On Universal Colexifications

Hongchang Bao **Bradley Hauer** **Grzegorz Kondrak**
Alberta Machine Intelligence Institute, Department of Computing Science
University of Alberta, Edmonton, Canada
{hongchan,bmhauer,gkondrak}@ualberta.ca

Abstract

Colexification occurs when two distinct concepts are lexified by the same word. The term covers both polysemy and homonymy. We posit and investigate the hypothesis that no pair of concepts are colexified in every language. We test our hypothesis by analyzing colexification data from BabelNet, Open Multilingual WordNet, and CLICS. The results show that our hypothesis is supported by over 99.9% of colexified concept pairs in these three lexical resources.

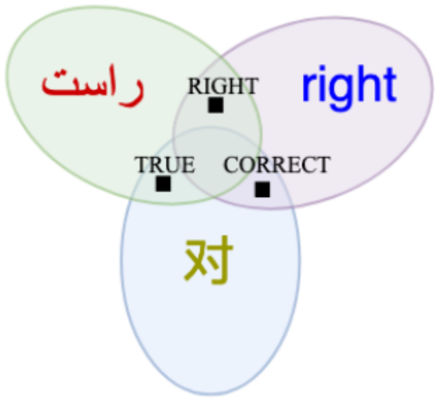

Figure 1: Three concepts (RIGHT, TRUE, CORRECT) that are colexified in Persian, English, and Chinese.

1 Introduction

Colexification refers to the phenomenon of multiple concepts in the same language being lexified by a single word (François, 2008). For example, the English word *right* colexifies the concepts of RIGHT (side) and CORRECT (Figure 1). The term covers both polysemy and homonymy (Pericliev, 2015). In this paper, we posit and investigate the hypothesis that there are no universal colexifications, or more precisely, that *no two distinct concepts are colexified in every language*.

The universal colexification hypothesis is relevant for the task of word sense disambiguation because it would imply that any sense distinction in any language could be disambiguated by translation into some language. It is also related to a famous proposal of Resnik and Yarowsky (1997) "to restrict a word sense inventory to those distinctions that are typically lexicalized cross-linguistically". If there are no universal colexifications, then a sense inventory based on cross-lingual translation pairs would also include all core concepts in existing lexical resources, which would cast doubt on the commonly expressed opinion that WordNet is too fine-grained (Pasini and Navigli, 2018).

We test our hypothesis by analyzing the colexification data from three different lexical resources: BabelNet (BN), Open Multilingual WordNet (OMWN), and the Database of Cross-Linguistic Colexifications (CLICS). Taken together, these resources contain over a million lexifications in three thousand languages. The results show that our hypothesis is supported by over 99.9% of colexified concept pairs in these three lexical resources.

The structure of the paper is as follows: In Section 2, we introduce terminology and background knowledge, formalize the concepts of lexification and colexification, and state our hypothesis. In Section 3, we summarize previous research related to colexification. In Section 4, we discuss the sources of colexification information which we use to test the hypothesis. Section 5 describes how we construct a colexification database from each of these resources. In Section 6 we present the empirical verification of the colexification hypothesis and analyze these results further. Section 7 concludes the paper.

2 Colexification

We begin by describing the terminology and background knowledge which contextualizes our work. In particular, we discuss the phenomenon of colexification and how it relates to synonymy, translation, and WordNet. We then provide a formal treatment of these concepts, inspired by the formalization of homonymy and polysemy of Hauer and Kondrak (2020a). Finally, we formally state and discuss our hypothesis.

2.1 Background

Princeton WordNet (Fellbaum, 1998), and similarly structured resources, consist of *synsets*. Each synset contains one or more words that can be used to express a specific *lexicalized concept*, or simply *concept* (Miller, 1995). A word *lexifies* a concept if it can be used to express that concept; that is, if the corresponding synset contains that word. Each content word lexifies at least one concept. Each concept that a word can express corresponds to a *sense* of that word. *Word sense disambiguation*, the task of determining the sense of a word in context, is one of the central tasks in computational lexical semantics and natural language understanding (Navigli, 2018).

