

The CogALex Shared Task on Monolingual and Multilingual Identification of Semantic Relations

Rong Xiang

Hong Kong Polytechnic University
csrxiang@comp.polyu.edu.hk

Emmanuele Chersoni

Hong Kong Polytechnic University
emmanuelechersoni@gmail.com

Luca Iacoponi

Amazon
jacoponi@gmail.com

Enrico Santus

Bayer
esantus@gmail.com

Abstract

The shared task of the CogALex-VI workshop focuses on the monolingual and multilingual identification of semantic relations. We provided training and validation data for the following languages: English, German and Chinese. Given a word pair, systems had to be trained to identify which relation holds between them, with possible choices being synonymy, antonymy, hypernymy and no relation at all.

Two test sets were released for evaluating the participating systems. One containing pairs for each of the training languages (systems were evaluated in a monolingual fashion) and the other one proposing a surprise language to test the crosslingual transfer capabilities of the systems.

Among the submitted systems, top performance was achieved by a transformer-based model in both the monolingual and in the multilingual setting, for all the tested languages, proving the potentials of this recently-introduced neural architecture.

The shared task description and the results are available at <https://sites.google.com/site/cogalexvisharedtask/>.

1 Introduction

Determining whether two words are related and what kind of relations holds between them is an important task in Natural Language Processing, and it has inspired a lot of research for more than one decade (Santus, 2016). Discovering relations between words is essential also for the creation of linguistic resources, such as ontologies and thesauri (Grefenstette, 1994), and this is especially true for specialized domains.

Research on semantic relations benefited from the success of Distributional Semantic Models (Budanitsky and Hirst, 2006; Turney and Pantel, 2010), since they allow to easily generate semantic representations for words from text, in the form of semantic vectors. However, the semantic similarity measured by vector models is an underspecified relation, and it is not easy to tell, given two similar words, in which way they are similar (Baroni and Lenci, 2011; Chersoni et al., 2016; Schulte Im Walde, 2020).

In the previous edition of the CogALex workshop, co-located with COLING 2016 in Osaka, the organizers set up a shared task dedicated to the corpus-based identification of semantic relations for English (Santus et al., 2016c). For the first time, systems were being evaluated in a shared task on the classification of multiple relations at once and, not surprisingly, the task proved to be challenging for computational models. For this new edition of the workshop, we have decided to launch a new version of the same shared task, adding more languages to the evaluation and encouraging the participants to evaluate their system also in a multilingual setting. Among the three teams that submitted their systems, the top performance was achieved by a RoBERTa-based system, XLM-R, in all the four languages, and both in the monolingual and in the multilingual setting.

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

2 Related Work

The earlier methods for identifying semantic relations were based on patterns. Patterns are generally very precise for identifying relations such as hypernymy-hyponymy (Hearst, 1992; Snow et al., 2004) and meronymy (Berland and Charniak, 1999; Girju et al., 2006), or even multiple relations at once (Pantel and Pennacchiotti, 2006), but their limit is that the two related words have to occur together in a corpus, and thus their recall is limited (Shwartz et al., 2016).

Distributional Models, which do not suffer from such limitations, became then the first choice for the NLP research on semantic relations. In a first phase, researchers focused on the similarity metric, proposing alternatives to cosine that can be more efficient in setting apart a specific semantic relation from the others, e.g. hypernymy (Weeds and Weir, 2003; Clarke, 2009; Lenci and Benotto, 2012; Santus et al., 2014a), synonymy (Santus et al., 2016a) or antonymy (Santus et al., 2014b), or looked for specific differences in their distributional contexts (Scheible et al., 2013). In parallel, the first large datasets for evaluating the identification of semantic relations were being released, including relations such as hypernymy, cohyponymy and antonymy (Baroni and Lenci, 2011; Lenci and Benotto, 2012; Scheible and Schulte Im Walde, 2014; Weeds et al., 2014; Santus et al., 2015).

In a second phase, following the increasing popularity of publicly-available frameworks for training word embeddings such as Word2Vec (Mikolov et al., 2013) and Glove (Pennington et al., 2014), the focus quickly shifted on the usage of these vectors as features for supervised classifiers. Some of these methods train classifiers directly on pairs of vectors (Baroni et al., 2012; Weeds et al., 2014), while others compute DSMs-based metrics first and then use them as features (Santus et al., 2016b). Late attempts to conciliate similarity metrics and word embeddings brought to proposals such as APSyn (Santus et al., 2018).

