Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 7147–7154
Marseille, 20-25 June 2022
© European Language Resources Association (ELRA), licensed under CC-BY-NC-4.0

Lexical Resource Mapping via Translations

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

Aligning lexical resources that associate words with concepts in multiple languages increases the total amount of semantic information that can be leveraged for various NLP tasks. We present a translation-based approach to mapping concepts across diverse resources. Our methods depend only on multilingual lexicalization information. When applied to align WordNet/BabelNet to CLICS and OmegaWiki, our methods achieve state-of-the-art accuracy, without any dependence on other sources of semantic knowledge. Since each word-concept pair corresponds to a unique sense of the word, we also demonstrate that the mapping task can be framed as word sense disambiguation. To facilitate future work, we release a set of high-precision WordNet-CLICS alignments, produced by combining three different mapping methods.

Keywords: lexical resources, WordNet, concept mapping, lexical translations, multilingual resource

1. Introduction

A lexical resource associates words in one or more languages with concepts they can express. Each word-concept pair corresponds to a unique sense of the word. For example, the word *plant* in Figure 1 has distinct senses corresponding to the industrial and vegetation concepts it can lexicalize. Lexical resources vary in how they are constructed and how they represent concepts and senses, which makes it difficult to combine information from multiple resources. The task that we address in this paper is mapping (aligning) concepts or senses across lexical resources. Given a concept in one of the resources, such a mapping allows us to identify a matching concept in the other resource.

Aligning concepts between lexical resources facilitates several tasks. First, combining information from multiple resources increases coverage of words and languages. For example, Open Multilingual WordNet (Bond and Foster, 2013) contains the translations of a large number of senses in all the languages covered by the aligned multilingual resources from which it was constructed. Second, in addition to word senses, lexical resources may contain other types of information, such as relations between concepts, glosses, and usage examples. By mapping the senses in one resource to their equivalents in another, information about a given concept can be retrieved from both resources, increasing the total knowledge available about each concept. Finally, lexical resources are important for various tasks in natural language processing (NLP). Inter-resource concept mapping has been shown to yield performance improvements compared to using resources in isolation (Ponzetto and Navigli, 2010).

The task of concept mapping has been addressed in prior work with a wide variety of resources, techniques, and applications. Numerous works consider mapping between Princeton WordNet and other resources, including OmegaWiki, Wiktionary, and Wikipedia.

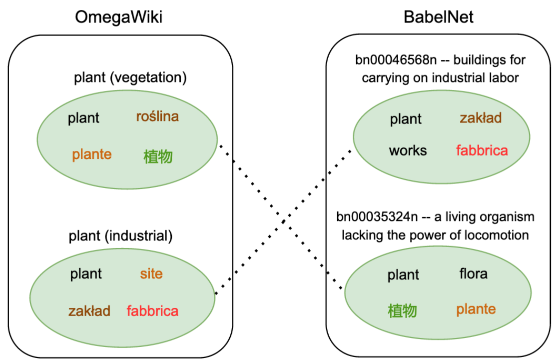

Figure 1: An overview of lexical resource mapping task

Similarity-based methods (Meyer and Gurevych, 2011) define similarity measures on pairs of concepts across the resources, which are often based on their glosses. Graph-based methods (Pilehvar and Navigli, 2014) identify equivalent senses by applying algorithms, such as Personalized PageRank, to graphs of concepts connected by various semantic relations. These paradigms take advantage of the fact that glosses and semantic relations are the two commonly-used ways to describe word senses in lexical resources.

In this paper, we propose two methods that leverage translation information to align multilingual lexical resources. We posit that the semantic similarity between concepts is strongly correlated with the number of shared lexicalizations. This idea is grounded in the sense translation synonymy formalization of Hauer and Kondrak (2020), which implies a one-to-one mapping between concepts in different languages. Our methods depend exclusively on lexicalization information, without relying on concept glosses, relations between concepts, or other structured information. We refrain from combining different types of information in order to assess how far translations can take us towards solv-

	BabelNet	CLICS	OmegaWiki
Languages	284	3050	143
Concepts	117,659	2919	50,785
Lexicalizations	15,771,915	1,377,282	248,166

Table 1: Statistics of the lexical resources.

ing the concept mapping task.