If two words in the same language lexify a single concept, such as *heart* and *core*, the words are *synonyms*. If two words in different languages lexify a single concept, such as *apple* and *pomme*, the words are *translational equivalents*. Synonymy and translational equivalence are the intra-lingual and inter-lingual components of the relation of *semantic equivalence*, or sameness of meaning (Hauer and Kondrak, 2020b). Indeed, multilingual wordnets (multi-wordnets) such as BabelNet (Navigli and Ponzetto, 2012) consist of multilingual synsets (multi-synsets), which contain words in many languages, each lexicalizing the concept that corresponds to that multi-synset. Multi-wordnets may be constructed by adding translations to the monolingual synsets of a pre-existing wordnet, typically WordNet itself (the *expand model*), or by linking the synsets of multiple independently constructed wordnets in different languages via an inter-lingual index (the *merge model*) (Vossen, 1996).

If two concepts are referred to by a single word, the concepts are *colexified* by that word. In WordNet terms, if two synsets have a non-empty intersection, each word in that intersection colexifies the concepts to which those synsets correspond. Some colexifications, such as the *bank* example above, are coincidental, arising only due to *homonymy*, the use of a single word to represent distinct, semantically unrelated entries in the lexicon (Hauer and Kondrak, 2020a). Other colexifications arise between concepts that are semantically related (Youn et al., 2016).

Lexification and colexification are language dependent. For any given concept, each language may have zero, one, or more synonymous words that lexify it, cases that correspond to the notions of a lexical gap, monolexical synset, and synonymy, respectively. For example, there is no Chinese word which colexifies the two concepts colexified by the English *right* in the example mentioned in Section 1. A language colexifies two concepts if it contains a word which colexifies them. For example, English colexifies the concepts RIGHT and CORRECT; Chinese does not.

2.2 Formalization

Let $\mathcal{C}$ be the set of all concepts. Let $\mathcal{L}$ be the set of all languages. For each language $E \in \mathcal{L}$, let $\mathcal{V}_E$ be the lexicon of E, the set of all words in E. Further, for each concept $c \in \mathcal{C}$, $w_E(c)$ is the set of words in E which lexify c; that is, w_E is a function from $\mathcal{C}$ to $\mathcal{P}(\mathcal{V}_E)$, where $\mathcal{P}$ denotes the *power set* of a set, the set of all the subsets of a set. If $w_E(c) = \emptyset$, c is a lexical gap in E; that is, no word in E lexifies c. Otherwise, if $w_E(c) \neq \emptyset$, c is lexified in E.

Two concepts $c_1, c_2 \in \mathcal{C}$ are colexified by language E if and only if $w_E(c_1) \cap w_E(c_2) \neq \emptyset$. We define $COL(c_1, c_2)$ as the set of languages that colexify c_1 and c_2, and $LEX(c_1, c_2)$ as the set of languages that lexify both c_1 and c_2:

$$COL(c_1, c_2) = \{E \in \mathcal{L} \mid w_E(c_1) \cap w_E(c_2) \neq \emptyset\}$$

$$LEX(c_1, c_2) = \{E \in \mathcal{L} \mid w_E(c_1) \neq \emptyset \neq w_E(c_2)\}$$

Obviously, $COL(c_1, c_2) \subseteq LEX(c_1, c_2)$.

For the purposes of analyzing colexification, we introduce the *colexification ratio*: for any pair of concepts, their colexification ratio is equal to the number of languages which colexify the concepts divided by the number of languages which lexify both concepts. Formally, we define the colexification ratio between two concepts as:

$$r(c_1, c_2) := \frac{|COL(c_1, c_2)|}{|LEX(c_1, c_2)|}$$

$r(c_1, c_2)$ is undefined if $LEX(c_1, c_2) = \emptyset$.

2.3 Hypothesis

We propose the following hypothesis: no pair of concepts is colexified in every language. More precisely, for any pair of concepts that are colexified in some language, there exists another language that lexifies both concepts but does not colexify them. Formally:

$$\forall c_1, c_2 \in \mathcal{C},\ \exists E \in \mathcal{L}\, s.t.\, w_E(c_1) \cap w_E(c_2) \neq \emptyset$$

$$\Rightarrow \exists F \in \mathcal{L}\, s.t.\, w_F(c_1) \neq \emptyset \neq w_E(c_2)$$

$$\wedge\ w_F(c_1) \cap w_F(c_2) = \emptyset$$

Equivalently, our hypothesis predicts that for every pair of concepts, the colexification ratio is either undefined or less than one:

$$\forall c_1, c_2 \in \mathcal{C}\ |LEX(c_1, c_2)| > 0$$

$$\Rightarrow r(c_1, c_2) < 1$$

This equivalence can be seen by simply substituting r, LEX and COL with the definitions given in Section 2.2, and applying some basic principles of set theory.

