Towards a Comprehensive English Wordnet-Wikidata Mapping

John P. McCrae[1], Johann Bergh[2] , Krasimir Angelov[3]
[1] Research Ireland Insight and ADAPT Research Centre,
Data Science Institute, University of Galway, Galway, Ireland
[2] adorsys GmbH, Nuremberg, Germany
[3] Chalmers and University of Gothenburg, Sweden
john@mccr.ae, johann.bergh@adorsys.com , krasimir@chalmers.se

Abstract

In this study, we present a comprehensive investigation into the mapping of English Wordnet to Wikidata, focusing on the existing mappings created by different projects. We systematically analyze the current mapping methodologies and their effectiveness, highlighting the strengths and limitations of each approach. Through a comparative analysis, we identified overlaps and discrepancies among the mappings, revealing insights into the relationships between the data sets. Our findings underscore the need for a more unified dataset that consolidates disparate mappings into a comprehensive unified Wordnet-Wikidata mapping. We propose a novel construction methodology for this unified data set, taking advantage of existing mappings while addressing their shortcomings. In addition, we discuss future perspectives and advanced techniques for mapping the remaining unmapped records, such as machine learning algorithms. This work not only contributes to the enhancement of data interoperability between Wordnet and Wikidata but also sets the stage for future research aimed at refining mapping techniques and expanding coverage.

Keywords: Wordnet; Linking; Sense Alignment; Wikidata

1. Introduction

Lexical and encyclopedic resources play a fundamental role in enabling semantic interoperability across natural language processing (NLP) applications. Among the most widely used lexical resources is *Princeton WordNet* (Miller, 1995; Fellbaum, 2010), a large-scale lexical database that organizes English words into synsets and its more recent open source version Open English Wordnet (McCrae et al., 2019, 2020, OEWN)[1]. In parallel, *Wikidata* has emerged as a collaboratively curated, multilingual knowledge graph that captures a vast range of entities, properties, and relationships. Linking these two complementary resources would be helpful to a wide range of applications, such as word sense disambiguation, entity linking, semantic search, and knowledge-based reasoning.

Over the past decade, several projects have attempted to construct mappings between English Wordnet and Wikidata (or Wikipedia) using different strategies and resources, including Open English Wordnet, GF, BabelNet, and yovisto. These attempts have differences in their coverage and methodologies and in this paper, we try to compare the coverage and accuracy of these different resources and investigate whether a single linking of high coverage and accuracy can be created by consolidating these efforts into a unified, high-quality resource. We systematically compare the strengths and weaknesses of different mapping strategies, quantify overlaps and discrepancies, and examine common sources of alignment errors. Our analysis reveals that while coverage can be substantially increased by merging existing mappings, maintaining high accuracy requires careful treatment of conflicts, granularity mismatches, and conceptual ambiguities.

To this end, we propose a unified construction methodology that integrates existing mappings and leverages similarity measures and human-in-the-loop validation to improve quality. This work contributes not only a richer and more coherent mapping between English Wordnet and Wikidata but also insights into the nature of cross-resource alignment, informing future research on knowledge integration for multilingual and semantically aware NLP applications.

2. Related Work

Other works have considered the linking of English Wordnet with Wikidata or Wikipedia, one of the most notable being UBY (Gurevych et al., 2012), which integrates senses alignments between several resources, including between English Wordnet and Wikipedia, which was described in Niemann and Gurevych (2011). This work constructed a small testset, which has since been included in other resources, studied and was able to align senses with accuracies between 90.7%-94.5%. The same team also mapped Wiktionary to Wikipedia (Meyer and Gurevych, 2011) with 91.4% accuracy of the linking. A similar previous study (Toral et al., 2009) on English Wordnet and Wikidata showed only an accuracy of 77.1% on the

[1]We use *English Wordnet* to refer to both resources, "WordNet" is a trademark of Princeton University

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 11916–11925
11–16 May 2026.

task, underlining the difficulties in tackling this task.

The work of Ahmadi (2022) on sense alignment, including providing a large multilingual benchmark for linking between dictionaries as part of the ELEXIS project (Ahmadi et al., 2020). He also explored the use of b-Matching to improve the quality of linking between lexicographic resources (Ahmadi et al., 2019).

3. Datasets

We describe the production of the four main datasets of links between English Wordnet and Wikidata. We note that two of these (GF and BabelNet) were originally links to Wikipedia, which were then projected to Wikidata using the index provided by MediaWiki. This is generally quite reliable, as all Wikipedia articles were assigned Wikidata identifiers when Wikidata was created.

3.1. Open English Wordnet

Open English Wordnet (McCrae et al., 2019, 2020, OEWN) is an open-source fork of Princeton WordNet (Miller, 1995; Fellbaum, 2010, PWN) that aims to enhance and extend the English lexical database. Hosted on GitHub, OEWN has released annual updates since 2019, each available in multiple formats—including the original PWN format for drop-in compatibility, and the standard formats recommended by the Global WordNet Association (GWA) (McCrae et al., 2021, GWA). These updates address a wide range of improvements, from typo corrections to the addition and merging of synsets.

