Monolingual Word Sense Alignment as a Classification Problem

Sina Ahmadi, John P. McCrae
Insight Centre for Data Analytics
Data Science Institute, National University of Ireland Galway
{sina.ahmadi,john.mccrae}@insight-centre.org

Abstract

Words are defined based on their meanings in various ways in different resources. Aligning word senses across monolingual lexicographic resources increases domain coverage and enables integration and incorporation of data. In this paper, we explore the application of classification methods using manually-extracted features along with representation learning techniques in the task of word sense alignment and semantic relationship detection. We demonstrate that the performance of classification methods dramatically varies based on the type of semantic relationships due to the nature of the task but outperforms the previous experiments.

1 Introduction

Dictionaries are valuable resources which document the life of words in a language from various points of view. Creating and maintaining such resources for a constantly changing phenomenon like human language requires much time and effort. With the expansion of collaboratively-curated resources such as Wiktionary, processing lexicographical resources automatically and efficiently is of high importance recently in computational lexicography, computational linguistics and natural language processing (NLP).

Senses, or definitions, are important components of dictionaries where dictionary entries, i.e. lemmata, are described in plain language. Therefore, unlike other properties such as references, comparisons (*cf.*), synonyms and antonyms, senses are unique in the sense that they are more descriptive but also highly contextualized. Moreover, unlike lemmata which remain identical through resources in the same language, except in spelling variations, senses can undergo tremendous changes based on the choice of the editor, lexicographer and publication period, to mention but a few factors. Therefore, the task of word sense alignment (WSA) will facilitate the integration of various resources and the creation of interlinked language resources.

Considering the literature, various components of the WSA task has been matters of research previously. However, a few of previous papers address WSA as a specific task on its own. In this paper, our focus is on providing explainable observations for the task of WSA using manually-extracted features and analyze the performance of traditional machine learning algorithms for word sense alignment as a classification problem. Despite the increasing popularity of deep learning methods in providing state-of-the-art results in various NLP fields, we believe that evaluating the performance of feature-engineered approaches is an initial and essential step to reflect the difficulties of the task and also, the expectations from the future approaches.

2 Related Work

The alignment of lexical resources has been previously of interest both to create resources and propose alignment approaches. In this section, we only focus on WSA techniques in the related literature.

Graph-based approaches have been widely used for the WSA task. Matuschek and Gurevych (2013) propose a graph-based approach, called Dijkstra-WSA, for aligning lexical-semantic resources, namely wordnet, OmegaWiki, Wiktionary and Wikipedia. In this approach, senses are represented as the nodes of a graph where the edges represent the semantic relation between them. Assuming that monosemous lemmata have a more specific meaning and therefore less ambiguous to match, a semantic relation is created among the senses of such lemmata when they appear in a sense of a polysemous lemma. Using

Dijkstra's shortest path algorithm along with semantic similarity scores and without requiring any external data or corpora, a set of possible sense matches are retrieved. In the same vein, Ahmadi et al. (Ahmadi et al., 2019) model the alignment task as a bipartite-graph where an optimal alignment solution is selected among the combination of possible sense matches in two resources. Although this algorithm performs competitively with the Dijkstra-WSA technique on the same datasets, no viable solution is provided regarding the tuning of the matching algorithm. Similarly, other authors (Nancy and Véronis, 1990; Pantel and Pennacchiotti, 2008; Meyer and Gurevych, 2010; Pilehvar and Navigli, 2014) focus on linking senses without considering semantic relationships.

Beyond aligning lexical resources, there has been much effort in inducing semantic relationships, particularly within more generic fields such as taxonomy extraction (Bordea et al., 2015), hypernym discovery (Camacho-Collados et al., 2018) and semantic textual similarity (Agirre et al., 2016). Although in these tasks the focus is on the relationship within words, there are a few works exploring how to induce semantic relationships between definitions. Heidenreich and Williams (2019) introduce an algorithm using a directed acyclic graph to construct a wordnet based on the Wiktionary data and enriched with synonym and antonym relationships. Using the semantic relationship annotations provided in Wiktionary, the method induces a semantic hierarchy by identifying a subset within each sense that can relate two lemmas together. In addition to graph-based methods, there are various other closely-related fields, such as word sense disambiguation (Maru et al., 2019) and sense embeddings (Iacobacci et al., 2015), which can potentially contribute to the task of WSA. However, we could not find any previous work exploring those approaches.