We evaluate our approach on the alignment of WordNet with two other lexical resources: CLICS and OmegaWiki. In both cases, our methods outperform three previous gloss-based methods. For the WordNet-OmegaWiki mapping, our best method matches the performance of a strong graph-based method. We also compare our methods to a gloss-based state-of-the-art word sense disambiguation method, and find that our methods achieve comparable performance on the WordNet-CLICS dataset, and easily outperform it on the WordNet-OmegaWiki dataset. To facilitate further work on this important task, we release a set of manually-verified high-precision WordNet-CLICS alignments, produced by combining three different mapping methods.[1]

2. Lexical Resources

In this section, we describe the lexical resources used in our experiments. Each resource consists of a set of concepts. Each concept is associated with a set of *lexicalizations*, words that can express the concept. Each lexicalization is considered to be a tuple composed of a lexical form and its language. For example, English *orange* is distinct from French *orange*.

Table 1 lists some statistics for each resource. The resources are diverse: CLICS contains data from more than 3000 languages, with about 400 words per language on average, while BabelNet (version 4.0.1) covers 284 languages, with about 55,000 words per language, many of which are proper nouns. The average number of lexicalizations per concept is 472 in CLICS vs. 134 in BabelNet.

Table 2 shows sample glosses and selected lexicalizations from each of the three resources. Note that the CLICS and OmegaWiki glosses are identical in this case; we found that 53.2% of CLICS concepts share their glosses with an OmegaWiki concept. This has no impact on our methods, which make no assumptions about the quality or availability of glosses, and instead exclusively use lexicalizations.

2.1. WordNet and BabelNet

Each concept in WordNet (Miller, 1995) is associated with a manually crafted set of synonyms (a *synset*) which lexicalize that concept in English. WordNet was constructed according to the principle that each synset should consist of words which are interchangeable in some context without altering the meaning of the expression. Each synset is associated with a gloss describing its concept, a part of speech (noun, verb, adjective, or adverb), and, optionally, one or more usage examples.

[1] https://www.cs.ualberta.ca/~kondrak

Resource	Gloss	Lexicalizations
WordNet + BabelNet	one skilled in caring for young children or the sick	IT: infermiere, infermiera FR: aide-soignant, infirmiere EN: nurse
CLICS	a person trained to provide care for the sick	EN: nurse NL: verpleegster ID: perawat, suster
OmegaWiki	a person trained to provide care for the sick	EN: nurse FR: infirmier, aide-soignant IT: infermiere

Table 2: Sample glosses and lexicalizations for the concept of NURSE from three lexical resources.

Lexical resources use different conventions to refer to concepts. In WordNet, a synset is typically referred by one of its lexicalizations, along with the part of speech and a number. For example, the synset $play_n^1$ contains the nouns {*play, drama, dramatic play*}, has the gloss "A dramatic work intended for performance by actors on a stage" and the usage example "He wrote several plays but only one was produced on Broadway."

BabelNet (Navigli and Ponzetto, 2012) expands WordNet's synsets with multilingual lexicalizations. It also adds entirely new concepts, including named entities. New lexicalizations and concepts are extracted from a variety of resources, such as Open Multilingual WordNet (Bond and Foster, 2013), Wikipedia (Mihalcea, 2007), and OmegaWiki (de Melo and Weikum, 2010). One important consequence of this construction method is that every WordNet synset is associated with a multilingual synset in BabelNet. Therefore, we can obtain from BabelNet a set of multilingual lexicalizations for any WordNet concept. For example, the multilingual synset associated with $play_n^1$ contains the French lexicalizations *piece de theatre* and *drame*, and the Chinese lexicalizations *xìjù* and *jùběn*.