OEWN had included information from other resources, including language-specific wordnets (e.g., plWordNet (Maziarz et al., 2016)) to improve coverage and accuracy in English. It also introduces editorial guidelines on term inclusion, modern sense distinctions, and issues such as the representation of gendered language. OEWN is designed to evolve with contemporary English usage and lexicographic standards.

The Open English Wordnet mapping was created in two stages. The first stage was primarily conducted in 2020 and considered linking between *hapax legomenon* terms that occur in only a single resource. The second phase was completed in 2024 and focused on the proper nouns in OEWN.

McCrae and Cillessen (2021)[2] introduce hapax linking as an initial method for aligning concepts between OEWN and Wikidata. This approach targets *hapax legomenon*, i.e., terms that appear only once in both resources, on the assumption that such unique occurrences are more likely to refer to the same concept. Specifically, they identify wordnet synsets that are associated with only a single noun sense and attempt to match them to Wikidata entities that have the same lemma as a unique English label. This strategy reduces ambiguity and provides a high-precision basis for linking the two resources.

To ensure quality, the authors applied several heuristics to exclude spurious matches, such as filtering out Wikidata entities with very high Q-numbers (indicating low relevance), disambiguation pages, and formulaic titles like "X by Y" (common for songs or films). Using this method, they identified over 16,000 candidate links and conducted a manual evaluation, finding a high accuracy rate of 96.1%, with even higher precision (98.4%) when excluding certain error-prone categories like media titles. The hapax linking approach thus offers a reliable foundation for bridging wordnet's lexical semantics with Wikidata's encyclopedic knowledge.

The second stage was focused on the linking of proper nouns in OEWN, in particular those synsets that have an `instance_hypernym` link to another synset rather than the normal `hypernym`. This was conducted as a human-in-the-loop evaluation, where the top scoring candidate was presented to the annotator, who could accept it or suggest an alternative. The candidates were generated in a two-step process where candidates were first selected and then each candidate was scored.

The candidate selection was performed by creating a database of all labels in Wikidata, including all the aliases of each concept. As the labels are often multiword expressions, we treated them as sentences and used a sentence embedding to represent the whole label rather than a word embedding. In particular, the labels were encoded using a sentence transformer model (Reimers and Gurevych, 2019)[3] and then the dimensionality was further reduced by a singular-value decomposition[4]. These vectors were then stored in a Faiss database (Douze et al., 2025) and a candidate was retrieved if it had a cosine similarity > 0.98 to a lemma in OEWN. This high threshold was intended to capture small variations such as changes in case, accents (e.g., the city of "Cumaná" is missing the accent in OEWN) or wording differences (e.g., "Siege of Port Arthur" in Wikidata against "Battle of Port Arthur" in OEWN).

The second part used the similarity of the definition. Firstly, we created an augmented definition for both Wikidata and OEWN. For Wikidata, this augmented definition consisted of the label and all aliases[5] followed by a colon. We then added the

[2]Note this paper concerns the first phase described here.

[3]In particular, we used `all-MiniLM-L6-v2`

[4]In this case, from 384 to 28 dimensions. $384 \times 20 = 7680$ labels were sampled to calculate the SVD.

[5]This may be a source of error and for future evalua-

description in Wikidata and then in square brackets, some selected properties from the context of the entry. An example of an augmented definition is as follows:

> Mobile River: river in the United States of America [river, in United States of America, in Alabama]

The OEWN entries were augmented in a similar way, except that no properties were added. The similarity between the candidates, w, o, was then calculated according to the following formula,

$$\mathrm{sim}(w, o) = \cos(f(w), f(o)) - \alpha \log_{10}(i(w))$$

Where f is the embedding of the augmented definition by sentence transformers[6] and i gives the identifier in Wikidata as a number and $\alpha = 0.05$ was a hand-tuned value. The intuition of the penalty was that very high Wikidata identifiers are assigned to individuals of less note and so are unlikely to occur in wordnet. We then used a human-in-the-loop validation to check that all of these links were correct. The evaluation consisted of a single annotator verifying the highest scoring link per wordnet synset according to the similarity score defined above. The evaluation was conducted using a Google Sheet. The human-in-the-loop acceptance rates, as sorted by this score, are given in Figure 1 and we can see that the similarity score was mostly very accurate and that higher similarity scores correlated well with human acceptance. These links are available on the GitHub page of the resources[7] and have been added to Wikidata as links to the Collaborative InterLingual Index (CILI) using the Wikidata property P5063.

3.2. GF

Grammatical Framework (GF) (Ranta, 2011) is a programming language for describing natural languages. It comes with a library of syntactic constructions (Ranta, 2009) for about 50 languages. This is complemented with a wordnet-based lexicon for 265 languages, where the lexicon for 48 of those is integrated with the syntactic library (Angelov (2020), Angelov (2025)).