One major limitation regarding previous work is with respect to the nature of the data used for the WSA task. Expert-made resources, such as the Oxford English Dictionary, require much effort to create and therefore, are not as widely available as collaboratively-curated ones like Wiktionary[1] due to copyright restrictions. On the other hand, the latter resources lack domain coverage and descriptive senses. To address this, Ahmadi et al. (2020) present a set of 17 datasets containing monolingual dictionaries in 15 languages, annotated by language experts with five semantic relationships according to the simple knowledge organization system reference (SKOS) (Miles and Bechhofer, 2009), namely, broader, narrower, related, exact and none. Our objective within this project is to explore the alignment of these open-source datasets using classification methods.

3 Problem Definition

Ignoring the differences in dictionary structures and formats such as XML, LMF (Francopoulo et al., 2006) and Ontolex-Lemon (McCrae et al., 2017), there are different lexicographic and logical ways for describing senses in a dictionary (Solomonick, 1996). As an example, Table 3 provides the senses available for ENTIRE (adjective) in various lexical resources where the predominant sense of "whole" or "complete" is provided in all resources. However, all resources do not equally cover specific domains such as botany and mathematics. Therefore, there are differences in the number of provided senses, e.g. one sense is provided in MACMILLAN while the Oxford Dictionary provides five.

We define our task of WSA and semantic induction as the detection of the semantic relationship between a pair of senses in two monolingual resources, as follows:

$$rel = sem(p, s_i, s_j) \quad (1)$$

where p is the part-of-speech of the lemma, s_i and s_j are senses belonging to the same lexemes in two monolingual resources and rel is a semantic relation, namely exact, broader, narrower, related and none. Our goal is to predict a semantic relation, i.e. rel given a pair of senses. Therefore, we define three classification problems based on the relation:

- **Binary classification** which predicts if two senses can possibly be aligned together. Otherwise, none is selected as the target class.
- **SKOS classification** which predicts a label among `exact`, `broader`, `narrower` and `related` semantic relationships.
- **SKOS+none classification** which predicts a label given all data instances. This is similar to the previous classifier, with `none` as a target class.

[1] www.wiktionary.org

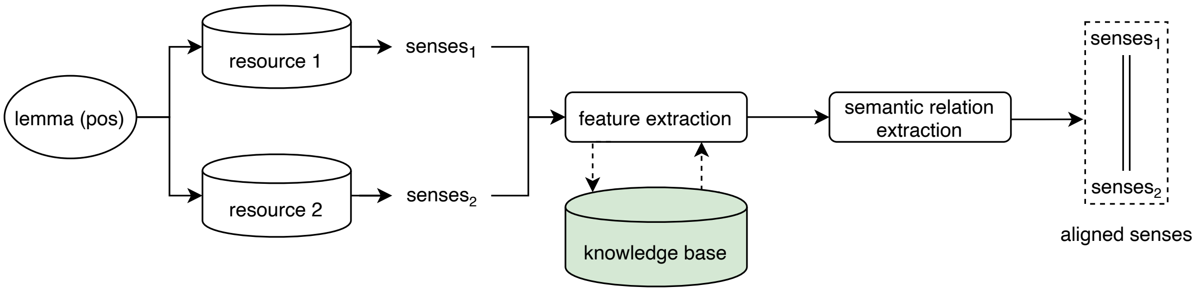

Figure 1: Our approach where features are extracted from word senses and external semantic resources

4 Approach

Assuming that the textual representation of senses as in definitions can be useful to align them, we define a few features which use the lengths of senses along with their textual and semantic similarities. In addition, we incorporate word-level semantic relationships to determine the type of relation that two senses may possibly have. To this end, we use CONCEPTNET (Speer et al., 2016), an openly-available and multilingual semantic network with relational knowledge from various other resources, such as Wiktionary and WordNet (Miller, 1995). A similar approach has been previously proposed for aligning bilingual with monolingual dictionaries (Saurí et al., 2019).

