Word Sense Disambiguation of French Lexicographical Examples Using Lexical Networks

Aman Sinha, Sandrine Ollinger, Mathieu Constant

ATILF, Université de Lorraine

WSD is a long standing research problem

Best deep learning models have performance less than 90%[1] for WSD.

- **Ping-pong I.1**
  - a sport activity
- **Ping-pong I.2**
  - the corresponding action (metonymic)
- **Ping-pong II**
  - an object used to play ping-pong (metonymic)
- **Ping-pong III**
  - an intellectual activity (metaphorical)

“ ***Il suit avec attention leur ping-pong argumentatif*** ”

(He carefully follows their argumentative ping-pong)

[1]http://nlpprogress.com/english/word_sense_disambiguation.html

- Several approaches includes supervised, unsupervised, knowledge-based and other mixed approaches (Navigli et. al. 2009)

- In our work, we focus on *knowledge-based approaches*.

  Some of the previous works in this direction:

  - Glosses (Huang et. al. 2019)
  - Sense embeddings (Kumar et. al. 2019)
  - Knowledge graphs (Bevilacqua and Navigli, 2020)

Lexical resources have always played a crucial role not only serving as sense inventories, but also as sources of information (Wilks and Stevenson, 1998)

- structure and lexical content of lexical networks (Agirre et. al. 2006)
- use of hypernym/hyponym/synonym relations (Kumar et al. 2019 ; Bevilacqua and Navigli 2020)
- implicit knowledge source from graph structure information of lexical networks along with pre-existing sense embeddings (Bevilacqua and Navigli, 2020)

Fr-LN[3] (Polguère, 2014), a formal model of the lexicon of contemporary French.

The complete fr-LN contains 29,220 word senses and 80,036 relations (LF-Arcs) between them.

DBLE-LN-fr : Collection of lexicographical usage examples.

Sources : Frantext[2] , FrWaC (Baroni et. al. 2009) , the Est-Républicain newspaper corpus (ATILF and CLLE, 2020).

Graph	#Word Senses	#Lemmas	#LF-Arcs	#LFs
Complete	29,220	18,400	62,641*	686
Verbs-only	5,237	2,559	9,854	399
Nouns-only	14,044	8,639	21,580	501

Table 1. Statistics on the fr-LN network.

[3]ORTOLANG platform: https://hdl.handle.net/11403/ examples-ls-fr/v2

[2]https://www.frantext.fr/

* Corresponds to paradigmatic and syntagmatic LFs only

Dataset: French Lexical Network (Fr-LN) and DBLE-LN-fr Database[1]

Based on the model of lexical systems (Polguère, 2014)

- The native structure of this resource is a graph
- It’s not a hierarchical graph, like WordNet
- All the edges are TYPED and ORIENTED
- All the edges have a semantic weight
- The resource is the result of a manual lexicographic work.

Corpus	#examples	#targets	#Word Senses	#Lemmas
Complete	31,131	51,347	27,343	17,161
Verbs-only	8,169	9,428	5,141	2,483
Nouns-only	19,644	27,105	13,601	8,131

Table 2. DBLE-LN-fr dataset

> Although WN has much larger coverage, it contains few relation types that are mainly paradigmatic relations whereas fr-LN contains various **syntagmatic, paradigmatic, copolysemic and phraseological** relations.

> fr-LN relations mainly involve **senses of different part-of-speech tags**, whereas WN relations quasi-exclusively involve nodes of the same part-of-speech. For instance, less than 6% of the relations involving verbs are between two verbs.

> Contrary to WN, fr-LN does not include glosses and the lexicographic definitions are still prototypical.

> fr-LN relations are associated with **semantic weights** depending to what extent the semantic content of the source node includes the semantic content of the target one.