2.2. CLICS

CLICS, The Database of Cross-Linguistic Colexifications (Rzymski et al., 2020) was constructed by integrating word lists representing thousands of languages. Each concept is associated with a gloss, a set of lexicalizations, and one of six categories (e.g., Person/Thing), but no part of speech. Unlike in WordNet, each concept is also assigned a unique name consisting of an English word or phrase which describes its meaning. No semantic relations between concepts are provided.

2.3. OmegaWiki

OmegaWiki is an online multilingual dictionary which can be freely edited through its website. Unlike other online multilingual dictionaries, such as Wiktionary, OmegaWiki is organized into concepts. Each concept is represented by a gloss, and associated with words from different languages, which are the lexicalizations of the concept. As with CLICS, semantic relations between concepts are not specified, but, different from CLICS and WordNet, glosses are translated into different languages, rather than being in English only.

3. Related Work

In this section we review prior work on cross-resource concept mapping. These papers vary in the resources they align, as well as in the nature of the mapping (e.g., one-to-one or many-to-many). In some cases, these differences preclude direct comparison to our own work.

There is a substantial amount of prior work on linking WordNet synsets to Wikipedia articles. There are various motivations for doing so: improving word sense disambiguation, obtaining multilingual lexicalizations, and evaluating mapping algorithms on a pair of highly dissimilar resources. Ponzetto and Navigli (2010) calculate the word similarity between "disambiguation contexts" constructed from the two resources. They compute English lexicalization overlap, but unlike ours, their approach is exclusively monolingual. Navigli and Ponzetto (2012) extend this approach by leveraging the graph of WordNet semantic relations to calculate the similarity between concepts. Finally, Pilehvar and Navigli (2014) propose a method for constructing graphs representing different lexical resources. The PageRank algorithm is then applied to these graphs to compute a similarity measure between pairs of concepts. Each of the papers considers a different subset of WordNet concepts, which makes comparison difficult.

Another sequence of papers aims at aligning senses in various resources, including WordNet, GermaNet, Wiktionary, and OmegaWiki. The method of Meyer and Gurevych (2011) is based on the similarity between sense definitions. Gurevych et al. (2012) extend this approach to align WordNet and German OmegaWiki. In particular, they use machine translation to translate the lemmas and glosses of one resource into the language of the other resource, and then compute the similarity between sense definitions. Matuschek and Gurevych (2013) propose a graph-based method which, different from Navigli and Ponzetto (2012), considers relations between all senses. A similar approach is applied by Matuschek et al. (2018) to align Wiktionary and OmegaWiki. Their method is again based on the similarity between sense definitions, and the application of the Personalized PageRank algorithm.

Translation information has also been leveraged in resource alignment. Pianta et al. (2002) propose a translation-based method to build an Italian WordNet. For each Italian word sense listed in the Italian-English bilingual dictionary, they first find a set of synset candidates in WordNet that contain at least one of the translations of that word sense. Then, they order these candidates based on several rules, such as gloss matching and synset intersection. Finally, they manually identify the correct WordNet synset among the candidates.

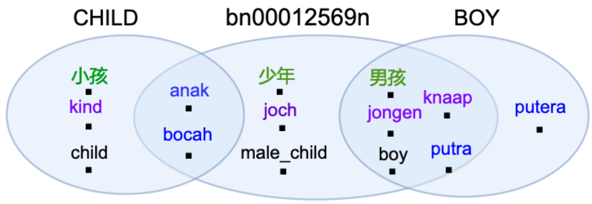

Figure 2: An example of concept lexicalization overlaps, with words from Chinese (green), Indonesian (blue), Dutch (purple), and English (black).

Similarly, Helou et al. (2016) demonstrate the effectiveness of using translations from bilingual dictionaries for concept mapping between wordnets in several languages. Our work generalizes these bilingual approaches by leveraging lexicalizations from a large number of languages.