Some of the more recent contributions proposed even more sophisticated classification approaches. (Shwartz et al., 2016; Roller and Erk, 2016) aim at integrating word embeddings with information coming from lexical patterns, which proved to be extremely accurate for detecting relations. Other researchers introduced modifications to the structure of the vector spaces with the goal of identifying a specific type of semantic relation, for example by modifying the objective function of the Word2Vec training to inject external knowledge from a lexical resource (e.g. WordNet) (Nguyen et al., 2016; Nguyen et al., 2017), or by adding an extra postprocessing step that projects the word vectors into a new space, expressly specialized for modeling the target relation (Vulić and Korhonen, 2018) or even more refined techniques of vector space specialisation (e.g. adversarial specialisation) (Kamath et al., 2019).

However, these contributions mostly tried to address one relation at a time, with rare attempts of tackling the problem in a multiclass setting. The shared task organized in coincidence with CogALex 2016 (Zock et al., 2016) was one of the few exceptions, and the low results achieved by most systems (the top F-score being 0.44) showed the difficulty of distinguishing between multiple relations at once. For this reason, we have decided to propose a similar challenge, yet including another factor of complexity: multilingualism. Considering the recent approaches that have been introduced for semantic relations in multilingual (Wang et al., 2019), crosslingual (Glavaš and Vulić, 2019) and meta learning (Yu et al., 2020) settings, we provided datasets in multiple languages (English, German, Chinese and Italian) and encouraged the participants to train their systems for both monolingual and multilingual evaluation.

3 Shared Task

The CogALex-VI shared task was organized as a friendly competition: participants had access to both training and testing datasets, which were respectively released on August 1 and September 1, 2020. The scores of the participating systems were evaluated with the official scripts, and each team had to submit a short paper containing the system description. Among the three participants that submitted their systems, one only submitted results for the English data.

3.1 Task Description

The shared task was split into two main subtasks. In subtask 1, training and validation data are provided for the following languages: English, German and Chinese. Participants are required to use the given datasets to train their model and then, utilize it to identify which relation – among synonymy, antonymy,

hypernymy and no relation at all – holds between two words in a testing set. Predictions are evaluated separately for each language. Subtask 2 aims at evaluating the crosslingual transfer capabilities of the participating systems by testing the already trained models on a surprise language, for which no training data was provided. The chosen evaluation language was Italian.

3.2 Datasets and Tasks

In order to build the CogALex-VI multilingual dataset, four data collections have been adopted: English (Santus et al., 2015), German (Scheible and Schulte Im Walde, 2014), Chinese (Liu et al., 2019) and Italian (Sucameli and Lenci, 2017). Data format was standardized across languages to obtain a word pair per line, followed by the semantic relation holding between the words. A description of the four semantic relations of the shared task is provided in Table 1.

Relation (label)	Description	Example
Synonymy (SYN)	$w1$ can be used with the same meaning of $w1$	new-novel
Antonymy (ANT)	$w1$ can be used as the opposite of $w2$	big-small
Hypernymy (HYP)	$w1$ is a kind of $w2$	cat-animal
Random (RANDOM)	$w1$ and $w2$ are not related	dog-fruit

Table 1: Description of the semantic relations

For each language, we tried to obtain a balanced distribution of pairs across classes. A stratified sampling is adopted for English, German and Chinese. 60% of the whole dataset is provided as training dataset, and 20% is used as validation set for above languages. Participants are expected to use the above data for model and parameter tuning. The remaining 20% is given as a test set, with no ground truth. Detailed class statistics can be found in Table 2 (no training and validation data was provided for Italian).