4.1 Feature Extraction

In this step, we extract sense instances from the MWSA datasets (Ahmadi et al., 2020), as $t = (p, s_i, s_j, r_{i_j})$. This instance is interpreted as sense s_i has relation r_{i_j} with sense s_j. Therefore, the order of appearance is important to correctly determine the relationship. It should also be noted that both senses belong to the same lemma with the part-of-speech p. Table 2 provides the basic statistics of the senses and their semantic relationships in various languages. # Entries and # SKOS refer to the number of entries and senses with a relationship within SKOS. In addition, the senses within the two resources which belong to the same lemma but are not annotated with a SKOS relationship, are included with a `none` relationship.

Given the class imbalance where senses with a `none` relationship are more frequent than the others, we carry out a data augmentation technique based on the symmetric property of the semantic relationships. By changing the order of the senses, also known as relation direction, in each data instance, a new instance can be created by semantically reversing the relationship. In other words, for each $t = (p, s_i, s_j, r_{ij})$ there is a $t' = (p, s_j, s_i, r'_{ij})$ where r'_{ji} is the inverse of r_{ij}. Thus, `exact` and `related` as symmetric properties remain the same, however, the asymmetric property of the `broader` and `narrower` relationships yields `narrower` and `broader`, respectively.

Once the senses extracted, we create data instances using the features in Table 1. Features 2 and 3 concern the length of senses and how they are different. Intuitively speaking, this regards the wordings used to describe two concepts and their semantic relationship. In features 4 to 11, we calculate this with and without function words, words with little lexical meaning. One additional step is to query CONCEPTNET to retrieve semantic relations between the content words in each sense pair. For instance, the two words "gelded" and "castrated" which appear in two different senses are synonyms and therefore, the whole senses can be possibly synonyms. In order to measure the reliability of the relationships, we sum up the weights, also known as *assertions*, of each relationship according to CONCEPTNET. Finally, features 12 and 13 provide the semantic similarity of each sense pair using word embeddings. For this purpose, we used GloVe (Pennington et al., 2014) and fastText[2]. The data instances are all standardized by scaling each feature to the range of [0-1].

4.2 Feature Learning

Restricted Boltzmann machine (RBM) is a generative model representing a probability distribution given a set of observations (Fischer and Igel, 2012). An RBM is composed of two layers, a visible one where the data instances according to the manually-created features are provided, and a latent one where a distribution is created by

[2]`https://fasttext.cc/`

#	feature	definition	possible values
1	POS_tag	part of speech of the headword	a one-hot vector of {N, V, ADJ, ADV, OTHER}
2	s_len_no_func_1/2	number of space-separated tokens in s_1 and s_2	$\mathbb{N}$
3	s_len_1/2	number of space-separated tokens in s_1 and s_2 without function words	$\mathbb{N}$
4	hypernymy	hypernymy score between tokens	sum of weights in CONCEPTNET
5	hyponymy	hyponymy score between tokens	sum of weights in CONCEPTNET
6	relatedness	relatedness score between tokens	sum of weights in CONCEPTNET
7	synonymy	synonymy score between tokens	sum of weights in CONCEPTNET
8	antonymy	antonymy score between tokens	sum of weights in CONCEPTNET
9	meronymy	meronymy score between tokens	sum of weights in CONCEPTNET
10	similarity	similarity score between tokens	sum of weights in CONCEPTNET
11	sem_sim	semantic similarity score between senses using word embeddings	averaging word vectors and cosine similarity [0-1]
12	sem_sim_no_func	semantic similarity score between senses without function words	averaging word vectors and cosine similarity excluding function words [0-1]
13	sem_bin_rel	target class	1 for alignable, otherwise 0
14	sem_rel_with_none	target class	{exact, narrower, broader, related, none}
15	sem_rel	target class	{exact, narrower, broader, related}

Table 1: Manually extracted features for semantic classification of sense relationships

the model by retrieving dependencies within variables. In other words, the relation of the features in how the target classes are predicted is learned in the training phase. We follow the description of (Hinton, 2012) in implementing and using an RBM for learning further features from our data instances. Regarding the classification problem, instead of training our models using the data instances described in the previous section, we train the models using the latent features of an RBM model. These new features have binary values and can be configured and tuned depending on the performance of the models.

