



**HAL**  
open science

# Compression de vocabulaire de sens grâce aux relations sémantiques pour la désambiguïsation lexicale - Version courte

Loïc Vial, Benjamin Lecouteux, Didier Schwab

## ► To cite this version:

Loïc Vial, Benjamin Lecouteux, Didier Schwab. Compression de vocabulaire de sens grâce aux relations sémantiques pour la désambiguïsation lexicale - Version courte. 2019. hal-02127237v1

**HAL Id: hal-02127237**

**<https://hal.science/hal-02127237v1>**

Preprint submitted on 13 May 2019 (v1), last revised 7 Jul 2019 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Compression de vocabulaire de sens grâce aux relations sémantiques pour la désambiguïsation lexicale

Loïc Vial, Benjamin Lecouteux, Didier Schwab

## RÉSUMÉ

---

En Désambiguïsation Lexicale (DL), les systèmes supervisés dominent largement les campagnes d'évaluation. La performance et la couverture de ces systèmes sont cependant rapidement limités par la faible quantité de corpus annotés en sens disponibles. Dans cet article, nous présentons une nouvelle méthode qui vise à résoudre ce problème en exploitant les relations sémantiques entre les sens tels que la synonymie, l'hyponymie et l'hyperonymie, afin de compresser le vocabulaire de sens de WordNet, et ainsi réduire le nombre d'étiquettes différentes nécessaires pour pouvoir désambiguïser tous les mots de la base lexicale. Notre méthode réduit considérablement le temps d'apprentissage et la taille des modèles de DL neuronaux, d'un facteur de 1,6 à 2,2 selon le corpus d'apprentissage, avec l'avantage d'améliorer leur couverture sans données supplémentaires. Avec notre méthode combinée aux récents travaux sur les représentations vectorielles de mots contextualisées, nous obtenons des résultats qui surpassent largement l'état de l'art sur toutes les tâches d'évaluation de la DL.

## ABSTRACT

---

### **Sense Vocabulary Compression through Semantic Knowledge for Word Sense Disambiguation**

In Word Sense Disambiguation (WSD), the predominant approach generally involves a supervised system trained on sense annotated corpora. The limited quantity of such corpora however restricts the coverage and the performance of these systems. In this article, we present a new method that tackles this problem by exploiting the semantic relationships between senses such as synonymy, hypernymy and hyponymy, in order to compress the sense vocabulary of WordNet, and thus reduce the number of different sense tags that must be observed to disambiguate all words of the lexical database. Our method greatly reduces the training time and the size of neural WSD models, by a factor of 1.6 to 2.2 depending on the training corpus, with the benefit of improving their coverage without additional training data. Our method combined with the recent advances in contextualized word embeddings leads to results that significantly outperform the state of the art on all WSD evaluation tasks.

---

**MOTS -CLÉS** : désambiguïsation lexicale, compression de vocabulaire, relations sémantiques.

**KEYWORDS** : Word Sense Disambiguation, Vocabulary Compression, Semantic Relationships.

---

## 1 Introduction

La Désambiguïsation Lexicale (DL) est une tâche qui vise à clarifier un texte en assignant à chacun de ses mots l'étiquette de sens la plus appropriée depuis un inventaire de sens prédéfini.

Il existe diverses approches pour la DL, telles que les approches à base de connaissances, qui s'appuient sur des dictionnaires, des bases de données lexicales ou des graphes de connaissances couplés à des algorithmes tels que les mesures de similarité lexicale (Lesk, 1986) ou des mesures basées sur les graphes (Moro et al., 2014), ou les méthodes supervisées, qui exploitent des corpus annotés en sens comme données d'apprentissage pour entraîner un classifieur multi-classe tel qu'un SVM (Chan et al., 2007; Zhong and Ng, 2010), ou plus récemment un réseau neuronal (Kågebäck and Salomonsson, 2016). Les méthodes supervisées sont de loin les plus représentées car elles offrent

généralement les meilleurs résultats dans les campagnes d'évaluation (par exemple (Navigli et al., 2007)). Les classifieurs état de l'art combinaient jusqu'à récemment des caractéristiques précises telles que les parties du discours et les lemmes des mots voisins, (Zhong and Ng, 2010), mais ils sont maintenant remplacés par des réseaux de neurones récurrents qui apprennent leur propre représentation des mots (Raganato et al., 2017; Le et al., 2018; Vial et al., 2018a).