EWISER : Neural WSD base + External Knowledge + Internal Knowledge

$$B = B_{-4} + B_{-3} + B_{-2} + B_{-1}$$
$$H_0 = \mathrm{BatchNorm}(B)$$
$$H_1 = \mathrm{swish}(H_0 W + \mathbf{b})$$

$$Z = H_1 O + \mathbf{b}$$

$$Q = Z A^T + Z$$

B_i : i^{th} BERT layers

O : Sense Embedding Matrix

A : Graph Adjacency Matrix

Our work : EWISER Improvisation

We don’t make use of Matrix *O* , but Matrix A

We removed the use of external pre-existing sense embedding matrix O, as our aim is to rely entirely on the **database of lexicographic examples** and the **French lexical network**

$$Z = H_1 O + \mathbf{b}$$
$$Q = ZA^T + Z$$

EWISER

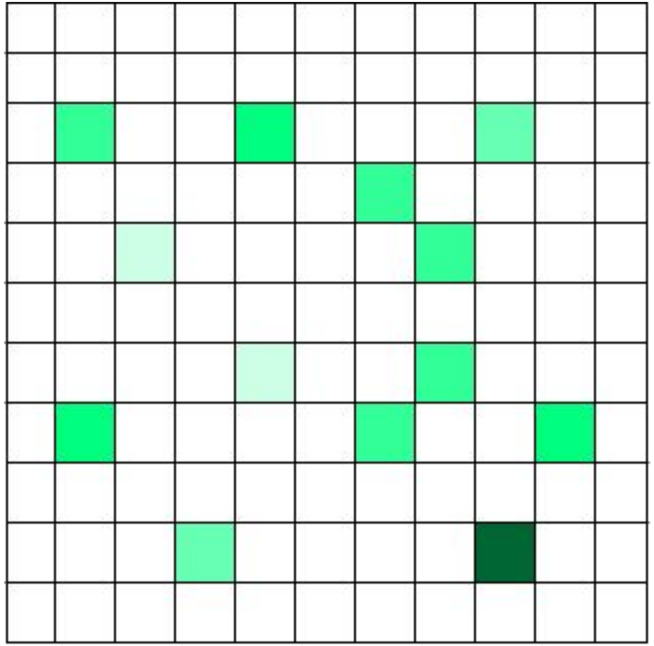
$$Q = H_1 A^T + H_1$$

Our work

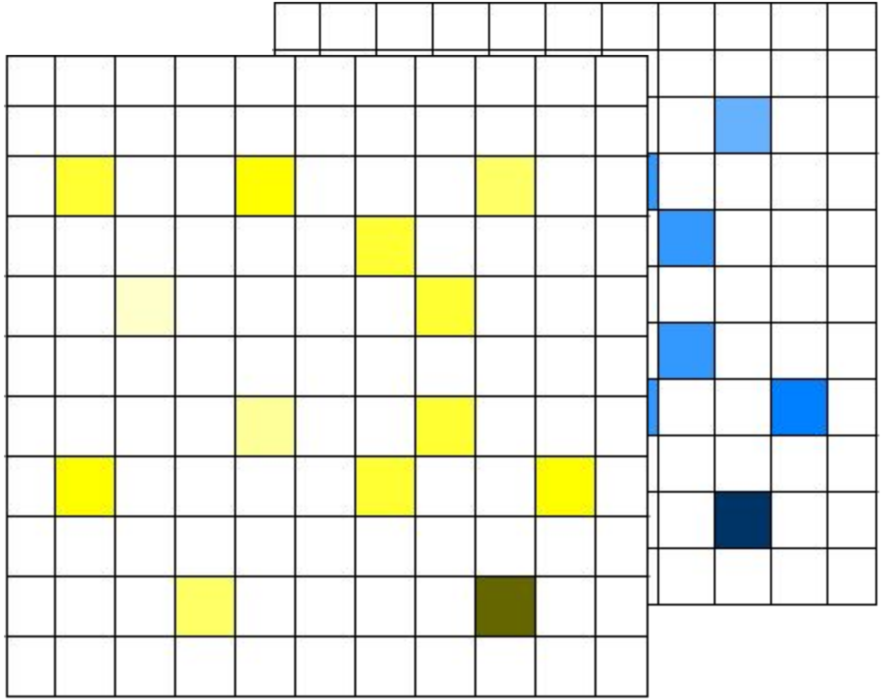
$$Z = H_1$$
$$Q = ZA^T + Z$$

Equivalence

We are interested in Graph Adjacency Matrix A



A^T



$A^T = A^T_{R1} + A^T_{R2} + A^T_{R3} + \dots + A^T_{Rn}$

STRUCT : $w(R) = 1$, R is a relation-function

SEM : $w(s) \in \{0,1,2\}$, s is a semantic strength information

Three variants :