One of the resources used for the development of the lexicon is Wikipedia. During the initial creation of the English-Swedish-Bulgarian part of the lexicon (Angelov, 2020), every lexical item was matched with the Wikipedia article whose titles in each language match as many of the existing translations as possible. Matches were then manually checked and used to extend the lexicon with more translations and more languages.

Later on, in Angelov (2025) the mapping to Wikipedia is projected to Wikidata. Wikidata has the property P8814 which refers to the synset ID in Princeton WordNet 3.1. Using the property, the community has made its own linking for some of the Wikidata entities. GF's own mapping was merged with this one. In the process of merging, mistakes on both sides were detected and fixed. Overall, this doubled the number of existing links.

Wikidata is one of the key resources used for extending GF Wordnet with more languages, which shows one of the potential applications of better mapping.

3.3. BabelNet

BabelNet[8] (Navigli and Ponzetto, 2010; Navigli et al., 2021) is a multilingual semantic network and knowledge base that integrates various lexical resources, such as English Wordnet and Wikipedia, to provide a comprehensive representation of concepts and their relationships across different languages.

A key component of this resource is the linking between Wikipedia and English Wordnet. The methodology to construct this (Ponzetto and Navigli, 2010) involves automatically mapping Wikipedia pages to corresponding English Wordnet senses through a probabilistic alignment algorithm; the Wikidata linking was then constructed from links between Wikipedia pages and Wikidata entities. For each Wikipedia page, the system considers its lemma and attempts to link it to the most appropriate English Wordnet sense based on contextual evidence. The disambiguation context for a Wikipedia page includes sense labels from its title, outgoing links to other pages, and category information, while the context for an English Wordnet sense incorporates synonyms, hypernyms/hyponyms, sister terms, and gloss definitions. The mapping algorithm computes the joint probability between a Wikipedia page and candidate English Wordnet senses by measuring the overlap between their respective disambiguation contexts, ultimately selecting the sense that maximises this probability.

The resulting resource forms the English component of BabelNet, which is further extended to multiple languages using Wikipedia inter-language links and machine translation. This approach enables the integration of encyclopedic knowledge from Wikipedia with the lexical-semantic structure of English Wordnet, creating a unified, multilingual semantic network that supports a wide range

tions, we would limit to 10 aliases

[6] Using `all-MiniLM-L6-v2` as before, but without SVD decomposition

[7] https://github.com/globalwordnet/english-wordnet

[8] https://babelnet.org/

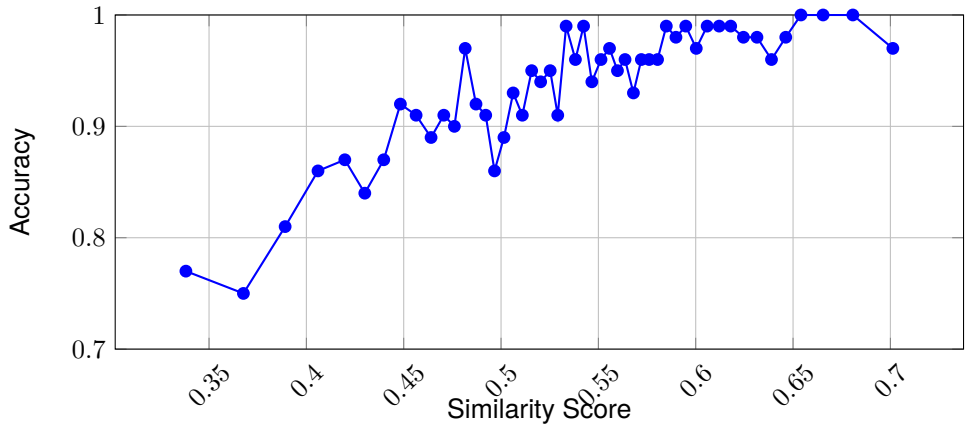

Figure 1: Proportion of accepted values by similarity scores. Average score grouped by 100 sorted elements.

of applications, including word sense disambiguation, semantic relatedness computation, and cross-lingual information access.

3.4. Yovisto

The Yovisto dataset was constructed using two annotators, DBpedia-Spotlight[9] and yovisto-KEA[10]. The two annotators identify DBpedia URIs within a given body of text. There is an existing mapping between the DBpedia URIs and Wikidata URIs, which makes it trivial to obtain the corresponding Wikidata URIs for the DBpedia URIs identified in the text. Our aim is to use these two annotators to map English Wordnet synsets to Wikidata URIs. We are faced with a challenge here because the annotators depend on context to identify the correct corresponding DBpedia entities. The only real direct textual information we can use here is the group of lemmas that are linked to the synset. The lemmas, however, are only single words, and do not provide enough context for the annotators to make an accurate selection (e.g., depending on the context, the noun 'rocket' can mean 'any vehicle self-propelled by a rocket engine' or 'herbal plant often used in salads'). Similarly, Siegel and Bergh (2023) were also confronted with this issue, but in a different context. They used machine translations from English to German to find the corresponding lemma for the Open German WordNet (OdeNet)[11], so that they could assign the correct Interlingual Index (ILI) to the German synset. By attaching the definition of the synset to the lemmas associated with the synset and then doing the machine translation, they were able to greatly improve the results. For example, the noun 'washer' in English can mean 'someone who washes things' or 'a seal used to prevent leaks'. The following two phrases were passed through the machine translation, which, contrary to previous attempts, rendered the correct results:

- `washer: someone who washes things for a living`
  → `Wäscher`
- `washer: seal consisting of a flat disk placed to prevent leakage`
  → `Unterlegscheibe`

The same method could be used to provide the annotators with more context so that an accurate selection can be made. For example, for the noun 'rocket' referring to a vehicle self-propelled by a rocket engine, we could run the annotators over the following phrase:

- `rocket: any vehicle self-propelled by a rocket engine`

We only look at the result that the annotators returned for the first word 'rocket', and discard the rest. This process can now be repeated for each one of the lemmas in the synset ('projectile', etc.). The results for this example are depicted in Table 1 for the DBpedia-Spotlight annotator. DBpedia-Spotlight returns a URI and confidence score for each lemma, and we then take the result with the highest confidence score. In this case, we would select Projectile as the most correct match for our synset. This process was repeated for each of the Noun-Synsets in the English Wordnet, for both DBpedia-Spotlight and yovisto-KEA. For DBpedia-Spotlight, 59,735 mappings were found using this method, and for yovisto-KEA, 67,438 mappings. This looks like quite a high recall when considering that the English Wordnet has about 83000 nouns in total. Upon closer inspection, however, it became clear that there were many duplicate mappings,

[9] https://www.dbpedia-spotlight.org/
[10] https://github.com/yovisto/kea-el
[11] https://github.com/hdaSprachtechnologie/odenet

ILI	Lemma	Result	Confidence
i58225	rocket	http://dbpedia.org/resource/Rocket_engine	0.59
i58225	projectile	http://dbpedia.org/resource/Projectile	0.99

Table 1: DBpedia-Spotlight results for 'rocket' and 'projectile' meaning 'self-propelled vehicle'

Annotator	Raw Mappings	Filtered Mapppings
DBPedia-Spotlight	59735	36530
yovisto-KEA	67438	34339

Table 2: Annotator Mappings out of a possible 83000 English Wordnet Nouns

meaning that the annotators mapped many synsets in English Wordnet to the same Wikidata URI.

For the Wordnet-Wikidata mapping, it is important to have a strict one-to-one mapping, i.e., one English Wordnet synset can only map to one Wikidata URI and vice versa. In order to enforce this, the mappings for both annotators were ordered by confidence score, and in case of duplicate mappings, the mapping with the highest confidence score was taken, and the other mapping(s) were discarded. This left us with 36,530 mappings for DBpedia-Spotlight and 34,339 mappings for yovisto-KEA, as depicted in Table 2.

Although all the duplicates have now been removed, we were still confronted with the fact that many of the mappings were not accurate. When compared to one of the existing hand-curated datasets as described in the work of McCrae and Cillessen (2021), the DBPedia-Spotlight and yovisto-KEA mappings only achieved a direct-match accuracy of 80%-85%, leaving considerable room for improvement. As a consequent step, we took the intersection of the two datasets, i.e., all the mappings for DBPedia-Spotlight and yovisto-KEA, where both of these datasets have the same mapping for the English Wordnet synset to the Wikidata URI. Because DBpedia-Spotlight and yovisto-KEA use two completely different algorithms to compute Wikidata URIs in a body of text, the theory was that the accuracy of the mappings would increase when confirmed by two diverse methods. This resulting dataset had only 13364 mappings. Although the number of mappings decreased significantly, the accuracy also increased markedly to 96%.

4. Analysis

We use UpSet plots to illustrate the comparison in the size of the existing Wikidata-Wordnet linkages. In an UpSet plot, the set intersections are visualized as column charts and the sets being intersected are shown by connected dots along the columns. In Figure 2, we compute the size of the resources, with BabelNet being the largest resource at 39,224 links and OEWN being the smallest with only 12,083 links. We also compute the size of the pair-wise intersections of the English Wordnet synsets covered between the resources and we see that for the most part, the different linkings are quite complementary, with at most 52.0% of the links being shared between any pair of resources. The complete intersection of all resources is only 3,017 synsets, while 49,219 synsets occur in at least one linking (60.0% of noun synsets), showing that there is still a lot of scope for improvement in the coverage of a single gold-standard linking.

In Figure 3, we consider how often the resources specify a different Wikidata target for the same synset in English Wordnet as a percentage of the total size of the intersection. In this case, we see that yovisto and OEWN have the strongest agreement but also the smallest intersection size. BabelNet and GF disagree the most with a total of 2,923/20,384 (=14.3%) disagreements.