Cependant, une des limitations majeures des systèmes supervisés est la quantité limitée de corpus manuellement annotés en sens. En effet, le SemCor (Miller et al., 1993), qui est le plus grand corpus manuellement annoté en sens disponible, contient 33 760 labels de sens différents, ce qui correspond à seulement environ 16% de l'inventaire de sens de WordNet<sup>1</sup>, la base de données lexicale de référence largement utilisée en DL. De nombreux travaux tentent de résoudre ce problème via la création de nouveaux corpus annotés en sens, générés soit automatiquement (Pasini and Navigli, 2017), semi-automatiquement (Taghipour and Ng, 2015), ou bien par *crowdsourcing* (Yuan et al., 2016), mais dans nos travaux, nous cherchons à résoudre ce problème en tirant parti des relations sémantiques présentes entre les sens de WordNet, en particulier les relations d'hyperonymie et d'hyponymie. Notre méthode est basée sur les observations suivantes :

1. Un sens, son hyperonyme et ses hyponymes partagent tous une même idée ou concept, à des niveaux d'abstraction différents.
2. En général, un mot peut être désambiguïsé en utilisant seulement les hyperonymes de ses sens.
3. Par conséquent, nous n'avons pas besoin de connaître tous les sens de WordNet pour désambiguïser tous les mots de WordNet.

Par exemple, considérons le mot « souris » et deux de ses sens : la souris *d'ordinateur* et la souris *l'animal*. Les notions plus générales comme « être vivant » (hyperonyme de souris/animal) et « appareil électronique » (hyperonyme de souris/ordinateur), permettent déjà de distinguer les deux sens, et toutes les notions plus spécialisées telles que « rongeur » ou « mammifère » sont, elles, superflues. En regroupant ces étiquettes de sens ensemble, on peut bénéficier de tous les autres exemples mentionnant un appareil électronique ou un être vivant dans un corpus d'entraînement, même si le mot « souris » n'est pas mentionné spécifiquement, pour désambiguïser le mot « souris ».

**Contributions :** Dans cet article, nous émettons l'hypothèse que seul un sous-ensemble des sens de WordNet peut être considéré pour pouvoir désambiguïser tous les mots de la base lexicale. Par conséquent, nous proposons une méthode pour construire ce sous-ensemble que nous appelons compression de vocabulaire de sens. En utilisant cette technique, nous sommes en mesure d'améliorer considérablement la couverture et la capacité de généralisation des systèmes de DL supervisés, en éliminant quasiment le besoin d'une stratégie de repli habituellement employée pour les mots jamais observés pendant l'entraînement. Nous présentons des résultats qui surpassent l'état de l'art de façon significative sur toutes les tâches d'évaluation de la DL, et nous fournissons à la communauté notre outil ainsi que nos meilleurs modèles pré-entraînés, sur un dépôt GitHub dédié<sup>2</sup>.

## 2 Compression de vocabulaire de sens

Les systèmes supervisés neuronaux état de l'art tels que Yuan et al. (2016); Raganato et al. (2017); Le et al. (2018); Vial et al. (2018a); Luo et al. (2018a,b) sont tous confrontés aux mêmes limitations :

1. La quantité de données annotés manuellement en sens étant très limitée, il se peut qu'un mot cible ne soit jamais observé pendant l'entraînement. Dans ce cas, le système ne peut pas être en mesure de l'annoter, et une stratégie de repli est généralement effectuée (par exemple utiliser le premier

---

1. <https://wordnet.princeton.edu/documentation/wnstats7wn>

2. Les liens seront disponibles dans la version non-anonymisée de l'article.

sens du mot dans WordNet).