[STRUCT/SEM] $A_{ij} := \Sigma\, w(R_k)$, where R_k is any edge between i and j; k = 1,2,3 …

[STRUCT/SEM]* $A_{ij} := \Sigma\, w(R_k)$, where $\Sigma(.)$ is trainable

[STRUCT/SEM]** $A_{ij} := \Sigma\, w(R_k)$, where $w(R_k)$ is trainable

- Frequency Baseline (Most FS /Least FS)
- Random Sense Baseline
- BARYCentre (cosine similarity of sense-representation)
- MLP (Neural Base (Bevilacqua and Navigli, 2020))

FlauBERT
(Vial et al. 2019)

System	VERB		NOUN	
	Dev	Test	Dev	Test
MFS	0.1145	0.1427	0.2026	0.2016
LFS	0.1178	0.1091	0.1973	0.1939
RS	0.1578	0.1654	0.2444	0.2357
BARYC.	0.3189	0.3178	0.5390	0.5454
MLP	0.2648	0.2822	0.5091	0.5163

CamemBERT
(Martin et al 2019)

Experiment Results

System	VERB		NOUN	
	Dev	Test	Dev	Test
MFS	0.1145	0.1427	0.2026	0.2016
LFS	0.1178	0.1091	0.1973	0.1939
RS	0.1578	0.1654	0.2444	0.2357
BARYC.	0.3189	0.3178	0.5390	0.5454
MLP	0.2648	0.2822	0.5091	0.5163
STRUCT	0.3513	0.3751	0.5061	0.5171
STRUCT*	0.3502	0.3708	**0.5521**	**0.5615**
STRUCT**	0.3372	0.347	0.5444	0.5516
SEM	0.3416	0.3676	0.5260	0.5309
SEM*	0.3556	0.3546	0.5379	0.5362
SEM**	**0.3610**	**0.3838**	0.5103	0.5274

Integration of lexical network knowledge systematically tends to **improve** the WSD performances

Better performance of SEM for verbs can be attributed to the #LF-Arcs – #Lemma ratio which is more for verbs (3.85) than nouns (2.49)

WSD on our dataset for French verbs is **harder** than for nouns.