3 Related Work

Approaches to colexification can be divided into three types, which are based on semantic maps, graphs, and databases, respectively.

The semantic-map approach to colexification is introduced by Haspelmath (2000), who focuses on distinguishing senses in the grammatical domain. Semantic maps are constructed by cross-linguistic comparison, and contain concepts that have distinct colexifications in at least two different languages. Their experiments show that 12 diverse languages are sufficient to build a stable semantic map. Our hypothesis relates this statement to entire lexicons of core concepts. François (2008) also uses colexification data to build a semantic map for studying the world's lexicons across languages. He observes that the more languages are considered, the more distinctions between senses need to be made. This finding is consistent with our hypothesis, and also raises another open question: is a given pair of colexified concepts colexified universally?

The graph-based approach is introduced by List and Terhalle (2013), who analyze cross-linguistic polysemy. They build a weighted colexification graph using data from 195 languages representing 44 language families, and find that clusters of closely-related or similar concepts are often densely connected. Youn et al. (2016) construct colexification graphs in the domain of natural objects to verify if human conceptual structure is universal. Analysis reveals universality of similar patterns in semantic structure, even across different language families.

The database approach is used by Pericliev (2015), who studies colexifications of 100 basic concepts, and introduces heuristics for distinguishing between homonymy and polysemy. Georgakopoulos et al. (2020) use a colexification database to study commonalities between languages in the domain of perception-cognition. They analyze the colexification of four concepts related to perception (SEE, LOOK, HEAR, and LISTEN) to reveal connections between verbs of vision and hearing.

4 Resources

In this section, we describe our three resources: BabelNet (BN), Open Multilingual WordNet (OMWN), and CLICS. Table 2 contains the number of concepts and languages that we consider in each of these resources. For instance, CLICS contains approximately one million words in 3050 languages, which express 2919 concepts. The other two resources have fewer languages, but a higher average number of words per language.

BabelNet (Navigli and Ponzetto, 2012) is a multi-wordnet automatically constructed using the expand model based on the Princeton WordNet. It combines data from Wikipedia, Wikidata, OmegaWiki, and various other resources, supplemented by machine translation, to cover nearly 300 distinct languages. Each of the multi-synsets in BN corresponds to a unique concept, with a unique eight-digit identifier, and an associated part of speech (noun, verb, adjective, or adverb), and contains one or more words which can express that concept in various languages. For instance, the nominal concept TREE is represented by synset `bn:00078131n` which includes the English words *tree* and *arbor*, as well as French *arbre* and Italian *albero*. We use BabelNet version 4.0.

Open Multilingual WordNet (Bond and Foster, 2013) is another multilingual wordnet, constructed by linking wordnets in 29 languages to WordNet version 3.0. Like BN, OMWN consists of multi-synsets, each containing one or more words from

Resource	Colexified Concept Pair	COL	LEX	Ratio
CLICS	LEG - FOOT	336	1038	0.324
	WOOD - TREE	335	1036	0.323
	MOON - MONTH	313	538	0.582
BN	town.n.01 - city.n.01	100	121	0.826
	painting.n.01 - image.n.01	89	93	0.957
	house.n.01 - dwelling.n.01	88	117	0.752
OMWN	book.n.02 (work) - book.n.01 (object)	23	25	0.920
	wing.n.02 (airplane) - wing.n.01 (animal)	22	22	1.000
	shout.v.02 (cry) - shout.v.01 (with loud voice)	22	24	0.917

Table 1: The concept pairs colexified by the most languages in each of the three databases.

one or more languages which lexify a particular concept. For example, *sign* and *mark* (English), and *signe*, *témoignage*, *preuve*, and *point* (French) all share a multi-synset.

OMWN is based on a set of 5000 *core concepts*, constructed by Boyd-Graber et al. (2006)[1]. This list was updated to WordNet 3.0 by the creators of OMWN[2]. Every WordNet 3.0 synset in this list corresponds to exactly one multi-synset in OMWN, and exactly one multi-synset in BN. Indeed, both resources are created by applying the expand model to WordNet 3.0. For the purposes of our work, we limit OMWN and BN to their respective 5000 synsets corresponding to these core concepts.