2. Pour la même raison, un mot peut être observé, mais pas tous ses sens. Dans ce cas, le système va être capable d'annoter ce mot, mais si le sens attendu n'a jamais été observé, le résultat sera faux, quelle que soit l'architecture sous-jacente du système supervisé.
3. L'empreinte mémoire des modèles neuronaux ainsi que leur temps d'entraînement et d'exécution augmente avec la quantité de données d'apprentissage et le nombre d'étiquettes de sens différentes prises en compte, jusqu'à 206 941 si l'on considère toutes les étiquettes de sens de WordNet.

Afin de résoudre ces problèmes, nous proposons une nouvelle méthode permettant de regrouper ensemble des étiquettes de sens qui se réfèrent à un même concept. En conséquence, le vocabulaire de sens diminue, le système est capable de mieux généraliser, et sa couverture augmente.

Dans la base de données lexicale WordNet, les sens sont organisés en ensembles de synonymes appelés *synsets*. Un *synset* est concrètement un groupe d'un ou plusieurs sens qui ont la même définition et donc la même signification. Par exemple, les premiers sens des mots « *eye* », « *optic* » et « *oculus* » appartiennent tous au même *synset* dont la définition est « l'organe de la vue ».

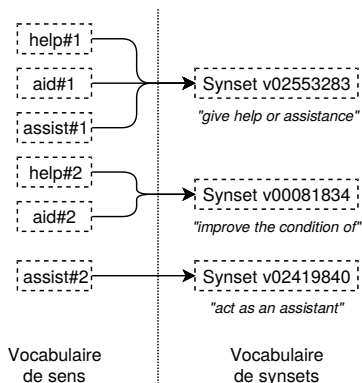


FIGURE 1 – Conversion des étiquettes de sens vers des étiquettes de *synsets*, appliqué aux deux premiers sens des mots « *help* », « *aid* » et « *assist* ». Le nombre total d'étiquettes de sens différentes passe ainsi de six à trois.

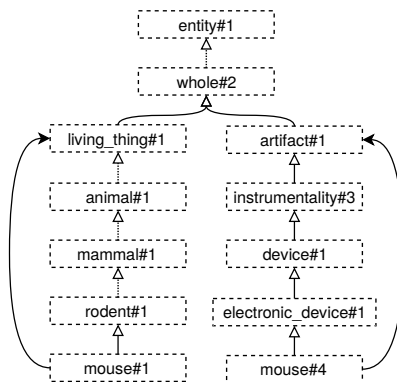


FIGURE 2 – Compression de vocabulaire utilisant la hiérarchie d'hyponymie, appliquée au premier et quatrième sens du mot « *mouse* ». Les lignes en pointillés indiquent que des nœuds ont été omis pour la clarté.

La conversion des étiquettes de sens (« Xème sens du mot N ») aux étiquettes de *synsets* (« *synset* numéro Y »), illustré dans la figure 1, est ainsi une façon de compresser le vocabulaire qui est déjà appliquée dans plusieurs travaux (Yuan et al., 2016; Vial et al., 2018a; Le et al., 2018) sans être toujours explicitement précisée. Cette méthode contribue pourtant clairement à améliorer la couverture des systèmes supervisés. En effet, si le verbe « *aid* » annoté avec son premier sens est observé dans les données d'apprentissage, le contexte autour du mot cible peut être aussi utile pour annoter ultérieurement les verbes « *assist* » ou « *help* » avec la même étiquette de *synset*.

La nouvelle méthode que nous proposons repose sur ce même principe de regroupement de sens, mais va plus loin, en exploitant les relations d'hyponymie et d'hyperonymie entre les sens.

Selon Polguère (2003), l'hyponymie et l'hyperonymie sont deux relations sémantiques qui correspondent à un cas particulier d'inclusion de sens : l'hyponyme d'un terme est une spécialisation de ce terme, alors que son hyperonyme est une généralisation. Par exemple, une « souris » est un type de « rongeur » qui est à son tour un type de « animal ». Dans WordNet, ces relations lient presque tous les

noms ensemble<sup>3</sup> allant de la racine générique, le nœud « entité » aux feuilles les plus spécifiques, par exemple le nœud « souris à pattes blanches ». Ces relations sont également présentes sur plusieurs verbes : ainsi, par exemple, « additionner » est une manière de « calculer » qui est à son tour une manière de « raisonner ».