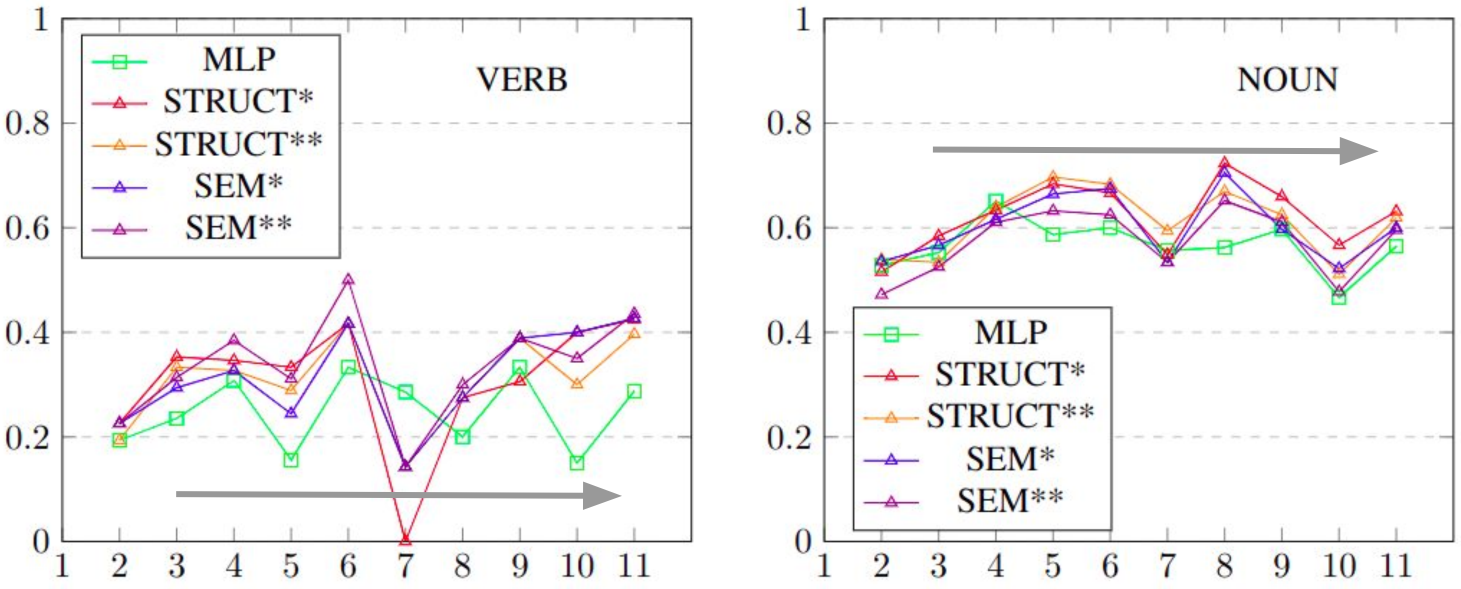


Figure 1: Polysemic performance analysis on dev set; x-axis: `sense-count` and y-axis : `accuracy`

Our proposed models tend to **more effectively** disambiguate polysemic lemmas with more than 3-4 senses than the MLP baseline

- A preliminary study of various word sense disambiguation systems on the French dataset, DBLE-LN-fr.
- Proposed a weighted training model in order to incorporate the richness of lexical and semantic information from the fr-LN network

In future work,

- The scarcity of A matrix: e.g. adding neighbors of various POS, or including transitive closures of relation
- Incorporation of definition embeddings
- Expansion on unknown senses

References

[1] Navigli, R. (2009). Word sense disambiguation: A survey. *ACM computing surveys (CSUR)*, *41*(2), 1-69.

[2] ATILF and CLLE. 2020. Corpus journalistique issu de l’est républicain. ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr.

[3] Bevilacqua, M., & Navigli, R. (2020, July). 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* (pp. 2854-2864).

[4] Vial, L., Lecouteux, B., & Schwab, D. (2019). Sense vocabulary compression through the semantic knowledge of wordnet for neural word sense disambiguation. *arXiv preprint arXiv:1905.05677*.

[5] Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation*, *43*(3), 209-226.

[6] Polguère, A. (2014). From writing dictionaries to weaving lexical networks. *International Journal of Lexicography*, *27*(4), 396-418.

[7] Huang, L., Sun, C., Qiu, X., & Huang, X. (2019). GlossBERT: BERT for word sense disambiguation with gloss knowledge. *arXiv preprint arXiv:1908.07245*.

[8] Wilks, Y., & Stevenson, M. (1998). Word sense disambiguation using optimised combinations of knowledge sources. *arXiv preprint cmp-lg/9806014*.

[9] Kumar, S., Jat, S., Saxena, K., & Talukdar, P. (2019, July). Zero-shot word sense disambiguation using sense definition embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 5670-5681).

[10] Agirre, E., Martínez, D., De Lacalle, O. L., & Soroa, A. (2006, July). Two graph-based algorithms for state-of-the-art WSD. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (pp. 585-593).

[11] Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., ... & Schwab, D. (2019). Flaubert: Unsupervised language model pre-training for french. *arXiv preprint arXiv:1912.05372*.

[12] Martin, L., Muller, B., Suárez, P. J. O., Dupont, Y., Romary, L., de La Clergerie, É. V., ... & Sagot, B. (2019). CamemBERT: a tasty French language model. *arXiv preprint arXiv:1911.03894*.

Thank you for your attention

GitHub: https://github.com/ATILF-UMR7118/GraphWSD
Contact : asinha@atilf.fr

Analysis : Polysemy in Fr-LN dataset