Recently, Tjuka et al. (2021) propose a frequency-based algorithm for mapping words to concepts in the Concepticon dataset, which lists words from psychology and linguistics in 40 languages. McCrae and Cillessen (2021) apply several similarity techniques to map WordNet synsets to entities in Wikidata. Yao et al. (2021) frame the task of mapping WordNet senses to dictionaries as maximum-weight bipartite graph matching. They use a gloss similarity measure to weight the edges of a bipartite graph, and then find an exact solution to the matching problem. We include this method in our evaluation.

4. Methods

In this section, we introduce two translation-based methods for aligning concepts across lexical resources. Our methods work by maximizing a similarity measure based on the lexicalizations shared by each pair of concepts. The two methods differ in that one considers the total number of shared lexicalizations, whereas the other considers only the number of languages that exhibit shared lexicalizations.

The first method, which we refer to as WordVote, maps a given source concept to the target concept by maximizing the lexical intersection between the concepts. Formally, we define a similarity measure between two concepts as follows:

$$s_W(c_s, c_t) = |\{lex(c_s, \mathcal{L}) \cap lex(c_t, \mathcal{L})\}|$$

where $lex(c, \mathcal{L})$ is a function that returns the set of lexicalizations of the concept c in a set of languages $\mathcal{L}$.

Figure 2 shows an example: the BabelNet synset bn00012569n shares five of its ten lexicalizations (from $\mathcal{L}$ = {English, Chinese, Indonesian, Dutch}) with the CLICS concept BOY, but only two of its lexicalizations with CHILD. So, the WordVote method maps bn00012569n to BOY.

Our second method, LangVote, maps a given source concept to the target concept by maximizing the number of languages in which the two concepts share at least one lexicalization. This can be viewed as a variant of the WordVote method, in which at most one lexicalization from each language may be included in the intersection. Formally, we define this similarity measure between two concepts as:

$$s_L(c_s, c_t) = |\{L \in \mathcal{L} : lex(c_s, L) \cap lex(c_t, L) \neq \emptyset\}|$$

Returning to the example in Figure 2, bn00012569n and BOY share at least one lexicalization in all four of the languages in $\mathcal{L}$. However, this synset and CHILD have a common lexicalization only in one language, Indonesian. (The fact that there are two common Indonesian lexicalizations is not relevant for this method.) So the LangVote method maps the BabelNet synset bn00012569n to the CLICS concept BOY.

We constrain our methods to produce a one-to-one mapping, so that each concept in one resource is aligned to at most one concept in the other resource. To do this, we first align the pair of concepts with the highest similarity value. We then remove the aligned concepts from consideration, and repeat until there are no remaining concept pairs with non-zero similarity.

Both methods maximize similarity measures that return an integer value, which can result in ties between candidate concepts. To break ties between multiple candidate concepts, we select the one with the highest number of lexicalizations in a given set of languages. The intuition is that this strategy should favor more frequent concepts, which in turn should have more reliable information. In our development experiments, we found that this tie-breaking strategy works better than normalizing the similarity value by the total number of lexicalizations.

5. Experiments

In this section, we describe experiments on concept mapping between WordNet, CLICS and OmegaWiki.

5.1. Comparison Methods

We compare our translation-based methods to five methods from prior work.

1. **MG11** (Meyer and Gurevych, 2011) creates a sparse, interpretable embedding for each concept, based on the bag of words in its gloss. Each dimension in these embeddings corresponds to a word, and the value of each dimension is the frequency of that word in the gloss. Each source concept is mapped to its most similar target concept, according to the cosine similarity of their embeddings. We use our own re-implementation of the method.

2. **SBERT** (Reimers and Gurevych, 2019) is a method for generating dense sentence embeddings. We apply SBERT to the concept mapping task by substituting them for the sparse embeddings of MG11. We compute the similarity between concepts as the cosine similarity between the embeddings of their glosses using the code provided by the authors.[2]

   For both MG11 and SBERT, we expand the glosses with additional information from the resources. For WordNet, we follow Meyer and Gurevych (2011) in combining the gloss of the synset with its synonyms and hypernyms. For CLICS, we expand the gloss with its concept name.