The Database of Cross-Linguistic Colexifications (CLICS) (Rzymski and Tresoldi, 2019) is an online lexical database containing information on cross-linguistic colexification patterns across thousands of languages from hundreds of language families. CLICS does not follow any wordnet model, but instead integrates word lists representing thousands of languages, which vary greatly in terms of lexicon coverage. Colexification patterns are represented in the form of a network, where the weights express the number of languages that colexify the concept pair. We obtained the data following the procedure of List (2018), which directly facilitates access to colexification data for any concept pair. CLICS also contains information on the family each language belongs to.

5 Method

For each of the resources described above, we use the following procedure to create a database of concept pairs and colexification information.

The first step is to extract from each resource the set of concepts it contains, and the set of words lexifying each concept. For CLICS, this is relatively straightforward, as the resource is already structured as a database of concepts and lexifications for each language. We access OMWN through NLTK[3], and BN via its Java API[4]. Each concept in these resources is represented by a multi-synset, which can be extracted using the aforementioned APIs.

The second step is to map each of the three sets of concepts to each other, so that identical concepts in distinct resources can be associated with one another for our analysis. This is done by using WordNet 3.0 as a pivot. As described in Section 4, each of the 5000 core concepts in BN and OMWN is already linked to a WordNet 3.0 synset. However, mapping CLICS to WordNet is not trivial because, unlike BN and OMWN multi-synsets, CLICS concepts have no intrinsic connection to WordNet synsets. Therefore, we use a Concepticon mapping created by List et al. (2016) which links a subset of CLICS concepts to WordNet. Unfortunately, the mapping is incomplete, covering only 1368 (46.9%) of CLICS concepts.

The third step is to enumerate all pairs of distinct concepts. There are approximately 4.3 million possible concept pairs in CLICS, and 12.5 million possible concept pairs in BN and OMWN. Although there are millions of concept pairs in each resource, only a subset are lexified by some language (i.e. there exists a language with at least

[1] `https://wordnetcode.princeton.edu/standoff-files/core-wordnet.txt`

[2] `http://compling.hss.ntu.edu.sg/omw/wn30-core-synsets.tab`

[3] `https://www.nltk.org/`

[4] `https://babelnet.org/guide`

Resource	Languages	Concepts	Lexifications	Colexifications	Exceptions	Support
CLICS	3050	2919	1,377,282	75,089	64	99.9%
BN	284	5000	1,441,990	88,907	3	99.9%
OMWN	29	5000	267,503	54,615	4	99.9%

Table 2: The statistics on the lexical resources, and the empirical validation of our hypothesis.

one word for each concept), and only a subset of those are colexified by some language (i.e. there exists a language with a single word for both concepts). So, we are working with a subset of a subset of all concept pairs.

The fourth step is to determine which concept pairs are colexified, that is, have words in common. This consists of testing whether the intersection of the corresponding synsets (for BN and OMWN) or the corresponding database entries (for CLICS) are non-empty. We report the number of concept pairs which are colexified in at least one language in Table 2. For each pair of concepts, we record the number of languages that colexify the pair. For example, the CLICS database lists 980 languages that lexify both RIGHT (side) and CORRECT. Taking the intersection of the words lexifying each concept, we find that 41 languages have a word which lexifies both concepts, that is, 41 languages colexify these concepts in the CLICS database. Therefore, the colexification ratio for this concept pair, in CLICS, is $41/980 \approx 0.042$.

Our hypothesis states that the colexification ratio for any concept pair, for any of our databases, is always less than 1, given that it is defined. That is, there is always some language that lexifies both concepts, but does not colexify them.

6 Results

In this section, we describe the empirical validation of our hypothesis on the colexification data from CLICS, BN, and OMWN. Our results are summarized in Table 2, which shows that all three resources provide very strong evidence for our hypothesis. Namely, 99.9% of all colexified concept pairs have a colexification ratio less than 1 in all three resources. We find only 71 apparent exceptions in the individual resources.