4.3 Classification Method

For this supervised classification problem, we use support vector machines (SVMs) using various hyper-parameters, as implemented in Scikit[3] (Pedregosa et al., 2011). After a preprocessing step, where the datasets are shuffled, normalized and scaled, we split them into train, test and validation sets with 80%, 10% and 10% proportions, respectively.

[3] https://scikit-learn.org

5 Experiments

Table 2 presents the best performance of the models trained for each language. In addition to an SVM, we also evaluated the usage of an RBM to learn features and classify them similarly using an SVM. Our baseline is based on the evaluation of Kernerman et al (2020) on the same datasets. The baseline provides accuracy for classifying sense pairs with a semantic relationship or none, i.e. SKOS+none, and precision, recall and F_1-measure for predicting whether two senses should be matched, i.e. binary classification. In the same vein, our evaluation is carried out using accuracy, precision, recall and as defined in (Powers, 2011), but for all classification setups.

Despite the high accuracy of the baseline systems for most languages, they do not perform equally efficiently for all languages in terms of precision and recall. Although our classifiers outperform the baselines for all the relation prediction tasks and perform competitively when trained for the binary classification and also given all data instances, there is a significant low performance when it comes to the classification of SKOS relationships. This can be explained by the lower number of instances available for these relations.

Language	# Entries	# SKOS	# SKOS+none	# All	Metric	Baseline	Binary	All	SKOS	RBM-Binary	RBM-all	RBM-SKOS
Basque	256	813	3661	4382	Accuracy	78.90	78.79	58.47	49.77	70.37	54.17	28.85
					Precision	21.10	71.40	59.21	43.65	62.14	59.08	20.73
					Recall	5.00	72.78	58.45	46.01	74.93	52.55	50.87
					F-measure	8.10	**72.08**	**58.83**	**44.80**	67.94	55.62	29.46
Bulgarian	1000	1976	3708	5656	Accuracy	72.80	70.60	65.91	34.05	73.51	63.38	36.47
					Precision	25.00	68.75	64.79	31.75	77.46	34.46	36.85
					Recall	1.10	69.32	65.44	31.83	72.91	49.87	24.86
					F-measure	2.00	69.03	**65.11**	**31.79**	**75.11**	40.76	29.69
Danish	587	1644	16520	18164	Accuracy	81.70	66.47	34.82	27.87	73.85	50.08	29.67
					Precision	3.00	74.54	23.70	36.49	60.59	60.96	30.47
					Recall	2.30	75.51	62.90	22.87	55.66	66.92	73.04
					F-measure	4.30	**75.02**	34.43	28.12	**58.02**	63.80	**43.00**
Dutch	161	622	20144	20766	Accuracy	93.60	82.55	59.99	24.75	83.90	51.47	36.34
					Precision	0.00	86.97	78.59	31.38	59.78	77.82	30.66
					Recall	0.00	88.24	79.22	33.10	67.33	39.65	66.03
					F-measure	0.00	**87.60**	**78.90**	32.22	63.33	52.54	41.88
English	684	1682	9269	10951	Accuracy	75.20	89.00	81.00	49.00	80.16	65.03	48.57
					Precision	0.00	82.35	73.03	39.31	64.36	63.67	55.53
					Recall	0.00	82.87	76.41	46.63	82.13	79.35	34.51
					F-measure	0.00	**82.61**	**74.68**	**42.66**	72.17	70.65	42.57
Estonian	684	1142	2316	3426	Accuracy	48.20	78.98	58.92	46.11	75.96	62.75	47.82
					Precision	54.50	76.06	68.83	40.81	63.53	60.67	36.63
					Recall	9.30	20.76	57.82	44.02	28.18	49.35	22.44