	English			German			Chinese			Italian
	train	valid	test	train	valid	test	train	valid	test	test
SYN	842	259	266	782	272	265	402	129	122	187
ANT	916	308	306	829	275	281	361	136	142	144
HYP	898	292	279	841	294	286	421	145	129	153
RANDOM	2554	877	887	2430	786	796	1330	428	445	523
TOTAL	5210	1736	1738	4882	1627	1628	2514	838	838	1007

Table 2: Dataset Statistics

3.3 Participating Teams

Three participants submitted their system to CogALex-VI shared task: HSemID (Colson, 2020), Text2CS (Wachowiak et al., 2020) and TransDNN (Karmakar and McCrae, 2020). All teams took part in subtask 1, while only two of them participated in subtask 2.

Text2TCS exploited a multilingual language model based on XLM-RoBERTa (Conneau et al., 2020), which is pretrained on 100 different languages using CommonCrawl data. To adapt the system to the task, the authors appended a linear layer to XLM-R, followed by a softmax for the classification. This system was fine-tuned on the three training set from different languages simultaneously.

TransDNN proposed an architecture combining BERT (Devlin et al., 2018), LSTM and CNN, in which the BERT embeddings are passed to an LSTM that helps to represent terms having multiple words, and finally reach a convolutional layer followed by a dense layer and a softmax, devised for the classification. This system was trained on the given English dataset and participated only in the first subtask.

HSemID proposed a multilayer perceptron combining 1st and 2nd order representations of semantic associations. The system was trained with default parameters and the representations were built on WaCky corpora for English, German, Italian and a translated WaCky corpus (Baroni et al., 2009) for Chinese.

The methods and corpora used are summarized in Table 3.

Team	Method	Corpus
Text2TCS	multilingual language model XLM-RoBERTa, multilingual training for English, German and Chinese simultaneously	XLM-RoBERTa is trained on 100 different languages using CommonCrawl data
TransDNN	BERT, LSTM, CNN	BERT is trained on 3.3 billion tokens using BookCorpus and English Wikipedia Data
HSemID	combination of 1st-order and 2nd-order representations of semantic associations, multilayer perceptron	WaCky corpora for English, German, Italian and a translated WaCky corpus based on similar seed words of Chinese.

Table 3: Description of the participating systems

4 Evaluation

For the evaluation, participants had to submit their predictions. The output files were expected to contain exactly the same pairs, in the same order, and using the same annotation labels of the gold standard.

Given the gold standard and the system output, our script calculates precision, recall and F1 score. The weighted performance average across the classes was calculated ignoring the RANDOM pairs, as our focus is on the system’s capability of detecting actual semantic relations. To this end, only SYN, ANT and HYP were averaged in the final score. The overall ranking was based on such a weighted average.

4.1 Subtask 1

Table 4 summarizes the performance of the systems in the three languages.¹ With the only exception of Text2TCS in Chinese, the relatively low F1 scores indicate that the task of identifying semantic relations is still hard to solve and that performance would benefit from more attention by the research community. Another interesting fact is that all participating systems show a similar pattern with regard to precision and recall. These systems tend to result in higher precision, while recall remains relatively low.

System	Overall Precision	Overall Recall	Overall F1
English			
Text2TCS	0.602	0.455	0.517
TransDNN	0.563	0.355	0.428
HSemID	0.400	0.276	0.320
German			
Text2TCS	0.592	0.435	0.500
HSemID	0.395	0.258	0.312
Chinese			
Text2TCS	0.904	0.860	0.881
HSemID	0.501	0.331	0.377

Table 4: Performance of participating systems for subtask 1

The best performing system is Text2TCS. It outperforms the others in every metric for three languages, achieving 0.52 F1 score for English, 0.50 for German, and 0.88 for Chinese. Due to the lack of non-English pre-trained models, TransDNN only provided results for English, ranking second with 0.43 F1 score. The gap between Text2TCS and TransDNN is lower for precision (0.039) than for recall (0.1). HSemID performs worse in subtask 1, lagging behind the other systems by a large margin.

As described in Table 3, all the systems utilize neural networks, although they differ in the architecture complexity and corpus size. In particular, transformer-based architectures demonstrate to outperform simpler approaches, such as the HSemID one.