3. **SemEq** (Yao et al., 2021) builds a bipartite graph where nodes represent concepts, and edge weights are computed with a cosine similarity measure on gloss embeddings. The best alignment is found by maximizing the total sum of the edge weights. We use the implementation made available by the authors.[3]

4. **SemAlign** (Pilehvar and Navigli, 2014) constructs two graphs using glosses and structural information from the two resources. The PageRank algorithm is applied to these graphs to compute a similarity measure between pairs of concepts. This method can map one concept to one other concept, or to multiple concepts, and can also leave a concept unaligned. Since the code for this method is not available, we report its results for the mapping between WordNet and OmegaWiki from the original paper.

5. **ESCHER** (Barba et al., 2021) is a state-of-the-art WSD system, which frames the WSD task as a text extraction problem. It concatenates the context of the target word with the glosses of the senses of the target word. The model is then trained to extract the span of text consisting of the gloss of the correct sense.

   The application of ESCHER to the resource mapping task via reduction to WSD is original to this paper. Given a concept to be aligned, we create an input context for ESCHER by concatenating the first word of the concept name with the verb “means” followed by its gloss. ESCHER then selects the best fitting WordNet sense of this target word, which implicitly identifies the output concept in either CLICS or OmegaWiki.

[2] https://huggingface.co/sentence-transformers/stsb-mpnet-base-v2
[3] https://github.com/tencent-ailab/EMNLP21_SemEq

Lang.	Overlap	ACC
EN	1881	0.799
EN & ID	3562	0.892
EN & NL	4069	0.869
EN & DE	3466	0.869
EN & RO	3899	0.860
EN & IT	3692	0.854
EN & GA	3546	0.851
EN & ES	3782	0.848
EN & PT	3693	0.843
EN & ZH	2705	0.843
EN & RU	2628	0.840
EN & FR	3538	0.837

Table 3: The mapping accuracy of the LANGVOTE method on the development set with different language sets, including the size of the lexical overlap between CLICS and BabelNet.

5.2. Evaluation Measures

As measures of the quality of concept mapping approaches, we report accuracy (ACC) and mean reciprocal rank (MRR). We designate one resource as the source, and the other as the target. Accuracy is simply the proportion of source concepts that are mapped to the correct target concept. MRR is calculated as follows: For each source concept, the target candidate concepts are ranked in order of their similarity values. The rank of the correct target concept in this ordering is identified, and its reciprocal is computed. The maximum reciprocal rank is therefore 1, and is attained if and only if the correct target concept is given the highest rank. If the correct target concept is not found in the ranking, the reciprocal rank is 0. The average of these reciprocal ranks over all source concepts yields the final MRR value.

5.3. Language Selection

The only tunable setting in our methods is the set of languages $\mathcal{L}$. As languages differ greatly in the quality and extent of their coverage in BabelNet, simply using all languages is suboptimal in terms of both running time and mapping accuracy. We establish the set of languages on our WordNet-CLICS development set, which we describe in Section 5.4.

Our language selection procedure is as follows: In addition to English (EN), we considered the languages with the highest lexicalization overlap between CLICS and BabelNet: Dutch (NL), Romanian (RO), Spanish (ES), Portuguese (PT), Italian (IT), Indonesian(ID), Irish (GA), French (FR), and German (DE). We also included Chinese (ZH) and Russian (RU), as these are typologically and orthographically different from the above languages.

First, we rank the 12 languages according to their performance on the development set when used in combination with only English. Table 3 shows the languages ordered by the accuracy of our LANGVOTE method, with ties broken randomly.