Pour la DL, tout comme le regroupement des synonymes en *synsets* aide à mieux généraliser, nous faisons l’hypothèse que le regroupement des sens faisant partie d’une même hiérarchie d’hyperonymie va aussi aider à mieux généraliser, et que les concepts les plus spécialisés de WordNet sont souvent superflus. En effet, si l’on considère un sous-ensemble de WordNet qui ne comprend que le mot « souris », avec son premier sens (le petit rongeur), son quatrième sens (le dispositif électronique), et tous leurs hyperonymes, tel qu’illustré dans la figure 2, on voit que les concepts « artefact » et « être vivant » suffisent à différencier les deux sens, et toutes les étiquettes plus spécialisées pourrait être ramenés à ces deux concepts. Ainsi, non seulement le vocabulaire de sens est réduit, mais en plus tous les autres « êtres vivants » donneront des exemples qui pourront ensuite permettre de différencier les deux sens de souris.

En considérant tout le vocabulaire de WordNet maintenant, l’objectif de notre méthode est ainsi de faire correspondre chaque sens de WordNet à son ancêtre le plus haut dans sa hiérarchie d’hyperonymie, avec les contraintes suivantes : Premièrement, cet ancêtre doit permettre de discriminer tous les différents sens du mot cible. Deuxièmement, nous devons conserver les hyperonymes qui sont indispensables pour discriminer les sens des autres mots du dictionnaire. Par exemple, en prenant tout WordNet en considération, nous ne pouvons pas faire correspondre « souris#1 » à « être vivant#1 », parce qu’une étiquette plus spécifique, « animal#1 » est nécessaire pour distinguer les deux sens du mot « proie » (un sens décrit une personne et l’autre un animal).

Notre méthode fonctionne donc en deux étapes :

1. Nous marquons comme « nécessaires » les enfants du premier ancêtre commun de chaque paire de sens de chaque mot de WordNet.
2. Nous faisons correspondre chaque sens à son ancêtre le plus bas dans sa hiérarchie d’hyperonymie ayant été précédemment marqué comme « nécessaire ».

En conséquence, les sens les plus spécifiques de l’arbre qui ne sont pas indispensables pour distinguer un mot de l’inventaire lexical seront automatiquement supprimés du vocabulaire. En d’autres termes, l’ensemble de sens qui reste dans le vocabulaire est le plus petit sous-ensemble de tous les *synsets* qui sont nécessaires pour distinguer chaque sens de chaque mot de WordNet.

Méthode de compression	Taille du vocabulaire	Taux de compression	Couverture du SemCor
Référence	206 941	référence	16%
Synonymes	117 659	43%	22%
Hyperonymes	39 147	81%	32%

TABLE 1 – Résultats de la compression de vocabulaire sur la taille du vocabulaire et la couverture du SemCor.

Le tableau 1 montre l’effet de la compression de vocabulaire via les synonymes (sens vers *synsets*) ainsi que de notre nouvelle méthode utilisant les hyperonymes, sur la taille du vocabulaire de sens de WordNet, et sur la couverture du SemCor. Comme nous pouvons le constater, la taille du vocabulaire des sens diminue considérablement grâce à notre méthode, et la couverture d’un même corpus est nettement améliorée.

3. Nous avons calculé que 41 607 sur les 44 449 noms polysémiques de WordNet (94%) font partie de cette hiérarchie.

### 3 Protocole expérimental

Afin d'évaluer notre méthode de compression de vocabulaire de sens, nous l'avons appliquée sur un système neuronal de DL état de l'art similaire à celui de Vial et al. (2018a). Notre réseau de neurones prend ainsi en entrée directement les mots sous formes vectorielles, à partir d'un modèle de vecteurs de mots pré-entraîné, il repose ensuite sur une couche de cellules *LSTM* bidirectionnelles (Hochreiter and Schmidhuber, 1997), puis sur une couche de sortie, qui associe à chaque mot une distribution de probabilité sur tous les sens du vocabulaire utilisé, à l'aide de la fonction softmax.