In order to measure the true accuracy of the mappings, we conducted a manual analysis of all the links that disagree between OEWN and the other three resources. We selected this resource as it is small and constructed manually. The results are shown in Table 3, where the results are divided into four cases: either OEWN was correct; the other resource was correct; both Wikidata links were different but correct; or both resources were incorrect and another Wikidata entity is the most appropriate. The results clearly confirm that OEWN is the most accurate resource; however, there are still a substantial number of errors in this resource. In Table 4, we illustrate some examples of the kinds of errors made by the resource. Firstly, we see that BabelNet has chosen an overly specific entity for 'Mandarin Chinese', rather than the correct entry for the language, similarly, we see the OEWN chose a Wikidata entry with the same lemma as the original however, an analysis of the definition shows that it should refer to an air-to-air attack aircraft, which BabelNet gets correct. A common issue was due

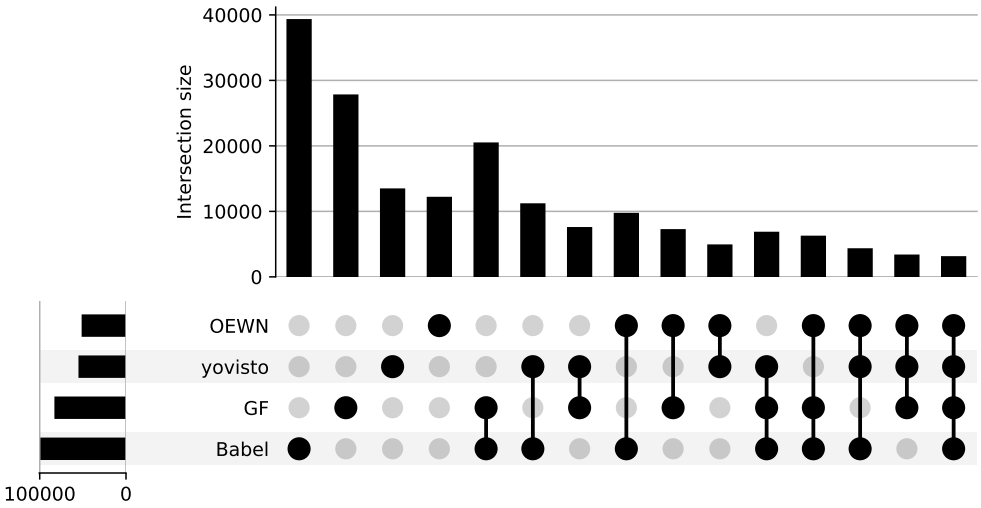

Figure 2: Comparative size of the resources

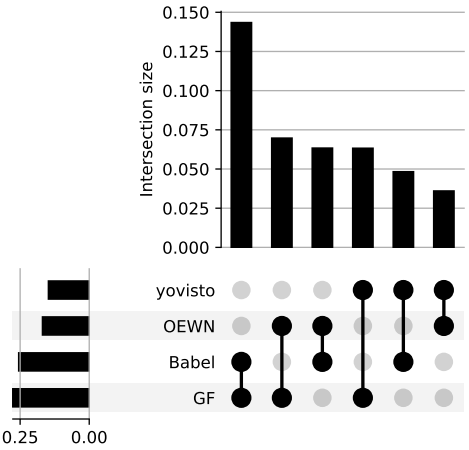

Figure 3: The percentage of disagreements in the value of links between different sources

to subtle changes in the type of the entity; English Wordnet and Wikidata both distinguish between the geographical concept of 'Marshall Islands' and the country named after the islands and BabelNet makes a mistake in this linking. In many cases, multiple links may be correct, for example, in English Wordnet, 'air pump' and 'vacuum pump' are in the same synset, while they are distinct concepts in Wikidata. Another common reason for this error was due to concepts that have been (or should be) merged in either resource, leading to different IDs redirecting to the same concept. Finally, in a small number of cases, neither resource was deemed correct, for example, both OEWN and BabelNet chose unusual mappings for the concept of 'railway junction', while a more obviously correct entity was available[12].

5. Discussion

This analysis reveals that the existing mappings from the OEWN, BabelNet, GF and yovisto are highly complementary mappings rather than overlapping, with only 3,017 synsets being covered by all four resources, while sixty percent of all nouns are covered by at least one mapping. This suggests that a unified approach would be required to consolidate these efforts and maximize the coverage. We also note that it is unlikely that all noun senses could be mapped to Wikidata, as the nature of the resources is quite different. For example, 'play' has 17 senses in Open English Wordnet, while Wikidata has 8 concepts labelled with 'play'[13] and many of these synsets represents senses that would be too abstract to appear in Wikidata, such as "in play" or "the play of the lights on the water."

