but they weren’t getting played because there was no place to play them.*

Schlechtweg et al. (2020)

How to make a word usage graph interpretable?

WUG of the word record, Schlechtweg et al. (2020)



- 0: **Context**: *The church records having been consumed when the parsonage house was burned*
  Definition: A book or other piece of writing which contains information about past events.
- 1: **Context**: *...a camera aboard the craft has surveyed the sky in the extreme ultraviolet and identified a record 385 bright sources in this wavelength band...*
  Definition: Of a historical highest level or achievement.
- 2: **Context**: *I was making records, but they weren't getting played because there was no place to play them.*
  Definition: A medium for recording sound, especially a vinyl phonograph and gramophone disc.

Contents

Contributions

Contributions

- updated dataset: enriched (labeled) WUGs with cluster definitions are released

- defintion generation (DefGen) models for Norwegian, English, and Russian are released on Huggingface Hub

- useful for NLP practitioners, lexicographers, linguists

Contents

WUGs we use

Word usage graphs that we use

Language	Time periods	Words	Clusters	Clusters annotated	Diachronic?
English Schlechtweg et al. (2020)	1810-1860, 1960-2010	46	819	120	True
German Schlechtweg et al. (2020)	1800–1899, 1946–1990	50	488	95	True
Norwegian-1 Kutuzov et al. (2022)	1929-1965, 1970-2013	40	99	23	True
Norwegian-2 Kutuzov et al. (2022)	1980-1990, 2012-2019	40	78	17	True
Russian Aksenova et al. (2022)	modern	24	90	39	False

Main statistics of the enriched WUGs. We labeled and annotated only clusters which feature at least three usage examples.

Contents

Definition generation methods

Definition generation methods

- Lesk algorithm (Lesk, 1986): choice from WordNet
- GlossReader (Rachinskiy and Arefyev, 2021): choice from WordNet
- DefGen: fine-tuned multilingual mT0-xl (Giulianelli et al., 2023): generation from scratch

Definition generation methods

GlossReader (Rachinskiy and Arefyev, 2021)

- multilingual word sense disambiguation system
- consists of a context encoder and a gloss encoder trained jointly

To select a definition for a cluster:

1. gloss embeddings for all WordNet definitions are built with the gloss encoder;
2. for each usage from the cluster:
   - 2.1 its contextualized embedding is built with the context encoder,
   - 2.2 k most similar glosses[a] are retrieved from English WordNet [b];
3. the definition retrieved for the largest number of usages in this cluster is selected to represent the cluster.

---

[a] as measured by the dot product

[b] we select glosses among all 117K glosses, this makes the method applicable for target words absent in WordNet, including words in other languages

Definition generation methods

DefGen

Encoder-decoder Transformer-based language model fine-tuned on the task of contextualized definition generation Giulianelli et al. (2023).

Input: *'The church records having been consumed when the parsonage house was burned. What is the definition of record?'*

Output: **'A book or other piece of writing which contains information about past events.'**

Fine-tuning data

Language	Datasets
Norwegian	Bokmålsordboka
English	Oxford dictionary Gadetsky et al. (2018), Wordnet Ishiwatari et al. (2019), CoDWoe (Wiktionary) Mickus et al. (2022)
Russian	CoDWoe (Wiktionary) Mickus et al. (2022)

Definition generation methods

WordNet choice vs generation from scratch

- **Context**: *The church records having been consumed when the parsonage house was burned*
  Lesk: anything (such as a document or a phonograph record or a photograph) providing permanent evidence of or information about past events
  GlossReader: anything (such as a document or a phonograph record or a photograph) providing permanent evidence of or information about past events
  DefGen: A book or other piece of writing which contains information about past events.
- **Context**: *...a camera aboard the craft has surveyed the sky in the extreme ultraviolet and identified a record 385 bright sources in this wavelength band...*
  Lesk: sound recording consisting of a disk with a continuous groove; used to reproduce music by rotating while a phonograph needle tracks in the groove
  GlossReader: establish as the highest level or best performance
  DefGen: Of a historical highest level or achievement.