Next, we constructed the final set of languages. We started with $\mathcal{L} = \{\text{English}\}$, and added languages to $\mathcal{L}$ one by one, in order of the ranking established in the first step, until a decrease in accuracy was observed. In short, we applied a greedy strategy of adding languages in the order of the accuracy they produced on our development set, with English being included by default. This process yielded the set of seven languages English, Indonesian, Dutch, German, Romanian, Italian, and Irish, which we use in all experiments that follow.

Method	ACC	MRR
MG11	0.517	0.654
SBERT	0.591	0.713
SEMEQ	0.667	0.657
ESCHER	0.715	–
WORDVOTE	0.711	0.823
LANGVOTE	0.706	0.818

Table 4: Accuracy (ACC) and MRR for the alignment of WordNet and CLICS, on the test set.

5.4. Aligning WordNet and CLICS

Our principal concept-mapping dataset comes from Concepticon (List et al., 2016), which includes a hand-crafted mapping between a subset of CLICS concepts and WordNet. We extracted the dataset by following the procedure described by List (2018). The mapping contains 1372 one-to-one pairings of CLICS concepts and WordNet synsets. As our development set, we use 343 concept pairs that include usage examples. The remaining 1029 concept pairs constitute our test set.

The lexicalization information our methods depend on is readily available in CLICS and BabelNet (Section 2.1). As mentioned in Section 2.2, CLICS concepts are associated with categories, whereas WordNet concepts are marked with a part of speech. We map CLICS categories to parts of speech as follows: Action/Process: Verb; Number, Person/Thing, or Classifier: Noun; Property: Adjective or Adverb; Other: Noun, Adjective, or Adverb.

While our evaluation is limited to the 1029 WordNet synsets and the corresponding 1029 CLICS concepts which comprise our test set, the experiment involves *all* the synsets and concepts in these resources, that is, we map all synsets/concepts between WordNet and CLICS. Since CLICS has fewer concepts than WordNet has synsets, this means that each method that we apply attempts to align each CLICS concept with a single WordNet synset. However, in some cases, no alignment is found, due either to not sharing any lexicalizations with a concept in the other resource, or to the one-to-one constraint. 317 out of 2919 CLICS concepts are not mapped by our WORDVOTE method because of these issues.

The results of our experiment on the test set are presented in Table 4. Our two translation-based methods perform well above the three gloss-based comparison methods, but slightly below ESCHER. (MRR cannot be calculated for ESCHER, because it outputs only a single WordNet synset for each CLICS concept.) WORDVOTE achieves slightly better accuracy and MRR than LANGVOTE, which suggests that the total number of shared lexicalizations provides useful information in addition to the number of shared languages.

Error analysis reveals two main sources of error. First, a number of CLICS concepts appear to be duplicates and/or combine multiple concepts. This complicates the identification of a correct one-to-one mapping. For example, CLICS contains separate concepts named "STONE OR ROCK" and "STONE". Second, many translations are missing from the resources. For example, the CLICS concept "TO DRIP" has no lexicalizations in Indonesian, Dutch, German, or Romanian, while the BabelNet synset $drop_n^1$ contains no Dutch words. We conclude that most of the apparent errors are due to issues with the resources rather than flaws in our methods. More principled methods of defining concepts, and improvements in the multilingual coverage of lexical resources, would likely improve resource alignment results, in addition to yielding other benefits.

5.5. High-Precision Concept Alignment

To facilitate further work on this task, we release an automatically-generated, manually-validated resource consisting of a set of mappings from CLICS concepts to WordNet synsets. In this section, we describe how the resource was produced and validated.

We prioritize precision over coverage when creating the resource. To this end, we combine three strong systems: WORDVOTE, ESCHER, and SEMEQ, by independently applying them to the CLICS-WordNet concept mapping task. The combined method outputs an alignment only if all three methods agree; otherwise, no alignment is produced for that concept.

The application of this consensus approach to the set of 1547 CLICS concepts that are not found in the existing gold data produces 370 new alignment pairs. We manually inspected all of these concept pairs, and removed 6 of them as incorrect. Based on this, we estimate the precision of the consensus approach to exceed 95%.