Nous avons ainsi entraînés deux systèmes :

1. un système « référence » dont le vocabulaire de sens est celui de tous les synsets vus pendant l'entraînement (utilisant ainsi la compression classique via les synonymes) ;
2. un système « compressé » entraîné dans les mêmes conditions, mais avec notre méthode de compression du vocabulaire via les hyperonymes appliquée sur le corpus d'entraînement.

En entrée de notre réseau, nous avons utilisé les vecteurs de mots contextualisés ELMo (Peters et al., 2018). Nous avons ainsi utilisé le modèle entraîné sur Wikipedia et les données monolingues de WMT 2008-2012<sup>4</sup>, dont la dimension des vecteurs est 512. Pour tous les autres paramètres du modèle, comme la taille des cellules LSTM, le nombre de phrase par mini-lot, le *dropout*, la méthode d'optimisation, etc., nous avons utilisé les mêmes paramètres que Vial et al. (2018a).

Nous avons comparé notre méthode sur deux ensembles de corpus d'entraînement : Le SemCor (Miller et al., 1993), le plus grand corpus annoté en sens utilisé pour l'apprentissage de la plupart des systèmes supervisés de DL, et la concaténation du SemCor et du WordNet Gloss Tagged<sup>5</sup>. Ce dernier est un corpus distribué dans WordNet depuis sa version 3.0, et il contient les définitions de tous les sens de WordNet, annoté manuellement ou semi-automatiquement en sens. Nous avons utilisé les versions de ces corpus fournies avec la ressource UFSAC 2.1<sup>6</sup> (Vial et al., 2018b).

Système	Temps d'entraînement		Nombre de paramètres	
	SemCor	SemCor+WNGT	SemCor	SemCor+WNGT
Référence	50mn	360mn (6h)	62.64M	149.35M
Compressé	30mn	160mn (2h40)	35.87M	68.41M

TABLE 2 – Temps d'entraînement et nombre de paramètres d'un modèle en fonction du corpus d'apprentissage et de l'utilisation de notre méthode de compression de vocabulaire.

Nous avons entraîné chaque modèle sur 20 passes de nos données d'entraînement. Nous avons ensuite évalué notre modèle à la fin de chaque passe sur un jeu de développement constitué de 4 000 phrases prises aléatoirement dans nos données d'entraînement, et nous avons au final conservé celui qui a obtenu le meilleur score F1 de DL. Les temps d'entraînement approximatifs et le nombre de paramètres des différents modèles, en fonction du corpus d'entraînement et de l'utilisation de notre méthode de compression du vocabulaire, sont affichés dans le tableau 2.

Nous avons évalué nos modèles sur tous les corpus d'évaluation de la DL de l'anglais des campagnes d'évaluation SensEval/SemEval, c'est-à-dire les corpus d'évaluation « grain fin » de SensEval 2 (Edmonds and Cotton, 2001), SensEval 3 (Snyder and Palmer, 2004), SemEval 2007 (tâche 17) (Pradhan et al., 2007), SemEval 2013 (Navigli et al., 2013) et SemEval 2015 (Moro and Navigli, 2015), ainsi que le corpus « ALL » constitué de leur concaténation. Nous avons également comparé nos résultats sur la tâche « gros grain » de SemEval 2007 (tâche 7) (Navigli et al., 2007).