The different resources have made different trade-offs between recall and precision, with OEWN using a high-precision methodology with human-in-the-loop validation, resulting in high accuracy

[12]railway junction (*place at which two or more rail routes converge or diverge*)

[13]A further 91 concepts are labelled as 'Play', however these are mostly proper nouns such as the 1999 Album by Moby and English Wordnet has no proper noun with this lemma

	OEWN	Other	Both	Neither	N
GF	59.4%	19.1%	18.7%	2.8%	498
BabelNet	64.4%	18.7%	15.6%	1.3%	611
yovisto	62.8%	19.2%	16.3%	1.7%	171

Table 3: Evaluation of disagreements between OEWN and the other three resources

Wordnet synset	OEWN Wikidata Target	BabelNet Wikidata Target	Label
Mandarin Chinese (*the dialect of Chinese spoken in Beijing...*)	Mandarin Chinese, Guanhua (*major branch of Chinese...*)	Beijing Mandarin dialect (*dialect of Beijing Mandarin...*)	OEWN
attack aircraft, fighter (*a [...] airplane designed to destroy enemy aircraft in the air*)	attack aircraft (*aircraft class designed for air-to-surface warfare*)	combat aircraft (*aircraft class designed for aerial warfare*)	BabelNet
Marshall Islands (*a group of coral islands in eastern Micronesia*)	Marshall Islands (*group of islands in the North Pacific Ocean*)	Marshall Islands (*country near the equator in the Pacific Ocean*)	OEWN
air pump, vacuum pump (*a pump that moves air in or out of something*)	air pump (*device for pushing air*)	vacuum pump (*equipment producing low air pressure*)	Both
railway junction (*a junction where two or more railway lines meet or cross*)	switches and crossings (*railroad elements*)	flat crossing, diamond crossing (*crossing of two rail tracks without grade separation...*)	Neither

Table 4: Examples of mapping errors in OEWN and BabelNet

Resources	Excl. Size	Excl. Accuracy	Est. Accuracy
BabelNet	10,298	95%	97.3%
GF	3,874	91%	96.9%
yovisto	1,026	85%	97.1%
Multiple	14,121	99%	-

Table 5: Comparison of size and accuracy of mappings exclusively suggested by one resource and by multiple resources.

but limited coverage. However, manual checking has not proven by itself to be a guarantee and the automatic methods are close in accuracy, with the automatically created resource GF being of similar quality to the automatically created BabelNet and yovisto datasets. Our results may understate the accuracy of BabelNet and GF as these resources frequently gave secondary links, which appeared to be more error-prone than the primary links, such as the Mandarin example in Table 4. In fact, the assumption that this linking is a strict one-to-one conceptual mapping has been disproved by the fact that the annotators often accepted both links.

Further challenges in the linking are due to granularity mismatches, such as the difference between 'air pump' and 'vacuum pump' and conceptual drift of entities over time. However, we note that the mappings are generally very accurate, with errors often being small *ontological* issues such as referring to 'circus' as a group of performers, not the performance made by this group.

In order to produce a consolidated mapping, it is necessary to merge the mappings and in Table 5, we consider the size and the accuracy of mappings that could be integrated. We consider the *exclusive* accuracy of links proposed only by this resource, which were estimated by a sample of 100 mappings from each of the resources, as well as a sample of 100 suggested by one or more resource. We also then attempt to estimate the resource's accuracy based on the weighted average of the resource-exclusive accuracy and multiple accuracy. We do not consider OEWN for this analysis as all the disagreements were already handled in Section 4.

We notice that a very large number of mappings are supported by multiple resources with about 14,000 links and that these links are of very high accuracy, however there are some mistakes that have still been made in multiple resources, such as the concept for 'aleph' which was mapped by both BabelNet and GF to the concept Q15033, which refers to the first letter of many Semitic alphabets rather than Q16725860, the first letter of the *Hebrew* alphabet. As such, human linking is still very

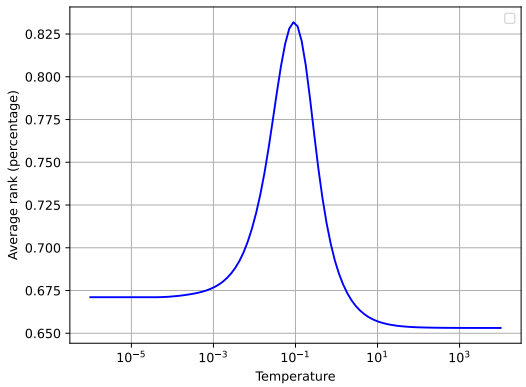

Figure 4: Average Rank of Rejected Mappings Reranked by Softmax with Temperature

important to detect these errors and the development of algorithms that better detect when there is a conflict between the potential candidates.

Further, the links are recorded in Wikidata using two different properties, P5063, which use the CILI ID and P8814, which targets the Princeton WordNet 3.1 synset offset identifier, and these two links sometimes display contradictory information. We also note that both OEWN and Wikidata are open resources, so anyone can contribute new links to either resource and new concepts can be added to cover gaps in either resource[14].