The three most frequently colexified concept pairs in each resource are shown in Table 1. For example, the concepts LEG and FOOT are both lexified in 1038 languages (i.e. CLICS contains words for them in those languages) but only 336 languages colexify both concepts (i.e. have a single word that can express both of them). So, the colexification of LEG and FOOT is far from universal. In fact, approximately 76% of the 75,089 colexified concept pairs in CLICS are colexified in only a single language.

6.1 Analysis

The 71 apparent exceptions to our hypothesis must be qualified by the fact that none of the three resources makes any claim of completeness. For each seemingly universal colexification, it may be the case that there exists a language that lexifies both concepts, and does not colexify them, but this fact is not recorded in the corresponding database. In this section, we perform a cross-database analysis, to investigate how many, if any, of these apparent exceptions are actual counterexamples to our hypothesis, and how many are simply the result of resource incompleteness.

For example, there are only six languages[5] which lexify both of the concepts DULL and BLUNT in CLICS. This is surprising, as English words lexifying these concepts are, in fact, used to name them. However, the concept DULL does not have the English word "dull" listed in CLICS. All six of the languages which do lexify both of these concepts have a single word which lexifies both; based on our criteria, this would represent a universal colexification, if CLICS was fully complete and correct. However, by cross-checking this example against the information in the other two resources, we find several languages that do not colexify the two concepts.

The 64 apparent exceptions in CLICS involve 113 distinct concepts. Unfortunately, in all 64 cases, at least one of the concepts is not mapped any of the WordNet core synsets. To remedy this, we manually map a subset of the 64 exceptions to OMWN and BabelNet. We choose all four instances that are colexified in more than two languages, plus ten more instances that are selected at random. We find that none of the these 14 pairs are

[5] Indonesian, Klon, Lavukaleve, Mbaniata, Mbilua, Savosavo

Colexified Concept Pair	CLICS Ratio	BN Ratio	OMWN Ratio
RUN_AWAY - FLEE	**10/10**	24/36	13/17
DULL - BLUNT	**6/6**	34/37	6/8
RIVER - FLOWING_BODY_OF_WATER	**4/4**	2/69	0/20
FISHING - CASSOWARY	**3/3**	0/45	0/12
SKIN (human) - SKIN (animal)	**3/3**	10/13	7/10
SAME_SEX_OLDER_SIBLING - BROTHER	**2/2**	44/96	9/16
PIMPLE - BOIL (of skin)	**2/2**	37/63	5/15
MALE - BRASS_INSTRUMENT	**1/1**	0/41	0/13
GAZELLE - DEER	**1/1**	4/79	0/17
WRAPPER - DRESS	**1/1**	1/51	0/12
HYENA - CART	**1/1**	0/55	0/16
ECHIDNA - ANTEATER	**1/1**	6/58	4/10
STRIKE - CAST	**1/1**	0/20	0/14
WRAPPER - CLOTH	**1/1**	0/53	0/12
intention.n.03 - purpose.n.01	n/a	**19/19**	14/15
reserve.v.03-reserve.v.04 (hold)	n/a	**20/20**	14/16
increase.n.04 - increase.n.03 (increment)	n/a	**26/26**	20/22
wing.n.02 (airplane) - wing.n.01 (animal)	n/a	31/47	**22/22**
short.a.01 (time) - short.a.02 (length)	n/a	36/37	**20/20**
probability.n.01 - probability.n.02 (event)	n/a	32/33	**18/18**
new.a.01 (time) - new.s.11 (unfamiliar)	n/a	18/19	**16/16**

Table 3: The concept pairs with the ratio of 1 represent possible exceptions to our hypothesis. The fact that the corresponding ratio is less than 1 in another resource provides evidence against the exception.

exceptions in OMWN or BN (Table 3). In other words, there is at least one language in each of OMWN and BN that lexifies the pairs but does not colexify them. Based on this analysis, we conclude that the 14 exceptions are caused by data sparsity.

In BabelNet, there are only three apparent exceptions to our hypothesis (Table 3). Considering BabelNet alone, they appear to be counterexamples to our hypothesis. Unfortunately, the corresponding WordNet concepts are not mapped to CLICS concepts. However, we find that none of these three pairs are exceptions in OMWN; for all three, the OMWN colexification ratio is less than 1. For example, Chinese lexifies `reserve.v.03` as *liu* and `reserve.v.04` as *ding*. Based on this analysis, we conclude that the three apparent exceptions in BabelNet are artifacts of data sparsity.