Definition generation methods

prompted with English for German without fine-tuning

Input: *'Ist eine Prüfung erforderlich, so erfolgt eine Entscheidung über den Antrag durch die zuständige Behörde. What is the definition of Entscheidung?' ('If an examination is necessary, a decision on the application will be made by the responsible authority. What is the definition of decision?')*

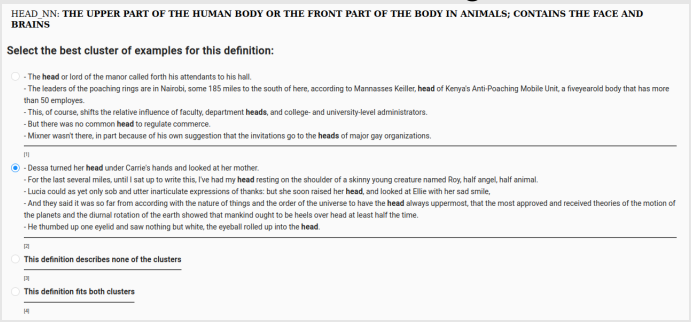
Output: *'The act of making up your mind about something; a decision.'*

Contents

Evaluation setup

Human evaluation as the only choice

- no existing benchmarking datasets for the task of WUG cluster labeling
- we invented and used a novel task: **guess the cluster by definition**



Krippendorff's $\alpha = 0.314$, fair agreement according to (Landis and Koch, 1977).

Contents

Results and analysis

Methods comparison

System	Accuracy, %	Fits both, %	Fits none, %
English DWUG, English definitions			
Lesk	21.67	5.00	53.33
GlossReader	50.00	9.17	37.50
DefGen	**69.17**	10.83	11.67
German DWUG, English definitions			
GlossReader	53.68	13.68	27.37
DefGen	**57.89**	16.84	12.63
Russian WUG RuDSI, English definitions			
GlossReader	64.10	10.26	17.95
DefGen	**71.79**	15.38	2.56

Results and analysis

DefGen only

Dataset	Definitions	Accuracy, %	Fits both, %	Fits none, %
Norwegian-1 DWUG	English	60.87	13.04	21.74
	Norwegian	73.91	4.35	21.74
Norwegian-2 DWUG	English	88.24	5.88	5.88
	Norwegian	76.47	11.76	11.76
Russian WUG RuDSI	Russian	48.72	7.69	15.38

Results and analysis

Again links to the models and enriched DWUGs

- defintion generation (DefGen) models for Norwegian, English, and Russian on Huggingface Hub

- enriched WUGs with cluster definitions

References I

Aksenova, A., Gavrishina, E., Rykov, E., and Kutuzov, A. (2022). RuDSI: Graph-based word sense induction dataset for Russian. In *Proceedings of TextGraphs-16: Graph-based Methods for Natural Language Processing*, pages 77–88, Gyeongju, Republic of Korea. Association for Computational Linguistics.

Gadetsky, A., Yakubovskiy, I., and Vetrov, D. (2018). Conditional generators of words definitions. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 266–271, Melbourne, Australia. Association for Computational Linguistics.

Giulianelli, M., Luden, I., Fernandez, R., and Kutuzov, A. (2023). Interpretable word sense representations via definition generation: The case of semantic change analysis. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3130–3148, Toronto, Canada. Association for Computational Linguistics.

References II

Ishiwatari, S., Hayashi, H., Yoshinaga, N., Neubig, G., Sato, S., Toyoda, M., and Kitsuregawa, M. (2019). Learning to describe unknown phrases with local and global contexts. 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 3467–3476, Minneapolis, Minnesota. Association for Computational Linguistics.

Kutuzov, A., Touileb, S., Mæhlum, P., Enstad, T., and Wittemann, A. (2022). NorDiaChange: Diachronic semantic change dataset for Norwegian. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2563–2572, Marseille, France. European Language Resources Association.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.

References III

Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on Systems documentation*, pages 24–26.

Mickus, T., Van Deemter, K., Constant, M., and Paperno, D. (2022). Semeval-2022 task 1: CODWOE – comparing dictionaries and word embeddings. In *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, pages 1–14, Seattle, United States. Association for Computational Linguistics.

Rachinskiy, M. and Arefyev, N. (2021). GlossReader at SemEval-2021 task 2: Reading definitions improves contextualized word embeddings. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 756–762, Online. Association for Computational Linguistics.

References IV

Schlechtweg, D., McGillivray, B., Hengchen, S., Dubossarsky, H., and Tahmasebi, N. (2020). SemEval-2020 task 1: Unsupervised lexical semantic change detection. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1–23, Barcelona (online). International Committee for Computational Linguistics.