In addition, we evaluated the accuracy of the consensus approach on the gold test set. In 64 test instances, the three methods agree with each other, but disagree with the gold mapping. However, in 49 of these 64 instances, we found that the mapping produced by the consensus approach is better than the one in the gold data. This suggest that the accuracy values reported in Table 4 may be underestimates.

5.6. Aligning WordNet and OmegaWiki

To validate the generality of our translation-based approach, we carry out an experiment on WordNet and OmegaWiki (Section 2.3). The gold data for this experiment was originally developed on the German part of OmegaWiki, which consists of German lexicalizations and concept glosses (Gurevych et al., 2012). Building upon this, Matuschek and Gurevych (2013) evaluate their mapping algorithm on this dataset directly, as each German OmegaWiki concept has at least one English lexicalization associated with it. For our evaluation, we use the version of the dataset provided by Pilehvar and Navigli (2014), who added additional English OmegaWiki candidates.

Method	ACC
MG11	0.840
SBERT	0.854
SEMEQ	0.853
SemAlign	0.893
ESCHER	0.695
WORDVOTE	0.894
LANGVOTE	0.879

Table 5: Accuracy for the alignment of WordNet and OmegaWiki.

The gold dataset is not a one-to-one mapping, but rather is composed of a set of binary mapping classifications on a number of pairs of concepts across the two resources. Any given concept may be aligned to zero, one, or multiple concepts in the other resource. For this reason, MRR is not well defined in this experiment. Following Pilehvar and Navigli (2014), we compute accuracy by dividing the number of correct binary classifications by the total number of concept pairs in the gold dataset (686).

The original dataset involves 315 WordNet synsets, but only 215 of them are aligned with OmegaWiki concepts. Since the dataset contains no lexicalizations of OmegaWiki concepts, we obtain them directly from OmegaWiki. Unfortunately, the version of OmegaWiki that served as the basis for this dataset is no longer available; we therefore use a more recent version (from 16 September 2021). Because of the dynamic nature of OmegaWiki, some concepts in the gold data are missing from the current version. We therefore restrict our evaluation to those OmegaWiki concepts that still have identical glosses as the current version. This yields a test set consisting of 276 WordNet synsets, of which 148 are aligned to OmegaWiki.

We used no part of this dataset in the development of our methods, made no changes to adapt them to this dataset, and attempted to keep our experimental setup as close as possible to the one described in Section 5.4.

Table 5 shows the results of the experiment. ESCHER underperforms on this dataset, because it often maps multiple OmegaWiki concepts to a single WordNet synset. MG11, SBERT, and SEMEQ all achieve accuracy values below those of our methods. Our WORDVOTE method performs comparably to SemAlign, This is remarkable considering that our approach is based exclusively on translation information, whereas SemAlign leverages both glosses and semantic relations between concepts. We conclude that this result provides evidence for the generality of our approach.

In our error analysis, we found three main types of errors. Most errors are due to missing lexicalizations in the resources. For example, the OmegaWiki concept VICTIM has no lexicalizations in Indonesian, Romanian, or Irish. Second, while our approach is designed to produce a one-to-one alignment, the data contains both one-to-many alignments and unaligned concepts. For example, the OmegaWiki concepts VICTIM and CASUALTY are both mapped to the WordNet concept with glossed as "an unfortunate person who suffers from some adverse circumstance." Third, due to the volatile nature of OmegaWiki, some concepts in the gold data are no longer in its current version. For example, for the WordNet synset $terminology^1_n$, three out of six candidates, including the correct one, are no longer available. Such errors are not caused by flaws in our methods; rather, they highlight the risk in using volatile online resources as a source of gold-standard data.