Les scores obtenus par nos systèmes en comparaison avec les meilleurs systèmes de l'état de l'art

4. <https://allennlp.org/elmo>

5. <http://wordnetcode.princeton.edu/glosstag-files/glosstag.shtml>

6. <https://github.com/getalp/UFSAC>

Système	SE2	SE3	SE07 17	SE13	SE15	ALL	SE07 07
SemCor, référence	91.15	96.76	97.58	91.06	94.78	93.23	92.84
SemCor, compressé	98.03	99.19	99.78	99.15	98.39	98.75	98.85
SemCor+WNGT, référence	97.81	98.92	99.34	97.63	99.34	98.26	98.45
SemCor+WNGT, compressé	99.74	99.95	100	99.76	99.91	99.83	99.91

TABLE 3 – Couverture du corpus d’entraînement en fonction du corpus d’apprentissage et de l’utilisation de notre méthode de compression de vocabulaire.

et l’étalon du premier sens sont présents dans le tableau 4, et le tableau 3 montre la couverture de nos systèmes sur les tâches d’évaluation. Comme nous pouvons le constater, notre méthode de compression améliore grandement la couverture de nos systèmes : Sur un total de 7 253 mots à annoter pour le corpus « ALL », le système de base entraîné sur le SemCor n’est pas capable d’annoter 491 d’entre eux, alors qu’avec la compression du vocabulaire, ce nombre descend à 91. Lors de l’ajout du WordNet Gloss Tagged aux données d’entraînement, seulement 12 mots ne peuvent pas être annotés.

Système	SE2	SE3	SE07 17	SE13	SE15	ALL	SE07 07
Étalon du premier sens	65.6	66.0	54.5	63.8	67.1	65.5	78.9
HCAN (Luo et al., 2018a)	72.8	70.3	-	68.5	72.8	71.1	-
LSTMLP (T-SemCor, U-1K) (Yuan et al., 2016)	73.8	71.8	63.5	69.5	72.6	71.5	83.6
SemCor, référence	74.93	73.84	64.84	69.83	72.70	72.55	85.01
SemCor, compressé	74.67	73.46	64.40	72.45	73.78	73.09	86.1
SemCor+WNGT, référence	<b>77.13</b>	<b>75.08</b>	67.25	73.72	76.71	<b>75.16</b>	87.88
SemCor+WNGT, compressé	76.47	73.08	<b>67.91</b>	<b>74.33</b>	<b>77.01</b>	74.65	<b>88.41</b>

TABLE 4 – Scores F1 (%) sur les tâches de DL de l’anglais des campagnes d’évaluation SensEval/SemEval. La stratégie de repli est appliquée sur les mots dont aucun sens n’a été observé pendant l’entraînement. Les scores prefixés par une obélisque (†) ne sont pas fournis par les auteurs mais sont déduits des autres scores.

Concernant les résultats, nos systèmes qui utilisent la compression de vocabulaire obtiennent des scores qui sont globalement équivalents aux systèmes qui ne l’utilisent pas, et par rapport aux autres travaux, nous obtenons des résultats surpassant significativement l’état de l’art dans toutes les tâches, notamment grâce à l’ajout du WordNet Gloss Tagged aux données d’entraînement, et des vecteurs ELMo en entrée de notre système.

## 4 Conclusion

Dans cet article, nous avons présenté une nouvelle méthode qui améliore la couverture et la capacité de généralisation des systèmes de DL supervisés, en réduisant le nombre d’étiquettes de sens différentes dans WordNet afin de ne conserver que celles qui sont essentielles pour différencier les sens de tous les mots présents dans la base lexicale.

Nous avons entraîné un système de DL neuronal état de l’art afin de montrer que cette méthode permettait de réduire la taille et le temps d’entraînement des modèles par un facteur de 1,6 à 2,2 sans dégrader les performances. Notre méthode combinée avec les récentes avancées en terme de vecteurs de mots pré-entraînés permet à notre système de surpasser nettement l’état de l’art dans toutes les tâches d’évaluation de la DL de l’anglais, avec une bien meilleure couverture.

En perspective, nous envisageons d’étendre notre méthode de compression de vocabulaire en groupant ensemble les sens reliés par d’autres relations sémantiques telles que l’antonymie (opposé de) ou la méronymie (partie de). Nous pensons que cette approche offre une alternative solide à la création automatique ou semi-automatique de corpus annotés en sens, car elle élimine presque totalement la nécessité d’une stratégie de repli dans les systèmes de DL supervisés, sans impacter la précision.