					F-measure	15.90	32.62	**62.85**	**42.35**	**39.05**	54.43	27.83
German	537	1211	4975	6185	Accuracy	77.77	73.14	61.99	49.58	77.97	43.23	44.21
					Precision	0.00	77.72	64.74	41.89	80.44	66.34	40.99
					Recall	0.00	54.41	59.95	43.73	22.88	27.92	48.99
					F-measure	0.00	**64.01**	**62.25**	42.79	35.63	39.30	**44.63**
Hungarian	143	949	15774	16716	Accuracy	94.00	79.65	58.40	22.95	81.46	36.27	15.20
					Precision	5.30	49.96	30.14	23.41	68.50	59.80	26.58
					Recall	1.20	54.47	37.95	68.08	56.72	73.85	29.23
					F-measure	2.00	52.12	33.60	**34.85**	**62.05**	**66.09**	27.84
Irish	680	975	2816	3763	Accuracy	58.30	75.00	55.75	26.27	79.61	60.84	24.75
					Precision	68.00	84.42	46.58	31.84	79.03	42.52	30.25
					Recall	18.50	84.46	39.85	46.15	52.47	54.65	25.40
					F-measure	29.10	**84.44**	42.95	37.68	63.06	**47.83**	27.61
Italian	207	592	2173	2758	Accuracy	69.30	59.08	55.43	44.48	77.23	46.26	43.01
					Precision	0.00	52.55	42.98	28.80	75.69	46.31	40.56
					Recall	0.00	66.47	52.64	42.16	45.05	68.67	31.27
					F-measure	0.00	**58.69**	47.32	**34.22**	56.49	**55.32**	35.32
Serbian	301	736	5808	6542	Accuracy	59.90	80.05	32.53	27.55	82.35	41.43	32.96
					Precision	19.00	76.78	48.57	43.06	73.51	37.70	21.49
					Recall	46.40	65.73	69.40	27.10	77.46	48.45	55.53
					F-measure	26.90	70.83	**57.15**	**33.26**	**75.43**	42.40	30.99
Slovenian	152	244	1100	1343	Accuracy	44.20	84.29	36.13	26.13	78.93	39.57	31.63
					Precision	17.30	73.08	23.19	46.98	78.62	38.59	20.97
					Recall	58.70	83.22	45.07	28.61	41.64	28.09	33.02
					F-measure	26.80	**77.82**	30.62	**35.56**	54.45	**32.51**	25.65
Spanish	351	1071	4898	5919	Accuracy	-	73.79	54.67	30.28	80.71	54.38	58.48
					Precision	-	79.78	55.07	33.21	79.40	42.54	39.57
					Recall	-	80.37	53.15	40.04	60.18	20.68	38.59
					F-measure	-	**80.07**	**54.10**	36.31	68.47	27.83	**39.07**
Portuguese	147	275	2062	2337	Accuracy	92.10	71.31	66.62	51.71	73.14	55.69	42.87
					Precision	8.30	49.29	58.23	53.52	77.72	69.41	40.45
					Recall	2.40	37.47	70.41	53.47	54.41	22.32	38.15
					F-measure	3.70	42.57	**63.74**	**53.49**	**64.01**	33.78	39.26
Russian	213	483	3376	3845	Accuracy	75.40	60.88	58.90	37.75	75.80	59.76	33.10
					Precision	43.80	72.92	63.83	27.28	73.38	73.77	32.71
					Recall	17.90	82.21	44.43	36.74	68.23	70.39	47.75
					F-measure	25.50	**77.29**	52.39	31.31	70.71	**72.04**	**38.82**

Table 2: Basic statistics of the datasets and the best classification results with and without an RBM. # refers to the number

Moreover, distinguishing certain types of relationships, such as `related` versus `exact`, is a challenging task even for an expert annotator. For instance, the relationship between two senses of ENTIRE in Table 3, "constituting the undiminished entirety" and "complete in all parts; undivided; undiminished; whole" is annotated as `narrower` and `exact` by two different annotators [4].

Regarding the performance of RBM, we do not observe a similar improvement in the results of all classifiers. The precision of the models which learn features with an RBM is higher in the majority of cases. Our optimal models where trained with 50 iterations, a learning rate within [0.05-0.2] and a hidden unit number within the range of 400 and 600.