As for the different languages, it is unexpected that systems perform best in Chinese. In order to gain insights into this surprising result, we investigated misclassification for both English and Chinese. In English, 287 pairs out of 1738 were misclassified by all the systems. Errors concerned 79 SYN, 76 ANT and 90 HYP, which were in most cases misclassified as RANDOM. This indicates that recall still needs to

¹The detailed scores by class can be found in the system description papers.

be improved. As it is shown in Table 5, RANDOM represents an important interference factor in every relation type. On the contrary, we did not find any RANDOM instance which was incorrectly classified as SYN. Another common confusion of the system was between SYN and HYP, because of the similar nature of these semantic relations (see for example the pairs "gauge-test" and "confine-constrain").

English				Chinese			
W1	W2	Gold	Pred	W1	W2	Gold	Pred
fan	blow on	SYN	RANDOM	私人(private)	公立(public)	ANT	RANDOM
workforce	loiterers	ANT	RANDOM	花(diverse)	眼花繚亂(dazed)	SYN	RANDOM
zone	reguale	HYP	RANDOM	打氣(cheer)	勉(encourage)	HYP	RANDOM
swat	hit	SYN	RANDOM	哪裏(where)	今(now)	RANDOM	SYN
misconstrue	gesture	RANDOM	HYP	貼(stick)	黏(sticky)	HYP	SYN
maze	path	ANT	HYP	狹隘(narrow)	窄(narrow)	HYP	ANT
guage	test	SYN	HYP	輕聲(whisper)	低聲(low voice)	SYN	ANT
confine	constrain	HYP	SYN	斥(blame)	使(enable)	HYP	SYN
chap	dude	SYN	ANT	醫學(medicine)	學術(academy)	HYP	SYN
arrive	rest	RANDOM	ANT	糕(cake)	糕點(cake)	HYP	SYN

Table 5: Sample of pairs that were misclassified by all systems

The common misclassified pairs for Chinese were less than for English. Only 47 instances out of 838 were wrongly classified by both Text2TCS and HSemID. It can be found in Table 5 that most errors are related to SYN and HYP. The Chinese dataset seems to show a neater distinction between semantic relations, compared to other tested languages. Moreover, it was not possible to identify any evidence for regular errors in the Chinese dataset (i.e. no relation types were more prone to be confused).

4.2 Subtask 2

Table 6 summarizes the results for the subtask 2, that is in the identification of semantic relations in Italian, proposed in the test as the surprise language. Both systems obtain metrics that are comparable to those obtained in subtask 1 for English and German. The small gap between trained models in subtask 1 and zero-shot learning approaches subtask 2 implies a certain degree of commonalities between European languages. Also in this case, Text2TCS largely outperforms HSemID. Once more, both systems retain higher precision than recall. Interestingly, however, the gap between the systems is this time higher for precision (0.192) than for recall (0.133).

System	Overall Precision	Overall Recall	Overall F1
Text2TCS	0.557	0.429	0.477
HSemID	0.365	0.296	0.325

Table 6: Performance of participating systems for subtask 2 (Italian)

5 Conclusions

In this paper, we have described the CogALex-VI shared task, which focused on monolingual and multilingual identification of semantic relations. Three teams have submitted their system. All of them have addressed subtask 1 (i.e. identifying relations in languages for which a training set was released: English, German and Chinese), while only two teams addressed subtask 2 (i.e. identifying relations in a surprise language: Italian). All the submitted systems utilized neural networks, but with different level of complexity. The evaluation shows that transformer-based approaches obtain better performance than other neural methods. These approaches, in particular, are also behind the recent developments in natural language processing, showing incredible capabilities in a large set of domains and applications. Probably because of such capabilities, traditional NLP tasks, such as the identification of semantic relations, are gradually losing traction in the community. This shared task meant to show how such core NLP tasks remain a big challenge, which would require more attention by the community to possibly generate even more powerful and robust semantic representations.