# Références

- Yee Seng Chan, Hwee Tou Ng, and Zhi Zhong. 2007. Nus-pt: Exploiting parallel texts for word sense disambiguation in the english all-words tasks. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, SemEval '07, pages 253–256, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Philip Edmonds and Scott Cotton. 2001. Senseval-2: Overview. In *The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems*, SENSEVAL '01, pages 1–5, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8) :1735–1780.
- Mikael Kågebäck and Hans Salomonsson. 2016. Word sense disambiguation using a bidirectional lstm. In *5th Workshop on Cognitive Aspects of the Lexicon (CogALex)*. Association for Computational Linguistics.
- Minh Le, Marten Postma, Jacopo Urbani, and Piek Vossen. 2018. A deep dive into word sense disambiguation with lstm. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 354–365. Association for Computational Linguistics.
- Michael Lesk. 1986. Automatic sense disambiguation using mrd : how to tell a pine cone from an ice cream cone. In *Proceedings of SIGDOC '86*, pages 24–26, New York, NY, USA. ACM.
- Fuli Luo, Tianyu Liu, Zexue He, Qiaolin Xia, Zhifang Sui, and Baobao Chang. 2018a. 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. Association for Computational Linguistics.
- Fuli Luo, Tianyu Liu, Qiaolin Xia, Baobao Chang, and Zhifang Sui. 2018b. Incorporating glosses into neural word sense disambiguation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1 : Long Papers)*, pages 2473–2482. Association for Computational Linguistics.
- George A. Miller, Claudia Leacock, Randee Teng, and Ross T. Bunker. 1993. A semantic concordance. In *Proceedings of the workshop on Human Language Technology*, HLT '93, pages 303–308, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Andrea Moro and Roberto Navigli. 2015. 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. Association for Computational Linguistics.
- Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity linking meets word sense disambiguation : a unified approach. *TACL*, 2 :231–244.
- Roberto Navigli, David Jurgens, and Daniele Vannella. 2013. 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.
- Roberto Navigli, Kenneth C. Litkowski, and Orin Hargraves. 2007. Semeval-2007 task 07 : Coarse-grained english all-words task. In *SemEval-2007*, pages 30–35, Prague, Czech Republic.
- Tommaso Pasini and Roberto Navigli. 2017. 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. Association for Computational Linguistics.



- Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proc. of NAACL*.
- Alain Polguère. 2003. *Lexicologie et sémantique lexicale*. Les Presses de l'Université de Montréal.
- Sameer S. Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. 2007. Semeval-2007 task 17: English lexical sample, srl and all words. In *Proceedings of the 4th International Workshop on Semantic Evaluations, SemEval '07*, pages 87–92, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Alessandro Raganato, Claudio Delli Bovi, and Roberto Navigli. 2017. Neural sequence learning models for word sense disambiguation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1167–1178. Association for Computational Linguistics.
- Benjamin Snyder and Martha Palmer. 2004. 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*.
- Kaveh Taghipour and Hwee Tou Ng. 2015. One Million Sense-Tagged Instances for Word Sense Disambiguation and Induction. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 338–344, Beijing, China. Association for Computational Linguistics.
- Loïc Vial, Benjamin Lecouteux, and Didier Schwab. 2018a. Approche supervisée à base de cellules LSTM bidirectionnelles pour la désambiguïsation lexicale. In *25e conférence sur le Traitement Automatique des Langues Naturelles*, Rennes, France.
- Loïc Vial, Benjamin Lecouteux, and Didier Schwab. 2018b. UFSAC: Unification of Sense Annotated Corpora and Tools. In *Language Resources and Evaluation Conference (LREC)*, Miyazaki, Japan.
- Dayu Yuan, Julian Richardson, Ryan Doherty, Colin Evans, and Eric Altendorf. 2016. Semi-supervised word sense disambiguation with neural models. In *COLING 2016*.
- Zhi Zhong and Hwee Tou Ng. 2010. It makes sense: A wide-coverage word sense disambiguation system for free text. In *Proceedings of the ACL 2010 System Demonstrations, ACLDemos '10*, pages 78–83, Stroudsburg, PA, USA. Association for Computational Linguistics.