6 Conclusion and Future Work

This paper presents a preliminary study on the task of word sense alignment using monolingual lexicographic datasets from 15 languages. The task is modeled as a classification task where data

[4]According to the datasets available at `https://github.com/elexis-eu/MWSA`

instances are extracted using various manually-defined features. The classification task aims at classifying sense matches across dictionaries and also, prediction of the semantic relationship between two given senses, namely narrower, broader, exact and related. The results indicate a better performance of the proposed approach with respect to the baselines reported previously.

One major limitation of the current approach is the usage of crafted features. We believe that as a future work further techniques can be used, particularly thanks to the current advances in word representations and neural networks. In addition, incorporating knowledge bases and external language resources such as corpora can be beneficial in improving to address sense ambiguity for polysemous entries.

Acknowledgements

The authors would like to thank the three anonymous reviewers for their insightful suggestions and careful reading of the manuscript. This work has received funding from the EU's Horizon 2020 Research and Innovation programme through the ELEXIS project under grant agreement No. 731015.

References

[Agirre et al.2016] Eneko Agirre, Aitor Gonzalez Agirre, Inigo Lopez-Gazpio, Montserrat Maritxalar, German Rigau Claramunt, and Larraitz Uria. 2016. Semeval-2016 task 2: Interpretable semantic textual similarity. In *SemEval-2016 Task 2: Interpretable semantic textual similarity. SemEval-2016. 10th International Workshop on Semantic Evaluation; 2016 Jun 16-17; San Diego, CA. Stroudsburg (PA): ACL; 2016. p. 512-24.* ACL (Association for Computational Linguistics).

[Ahmadi et al.2019] Sina Ahmadi, Mihael Arcan, and John McCrae. 2019. Lexical sense alignment using weighted bipartite b-matching. In *Proceedings of the LDK 2019 Workshops*. 2nd Conference on Language, Data and Knowledge (LDK 2019).

[Ahmadi et al.2020] Sina Ahmadi, John P McCrae, Sanni Nimb, Fahad Khan, Monica Monachini, Bolette S Pedersen, Thierry Declerck, Tanja Wissik, Andrea Bellandi, Irene Pisani, et al. 2020. A multilingual evaluation dataset for monolingual word sense alignment. In *Proceedings of the 12th Language Resource and Evaluation Conference (LREC 2020)*.

[Bordea et al.2015] Georgeta Bordea, Paul Buitelaar, Stefano Faralli, and Roberto Navigli. 2015. SemEval-2015 task 17: Taxonomy extraction evaluation (TExEval). In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 902–910, Denver, Colorado, June. Association for Computational Linguistics.

[Camacho-Collados et al.2018] Jose Camacho-Collados, Claudio Delli Bovi, Luis Espinosa-Anke, Sergio Oramas, Tommaso Pasini, Enrico Santus, Vered Shwartz, Roberto Navigli, and Horacio Saggion. 2018. SemEval-2018 task 9: Hypernym discovery. In *Proceedings of the 12th International Workshop on Semantic Evaluation (SemEval-2018); 2018 Jun 5-6; New Orleans, LA. Stroudsburg (PA): ACL; 2018. p. 712–24.* ACL (Association for Computational Linguistics).

[Fischer and Igel2012] Asja Fischer and Christian Igel. 2012. An introduction to restricted Boltzmann machines. In *Iberoamerican congress on pattern recognition*, pages 14–36. Springer.

[Francopoulo et al.2006] Gil Francopoulo, Monte George, Nicoletta Calzolari, Monica Monachini, Nuria Bel, Mandy Pet, and Claudia Soria. 2006. Lexical markup framework (LMF). In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy, May. European Language Resources Association (ELRA).

[Heidenreich and Williams2019] Hunter Heidenreich and Jake Williams. 2019. Latent semantic network induction in the context of linked example senses. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pages 170–180.

[Hinton2012] Geoffrey E Hinton. 2012. A practical guide to training restricted Boltzmann machines. In *Neural networks: Tricks of the trade*, pages 599–619. Springer.

[Iacobacci et al.2015] Ignacio Iacobacci, Mohammad Taher Pilehvar, and Roberto Navigli. 2015. Sensembed: Learning sense embeddings for word and relational similarity. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 95–105.