The situation in OMWN is similar: we find only four apparent exceptions, and none of them are exceptions in BabelNet. For example, according to BabelNet, Icelandic lexifies “new.a.01 (time)” as *nýr*, and “new.s.11 (unfamiliar)” as *óþekktur*, but no Icelandic word lexified both concepts.

7 Conclusion

We have proposed a novel hypothesis which states that there are no universal colexifications. We provided evidence that the few apparent exceptions to the hypothesis that we found in three multilingual resources are attributable to omission errors in the resources. In the future, we plan to leverage our hypothesis to improve the accuracy of multilingual word sense disambiguation.

The validation of our hypothesis provides novel insights into several open issues in lexical semantics. It implies that every sense distinction in every language can be disambiguated by translation into some language. It also provides support for the informal conjecture of Palmer et al. (2007) that every possible sense distinction can be identified by translation into multiple languages. Finally, it furnishes evidence that the fine-granularity of wordnets and multi-wordnets is necessary for distinguishing between lexical translations of concepts.

Acknowledgments

This research has been supported by the Natural Sciences and Engineering Research Council of

Canada (NSERC), and the Alberta Machine Intelligence Institute (Amii).

References

Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual Wordnet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1352–1362, Sofia, Bulgaria, August.

J. Boyd-Graber, C. Fellbaum, D. Osherson, and R. Schapire. 2006. Adding dense, weighted connections to wordnet. In *In: Proceedings of the Third Global WordNet Meeting, Jeju Island, Korea*.

Christiane Fellbaum. 1998. WordNet: An on-line lexical database and some of its applications. *MIT Press*.

Alexandre François. 2008. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages. *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations*, 163-215.

A. Georgakopoulos, E. Grossman, D. Nikolaev, and S. Polis. 2020. Universal and macro-areal patterns in the lexicon. *Linguistic Typology*.

Martin Haspelmath. 2000. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. *The new psychology of language*.

Bradley Hauer and Grzegorz Kondrak. 2020a. One homonym per translation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7895–7902.

Bradley Hauer and Grzegorz Kondrak. 2020b. Synonymy = translational equivalence. *arXiv preprint arXiv:2004.13886*.

Johann-Mattis List and Anselm Terhalle. 2013. Using network approaches to enhance the analysis of cross-linguistic polysemies. *Proceedings of the 10th International Conference on Computational Semantics*.

Johann-Mattis List, Michael Cysouw, and Robert Forkel. 2016. Concepticon: A resource for the linking of concept lists. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 2393–2400, Portorož, Slovenia, May. European Language Resources Association (ELRA).

Johann-Mattis List. 2018. Cooking with clics. *Computer-assisted language comparison in practice*, 14-18.

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

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Roberto Navigli. 2018. Natural language understanding: Instructions for (present and future) use. In *IJCAI*, pages 5697–5702.

Martha Palmer, Hoa Trang Dang, and Christiane Fellbaum. 2007. Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Natural Language Engineering*, 13(2):137–163.

Tommaso Pasini and Roberto Navigli. 2018. Two knowledge-based methods for high-performance sense distribution learning. In *Proc. of the 32th AAAI Conference on Artificial Intelligence*.

Vladimir Pericliev. 2015. On colexification among basic vocabulary. *Journal of Universal Language*, 63-93.

Philip Resnik and David Yarowsky. 1997. A perspective on word sense disambiguation methods and their evaluation. In *Tagging Text with Lexical Semantics: Why, What, and How?*, pages 79–86.

Christoph Rzymski and Tiago et al. Tresoldi. 2019. The database of cross-linguistic colexifications, reproducible analysis of cross- linguistic polysemies. *Scientific Data*.

PJTM Vossen. 1996. Right or wrong: combing lexical resources in the eurowordnet project. In *M. Gellerstam, J. Jarborg, S. Malmgren, K. Noren, L. Rogstrom, CR Papmehl, Proceedings of Euralex-96, Goetheborg, 1996*, pages 715–728. Vrije Universiteit.

Hyejin Youn, Logan Sutton, Eric Smith, Cristopher Moore, Jon F Wilkins, Ian Maddieson, William Croft, and Tanmoy Bhattacharya. 2016. On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences*, 113(7):1766–1771.