6. Conclusion

We presented two methods of leveraging translations for aligning concepts across lexical resources. They are based exclusively on multilingual lexicalization information, without any dependence on lexical relations, glosses, embeddings, or other sources of semantic or lexical knowledge. When tested on aligning two pairs of lexical resources, our methods match or exceed the accuracy of the best comparable methods from prior work. This demonstrates the utility of multilingual translation for resource mapping, while making the results easily interpretable. We provide a high-precision concept mapping between CLICS and WordNet concepts to facilitate comparison and further research.

7. Acknowledgements

We thank Robert Holte for comments on an earlier version of this paper.

This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Alberta Machine Intelligence Institute (Amii).

8. Bibliographical References

Barba, E., Pasini, T., and Navigli, R. (2021). ESC: Redesigning WSD with Extractive Sense Comprehension. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4661–4672.

Bond, F. and Foster, R. (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.

de Melo, G. and Weikum, G. (2010). Providing multilingual, multimodal answers to lexical database queries. In *Proceedings of the 7th Language Resources and Evaluation Conference (LREC 2010)*, pages 348–355, Paris, France. ELRA.

Gurevych, I., Eckle-Kohler, J., Hartmann, S., Matuschek, M., Meyer, C. M., and Wirth, C. (2012). UBY - a large-scale unified lexical-semantic resource based on LMF. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 580–590, Avignon, France.

Hauer, B. and Kondrak, G. (2020). Synonymy = translational equivalence. *arXiv preprint arXiv:2004.13886*.

Helou, M. A., Palmonari, M., and Jarrar, M. (2016). Effectiveness of automatic translations for cross-lingual ontology mapping. *J. Artif. Int. Res.*, 55(1):165–208.

List, J.-M., Cysouw, M., and Forkel, R. (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.

List, J.-M. (2018). Cooking with CLICS. *Computer-assisted language comparison in practice*, pages 14–18.

Matuschek, M. and Gurevych, I. (2013). Dijkstra-WSA: A graph-based approach to word sense alignment. *Transactions of the Association for Computational Linguistics*, 1:151–164.

Matuschek, M., Meyer, C. M., and Gurevych, I. (2018). Multilingual knowledge in aligned Wiktionary and Omegawiki for translation applications. In *Language technologies for a multilingual Europe*, pages 139–180.

McCrae, J. P. and Cillessen, D. (2021). Towards a linking between WordNet and Wikidata. In *Proceedings of the 11th Global Wordnet Conference*, pages 252–257, University of South Africa (UNISA).

Meyer, C. M. and Gurevych, I. (2011). What psycholinguists know about chemistry: Aligning Wiktionary and WordNet for increased domain coverage. In *Proceedings of 5th International Joint Conference on Natural Language Processing*, pages 883–892, Chiang Mai, Thailand.

Mihalcea, R. (2007). Using Wikipedia for automatic word sense disambiguation. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 196–203, Rochester, New York.

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

Navigli, R. and Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Pianta, E., Bentivogli, L., and Girardi, C. (2002). Multiwordnet: developing an aligned multilingual database. In *First international conference on global WordNet*, pages 293–302.

Pilehvar, M. T. and Navigli, R. (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, Baltimore, Maryland.

Ponzetto, S. P. and Navigli, R. (2010). Knowledge-rich word sense disambiguation rivaling supervised systems. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1522–1531, Uppsala, Sweden.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*.

Rzymski, C., Tresoldi, T., Greenhill, S. J., Wu, M.-S., Schweikhard, N. E., Koptjevskaja-Tamm, M., Gast, V., Bodt, T. A., Hantgan, A., Kaiping, G. A., et al. (2020). The database of cross-linguistic colexifications, reproducible analysis of cross-linguistic polysemies. *Scientific data*, 7(1):1–12.

Tjuka, A., Forkel, R., and List, J.-M. (2021). Linking norms, ratings, and relations of words and concepts across multiple language varieties. *Behavior Research Methods*, pages 1–21.

Yao, W., Pan, X., Jin, L., Chen, J., Yu, D., and Yu, D. (2021). Connect-the-dots: Bridging semantics between words and definitions via aligning word sense inventories. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7741–7751.