We note that in this work, we have taken a strict one-to-one approach with this mapping, which assumes that each English Wordnet synset can be uniquely mapped to a single entity in Wikidata. However, we note that this is not always the case and more complex mapping could be helpful. Firstly, it is often the case that the wordnet concept will have a different granularity compared to that in Wikidata, for example, the concept of 'air pump' and 'vacuum pump' are synonyms in wordnet, but have distinct concepts in Wikidata. In this case, many-to-many links using more properties, such as hypernym, would be useful. However, this represents only a very small amount of the links. The OEWN release contains a number of multiple links (6.03%), although it should be noted that these are mostly links to deprecated IDs in Wikidata. Further, recording the provenance of the links would be helpful and this is being made available in the Wikidata references.

6. Towards an Integrated Resource

In order to create a single integrated resource with high accuracy, it is necessary to combine all these resources using a human-in-the-loop validation procedure. We can see that links that occur in multiple resources are of very high quality, and so these will be the first priority in terms of manual validation, followed by the individual resources. In addition, we rank the results by the similarity formula used by OEWN; however, we have since improved this formula to the following:

$$\begin{aligned}\mathrm{sim}'(w,o) =& \mathrm{sim}(w,o)\times \\ & \max_{w'\in W_o}(\mathrm{softmax}_\tau(sim(w',o)))\end{aligned}$$

Where W_o are the candidate mappings in Wikidata and $\tau = 0.1$ is a temperature parameter. This measure calculates the uncertainty of a mapping by down-weighting scores if there are two or more close candidates. We measured the average rank of examples that were rejected during the evaluation of instance hypernym mappings described in Section 3.1. We considered the position where the rejected examples occurred in the mapping and this is shown in Figure 4. This rank is defined as:

$$\rho = \frac{\sum_{i\in I}\mathrm{rank}(i)}{N\times|I|}$$

Where N is the total number of linking suggestions, I is the set of rejected mappings, and $rank$ gives the rank between 1 and N. This method was able to reduce the average occurrence of a rejected example for about two-thirds down the list to the last fifth of the data and close to the optimal rank of 96.6%, where all rejected examples are suggested after the accepted examples. This metric will drive our manual review, which will lead to a single high-quality and high-coverage linking and currently over 15,000 links have been reviewed.

7. Conclusion

In this paper, we have presented a comprehensive investigation into existing mappings between English Wordnet and Wikidata, examining four major resources, namely Open English WordNet, GF, BabelNet, and yovisto. Our comparative analysis reveals that these mappings are largely complementary rather than overlapping, with only a small subset of synsets being linked consistently across all resources. While the current mappings together cover around sixty percent of noun synsets, there is substantial variation in accuracy, granularity, and conceptual alignment.

We have shown that combining multiple mappings can improve coverage and precision, as links supported by multiple resources demonstrate notably higher accuracy. However, the consolidation

[14]This is in fact something that we are doing

of these mappings also exposes a number of challenges, including mismatches in conceptual granularity, duplicate or conflicting links, and subtle semantic divergences. We are working towards a unified Wordnet–Wikidata mapping that builds on the strengths of existing approaches while resolving their inconsistencies.

8. Limitations

While this study provides a detailed comparative analysis of existing Wordnet–Wikidata mappings and proposes a unified integration methodology, several limitations remain. First, the evaluation was limited to noun synsets, excluding verbs, adjectives, and adverbs, which may exhibit distinct patterns of ambiguity and alignment difficulty. It is not clear what a suitable target for parts of speech other than nouns might be. Another limitation concerns the dynamic nature of Wikidata, which is continuously updated by a global community. Changes in entity definitions, merges, and deletions may lead to inconsistencies or link rot in the mapping over time. Developing mechanisms for continuous synchronization between Wordnet and Wikidata will be essential to maintain the long-term stability of the resource. Finally, the scope of this study did not include multilingual evaluation beyond English, although Wikidata is an inherently multilingual resource and wordnets are available for many languages. In fact, similarly to GF, these links could provide important tools for increasing the coverage of wordnets in other languages by exploiting translations in Wikidata.

Acknowledgements

John P. McCrae is supported by Research Ireland under Grant Number SFI/12/RC/2289_P2 Insight_2, Insight SFI Centre for Data Analytics and Grant Number 13/RC/2106_P2, ADAPT SFI Research Centre.

Bibliographical References

Sina Ahmadi. 2022. *Monolingual alignment of word senses and definitions in lexicographical resources*. Ph.D. thesis, University of Galway.

Sina Ahmadi, Mihael Arcan, and John P. McCrae. 2019. Lexical sense alignment using weighted bipartite b-matching. In *Proceedings of the Poster Session of the 2nd Conference on Language, Data and Knowledge (LDK 2019), Leipzig, Germany, May 21, 2019*, volume 2402 of *CEUR Workshop Proceedings*, pages 12–16. CEUR-WS.org.