References

- Marco Baroni and Alessandro Lenci. 2011. How We BLESSED Distributional Semantic Evaluation. In *Proceedings of the ACL Workshop on GEometrical Models of Natural Language Semantics*.
- Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language resources and evaluation*, 43(3):209–226.
- Marco Baroni, Raffaella Bernardi, Ngoc-Quynh Do, and Chung-chieh Shan. 2012. Entailment Above the Word Level in Distributional Semantics. In *Proceedings of EACL*.
- Matthew Berland and Eugene Charniak. 1999. Finding Parts in Very Large Corpora. In *Proceedings of ACL*.
- Alexander Budanitsky and Graeme Hirst. 2006. Evaluating Wordnet-Based Measures of Lexical Semantic Relatedness. *Computational Linguistics*, 32(1):13–47.
- Emmanuele Chersoni, Giulia Rambelli, and Enrico Santus. 2016. CogALex-V Shared Task: ROOT18. In *Proceedings of the COLING Workshop on Cognitive Aspects of the Lexicon*.
- Daoud Clarke. 2009. Context-Theoretic Semantics for Natural Language: An Overview. In *Proceedings of the EACL workshop on Geometrical Models of Natural Language Semantics*.
- Jean-Pierre Colson. 2020. Extracting Meaning by Idiomaticity: Description of the HSemID system at CogALex VI (2020). In *Proceedings of the COLING Workshop on Cognitive Aspects of the Lexicon*.
- Alexis Conneau, Alexei Baevski, Ronan Collobert, Abdelrahman Mohamed, and Michael Auli. 2020. Unsupervised cross-lingual representation learning for speech recognition. *arXiv preprint arXiv:2006.13979*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Roxana Girju, Adriana Badulescu, and Dan Moldovan. 2006. Automatic Discovery of Part-Whole Relations. *Computational Linguistics*, 32(1):83–135.
- Goran Glavaš and Ivan Vulic. 2019. Generalized Tuning of Distributional Word Vectors for Monolingual and Cross-lingual Lexical Entailment. In *Proceedings of ACL*.
- Gregory Grefenstette. 1994. Explorations in Automatic Thesaurus Discovery. In *PubMed Abstract| Publisher Full Text OpenURL*.
- Marti A Hearst. 1992. Automatic Acquisition of Hyponyms from Large Text Corpora. In *Proceedings of COLING*.
- Aishwarya Kamath, Jonas Pfeiffer, Edoardo Ponti, Goran Glavaš, and Ivan Vulić. 2019. Specializing Distributional Vectors of All Words for Lexical Entailment. In *Proceedings of the ACL Workshop on Representation Learning for NLP*.
- Saurav Karmakar and John P McCrae. 2020. CogALex-VI Shared Task: Bidirectional Transformer-Based Identification of Semantic Relations. In *Proceedings of the COLING Workshop on Cognitive Aspects of the Lexicon*.
- Alessandro Lenci and Giulia Benotto. 2012. Identifying Hypernyms in Distributional Semantic Spaces. In *Proceedings of *SEM*.
- Hongchao Liu, Emmanuele Chersoni, Natalia Klyueva, Enrico Santus, and Chu-Ren Huang. 2019. Semantic Relata for the Evaluation of Distributional Models in Mandarin Chinese. *IEEE access*, 7:145705–145713.
- Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.
- Kim Anh Nguyen, Sabine Schulte im Walde, and Ngoc Thang Vu. 2016. Integrating Distributional Lexical Contrast into Word Embeddings for Antonym-Synonym Distinction. In *Proceedings of ACL*.
- Kim Anh Nguyen, Maximilian Köper, Sabine Schulte im Walde, and Ngoc Thang Vu. 2017. Hierarchical Embeddings for Hypernymy Detection and Directionality. *Proceedings of EMNLP*.
- Patrick Pantel and Marco Pennacchiotti. 2006. Espresso: Leveraging Generic Patterns for Automatically Harvesting Demantic Relations. In *Proceedings of COLING-ACL*.

- Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of EMNLP*.
- Stephen Roller and Katrin Erk. 2016. Relations such as hypernymy: Identifying and exploiting hearst patterns in distributional vectors for lexical entailment. In *Proceedings of EMNLP*.
- Enrico Santus, Alessandro Lenci, Qin Lu, and Sabine Schulte Im Walde. 2014a. Chasing Hypernyms in Vector Spaces with Entropy. In *Proceedings of EACL*.
- Enrico Santus, Qin Lu, Alessandro Lenci, and Chu-Ren Huang. 2014b. Taking Antonymy Mask Off in Vector Space. In *Proceedings of PACLIC*.
- Enrico Santus, Frances Yung, Alessandro Lenci, and Chu-Ren Huang. 2015. Evaluation 1.0: An Evolving Semantic Dataset for Training and Evaluation of Distributional Semantic Models. In *Proceedings of the ACL Workshop on Linked Data in Linguistics: Resources and Applications*.
- Enrico Santus, Tin-Shing Chiu, Qin Lu, Alessandro Lenci, and Chu-Ren Huang. 2016a. What a Nerd! Beating Students and Vector Cosine in the ESL and TOEFL Datasets. In *Proceedings of LREC*.
- Enrico Santus, Alessandro Lenci, Tin-Shing Chiu, Qin Lu, and Chu-Ren Huang. 2016b. Nine Features in a Random Forest to Learn Taxonomical Semantic Relations. In *Proceedings of LREC*.
- Enrico Santus, Anna Rogers, Stefan Evert, and Alessandro Lenci. 2016c. The CogALex-V Shared Task on the Corpus-Based Identification of Semantic Relations. In *Proceedings of the COLING Workshop on Cognitive Aspects of the Lexicon*.
- Enrico Santus, Hongmin Wang, Emmanuele Chersoni, and Yue Zhang. 2018. A Rank-Based Similarity Metric for Word Embeddings. In *Proceedings of ACL*.
- Enrico Santus. 2016. *Making Sense: From Word Distribution to Meaning*. Ph.D. thesis, Hong Kong Polytechnic University.
- Silke Scheible and Sabine Schulte Im Walde. 2014. A Database of Paradigmatic Semantic Relation Pairs for German Nouns, Verbs, and Adjectives. In *Proceedings of the COLING Workshop on Lexical and Grammatical Resources for Language Processing*.
- Silke Scheible, Sabine Schulte Im Walde, and Sylvia Springorum. 2013. Uncovering Distributional Differences Between Synonyms and Antonyms in a Word Space Model. In *Proceedings of IJCNLP*.
- Sabine Schulte Im Walde. 2020. Distinguishing between Paradigmatic Semantic Relations across Word Classes: Human Ratings and Distributional Similarity. *Journal of Language Modelling*, 8:53–101.
- Vered Shwartz, Yoav Goldberg, and Ido Dagan. 2016. Improving Hypernymy Detection with an Integrated Path-Based and Distributional Method. In *Proceedings of ACL*.
- Rion Snow, Daniel Jurafsky, and Andrew Y Ng. 2004. Learning Syntactic Patterns for Automatic Hypernym Discovery. In *Proceedings of NIPS*.
- Irene Sucameli and Alessandro Lenci. 2017. PARAD-it: Eliciting Italian Paradigmatic Relations with Crowdsourcing. In *Proceedings of CLIC.it*.
- Peter D Turney and Patrick Pantel. 2010. From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, 37:141–188.
- Ivan Vulić and Anna Korhonen. 2018. Injecting Lexical Contrast into Word Vectors by Guiding Vector Space Specialisation. In *Proceedings of The ACL Workshop on Representation Learning for NLP*.
- Lennart Wachowiak, Christian Lang, Barbara Heinisch, and Dagmar Gromann. 2020. CogALex-VI Shared Task: Transrelation - A Robust Multilingual Language Model for Multilingual Relation Identification. In *Proceedings of the COLING Workshop on Cognitive Aspects of the Lexicon*.
- Chengyu Wang, Yan Fan, Xiaofeng He, and Aoying Zhou. 2019. A Family of Fuzzy Orthogonal Projection Models for Monolingual and Cross-lingual Hypernymy Prediction. In *The World Wide Web Conference*.
- Julie Weeds and David Weir. 2003. A general Framework for Distributional Similarity. In *Proceedings of EMNLP*.
- Julie Weeds, Daoud Clarke, Jeremy Reffin, David Weir, and Bill Keller. 2014. Learning to Distinguish Hypernyms and Co-hyponyms. In *Proceedings of COLING*.

Changlong Yu, Jialong Han, Haisong Zhang, and Wilfred Ng. 2020. Hypernymy Detection for Low-Resource Languages via Meta Learning. In *Proceedings of ACL*.

Michael Zock, Alessandro Lenci, and Stefan Evert. 2016. Proceedings of the COLING Workshop on Cognitive Aspects of the Lexicon (CogALex-V). In *Proceedings of the 5th Workshop on Cognitive Aspects of the Lexicon (CogALex-V)*.