[Kernerman et al.2020] Ilan Kernerman, Simon Krek, John P. McCrae, Jorge Gracia, Sina Ahmadi, and Besim Kabashi, editors. 2020. *Proceedings of the 2020 Globalex Workshop on Linked Lexicography*, Marseille, France, May. European Language Resources Association.

[Maru et al.2019] Marco Maru, Federico Scozzafava, Federico Martelli, and Roberto Navigli. 2019. 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.

[Matuschek and Gurevych2013] Michael Matuschek and Iryna Gurevych. 2013. Dijkstra-WSA: A graph-based approach to word sense alignment. *Transactions of the Association for Computational Linguistics*, 1:151–164.

[McCrae et al.2017] John P McCrae, Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano. 2017. The Ontolex-Lemon model: development and applications. In *Proceedings of eLex 2017 conference*, pages 19–21.

[Meyer and Gurevych2010] Christian M Meyer and Iryna Gurevych. 2010. Worth its weight in gold or yet another resource—a comparative study of Wiktionary, OpenThesaurus and GermaNet. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 38–49. Springer.

[Miles and Bechhofer2009] Alistair Miles and Sean Bechhofer. 2009. SKOS simple knowledge organization system reference. *W3C recommendation*, 18:W3C.

[Miller1995] George A Miller. 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.

[Nancy and Véronis1990] I Nancy and Jean Véronis. 1990. Mapping dictionaries: A spreading activation approach. In *6th Annual Conference of the Centre for the New Oxford English Dictionary*, pages 52–64. Citeseer.

[Pantel and Pennacchiotti2008] Patrick Pantel and Marco Pennacchiotti. 2008. Automatically Harvesting and Ontologizing Semantic Relations. In *Proceedings of the 2008 Conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, page 171–195, NLD. IOS Press.

[Pedregosa et al.2011] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

[Pennington et al.2014] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

[Pilehvar and Navigli2014] Mohammad Taher Pilehvar and Roberto Navigli. 2014. A robust approach to aligning heterogeneous lexical resources. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 468–478.

[Powers2011] David Martin Powers. 2011. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.

[Saurí et al.2019] Roser Saurí, Louis Mahon, Irene Russo, and Mironas Bitinis. 2019. Cross-Dictionary Linking at Sense Level with a Double-Layer Classifier. In *2nd Conference on Language, Data and Knowledge (LDK 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[Solomonick1996] Abraham Solomonick. 1996. Towards a comprehensive theory of lexicographic definitions. In *Euralex 1996 Conference Proceedings*, pages 1–8.

[Speer et al.2016] Robyn Speer, Joshua Chin, and Catherine Havasi. 2016. ConceptNet 5.5: An open multilingual graph of general knowledge. *arXiv preprint arXiv:1612.03975*.

Appendix

ENTIRE (adjective)	
WORDNET[5]	- (of leaves or petals) having a smooth edge; not broken up into teeth or lobes - constituting the full quantity or extent; complete - constituting the undiminished entirety; lacking nothing essential especially not damaged - (used of domestic animals) sexually competent
WEBSTER[6]	- complete in all parts; undivided; undiminished; whole; full and perfect; not deficient - without mixture or alloy of anything; unqualified; morally whole; pure; faithful - not gelded; – said of a horse - internal; interior.
WIKTIONARY[7]	- (sometimes postpositive) Whole; complete. - (botany) Having a smooth margin without any indentation. - (botany) Consisting of a single piece, as a corolla. - (complex analysis, of a complex function) Complex-differentiable on all of $\mathbb{C}$. - (of a male animal) Not gelded. - morally whole; pure; sheer
MACMILLAN[8]	- used for emphasizing that you mean all or every part of something
LONGMAN[9]	- used when you want to emphasize that you mean all of a group, period of time, amount etc
Oxford[10]	[attributive] - with no part left out; whole. - Without qualification or reservations; absolute. - Not broken, damaged, or decayed. - (of a male horse) not castrated. - Botany (of a leaf) without indentations or division into leaflets.
Cambridge[11]	- whole or complete, with nothing lacking, or continuous, without interruption

Table 3: Senses of ENTIRE (adjective) in various monolingual English dictionaries