Sina Ahmadi, John Philip McCrae, Sanni Nimb, Anas Fahad Khan, Monica Monachini, Bolette S. Pedersen, Thierry Declerck, Tanja Wissik, Andrea Bellandi, Irene Pisani, Thomas Troelsgård, Sussi Olsen, Simon Krek, Veronika Lipp, Tamás Váradi, László Simon, András Gyorffy, Carole Tiberius, Tanneke Schoonheim, Yifat Ben Moshe, Maya Rudich, Raya Abu Ahmad, Dorielle Lonke, Kira Kovalenko, Margit Langemets, Jelena Kallas, Oksana Dereza, Theodorus Fransen, David Cillessen, David Lindemann, Mikel Alonso, Ana Salgado, José-Luis Sancho, Rafael-J. Ureña-Ruiz, Jordi Porta-Zamorano, Kiril Simov, Petya Osenova, Zara Kancheva, Ivaylo Radev, Ranka Stankovic, Andrej Perdih, and Dejan Gabrovsek. 2020. A multilingual evaluation dataset for monolingual word sense alignment. In *Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020*, pages 3232–3242. European Language Resources Association.

Krasimir Angelov. 2020. A parallel WordNet for English, Swedish and Bulgarian. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3008–3015, Marseille, France. European Language Resources Association.

Krasimir Angelov. 2025. An Abstract Multilingual WordNet. In *Global WordNet Conference 2025*.

Matthijs Douze, Alexandr Guzhva, Chengqi Deng, Jeff Johnson, Gergely Szilvasy, Pierre-Emmanuel Mazaré, Maria Lomeli, Lucas Hosseini, and Hervé Jégou. 2025. The Faiss library.

Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, and Christian Wirth. 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. Association for Computational Linguistics.

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

John P. McCrae, Michael Wayne Goodman, Francis Bond, Alexandre Rademaker, Ewa Rudnicka, and Luis Morgado Da Costa. 2021. The GlobalWordNet formats: Updates for 2020. In *Proceedings of the 11th Global Wordnet Conference*,

pages 91–99, University of South Africa (UNISA). Global Wordnet Association.

Christian M. Meyer and Iryna Gurevych. 2011. What psycholinguists know about chemistry: Aligning wiktionary and wordnet for increased domain coverage. In *Fifth International Joint Conference on Natural Language Processing, IJCNLP 2011, Chiang Mai, Thailand, November 8-13, 2011*, pages 883–892. The Association for Computer Linguistics.

Elisabeth Niemann and Iryna Gurevych. 2011. The people's web meets linguistic knowledge: Automatic sense alignment of Wikipedia and WordNet. In *Proceedings of the Ninth International Conference on Computational Semantics (IWCS 2011)*.

Simone Paolo Ponzetto and Roberto Navigli. 2010. Knowledge-rich word sense disambiguation rivaling supervised systems. In *ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden*, pages 1522–1531. The Association for Computer Linguistics.

Aarne Ranta. 2009. The GF resource grammar library. *Linguistic Issues in Language Technology*.

Aarne Ranta. 2011. *Grammatical Framework: Programming with Multilingual Grammars*. CSLI Publications, Stanford. ISBN-10: 1-57586-626-9 (Paper), 1-57586-627-7 (Cloth).

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. 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 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Melanie Siegel and Johann Bergh. 2023. Connecting multilingual wordnets: Strategies for improving ILI classification in OdeNet. In *Proceedings of the 12th Global Wordnet Conference*, pages 363–368, University of the Basque Country, Donostia - San Sebastian, Basque Country. Global Wordnet Association.

Antonio Toral, Óscar Ferrández, Eneko Agirre, and Rafael Muñoz. 2009. A study on linking wikipedia categories to wordnet synsets using text similarity. In *Recent Advances in Natural Language Processing, RANLP 2009, 14-16 September, 2009, Borovets, Bulgaria*, pages 449–454. RANLP 2009 Organising Committee / ACL.

9. Language Resource References

Christiane Fellbaum. 2010. *WordNet*, pages 231–243. Springer Netherlands, Dordrecht.

Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, Stan Szpakowicz, and Paweł Kędzia. 2016. plWordNet 3.0 – a comprehensive lexical-semantic resource. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2259–2268, Osaka, Japan. The COLING 2016 Organizing Committee.

John P. McCrae, Alexandre Rademaker, Francis Bond, Ewa Rudnicka, and Christiane Fellbaum. 2019. English WordNet 2019 – an open-source WordNet for English. In *Proceedings of the 10th Global Wordnet Conference*, pages 245–252, Wroclaw, Poland. Global Wordnet Association.

John Philip McCrae, Alexandre Rademaker, Ewa Rudnicka, and Francis Bond. 2020. English WordNet 2020: Improving and extending a WordNet for English using an open-source methodology. In *Proceedings of the LREC 2020 Workshop on Multimodal Wordnets (MMW2020)*, pages 14–19, Marseille, France. The European Language Resources Association (ELRA).

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

Roberto Navigli, Michele Bevilacqua, Simone Conia, Dario Montagnini, and Francesco Cecconi. 2021. Ten years of babelnet: A survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 4559–4567. ijcai.org.

Roberto Navigli and Simone Paolo Ponzetto. 2010. Babelnet: Building a very large multilingual semantic network. In *ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden*, pages 216–225. The Association for Computer Linguistics.