Leveraging LLMs to Automatically Construct WordNets as Bilingual Resources

Johann Bergh
yovisto GmbH
Potsdam, Germany
johann@yovisto.com

Jörg Waitelonis
yovisto GmbH
Potsdam, Germany
joerg@yovisto.com

Melanie Siegel
Darmstadt University
of Applied Science
melanie.siegel@h-da.de

Abstract

The Princeton WordNet (OEWN[1]) is widely recognized as the gold standard for both the quality and quantity of synsets it contains, with over 120,000 manually curated entries, making it one of the most comprehensive WordNets available. In contrast, non-English WordNets within the Open Multilingual WordNet (OMW) project contain significantly fewer synsets, highlighting a substantial disparity. This gap poses challenges for researchers, particularly when working in multilingual settings or with under-resourced languages, where WordNets are often unavailable. This shortfall also limits the applicability of WordNets in multilingual and minority language contexts. In this paper, we propose automated methods to construct high-quality WordNets to bridge these gaps. Specifically, we introduce a method leveraging large language models (LLMs) to generate missing lemmas. Our approach was evaluated using a manually compiled dataset, demonstrating its potential to address the synset shortfall in non-English and low-resource languages.

1 Introduction

OEWN is a hand-curated WordNet that contains more than 120,000 synsets. Most other WordNets, especially those in the OMW project, contain far fewer synsets, and some of them were constructed by means of automation, or they were only partly hand-curated. Table 1[2] gives a brief summary of the number of synsets for OEWN and some of the other important European languages. Most of them have fewer than a third of the number of synsets that OEWN has. This poses a problem not only for researchers who want to work with these languages, but also for those who are interested in using WordNets as bilingual resources. Automated methods for addressing this problem have thus-far proved to be inadequate, mainly due to stumbling blocks encountered when using machine translations. Especially for languages with fewer resources, the effort of manually creating a WordNet is often too high and the need for automatic methods is urgent.

WordNet	Language	No. Synsets
OEWN	English	120,135
WOLF	French	59,091
OpenWN-PT	Portuguese	43,895
Multil. Central Repos.	Spanish	38,512
OdeNet	German	36,268
MultiWordNet	Italian	35,001
Open Dutch WordNet	Dutch	30,177

Table 1: Details of European WordNets in OMW

The Interlingual Index (ILI) is an identifier (Bond et al., 2016) that was introduced to the OMW project to enable a synset from one language to be linked to an equivalent synset of another language. It is very important for the usage of WordNets in bilingual contexts. We have tested various ways of automatically creating new and hybrid WordNets and propose a method that allows WordNets for new languages to be created with help of Large Language Models (LLMs) with reasonable quality within 2–3 days.

The advent of LLMs and Prompt Engineering has brought many advancements in solving lexical semantic tasks. WordNets are still relevant, however, and complement LLMs in several ways. With its explicit structure and fine-grained lexical relations, WordNets can complement LLMs by offering a precise semantic network that can help models reason about language. The explainability and transparency of WordNets, which allows for the referencing of explicit semantic relationships and traceable logic, provide an advantage over the black-box nature of LLMs. Researchers have started to use LLM technology for the construction and expansion of WordNets. For example, (Wojtasik et al., 2023) generated definitions for synsets in the Polish WordNet with an LLM. (Oliveira, 2023) have added synonym, antonym, hypernym, and hyponym relations to the Portuguese WordNet prompting BERT models.

In this paper, we address the current challenges in the automatic construction of WordNets. In addition to discussing machine-translated and hybrid WordNets, we

[1]The Open English WordNet (OEWN) is a copy of the Princeton WordNet and forms part of the Open Multilingual WordNet Project. When referring to OEWN in this article, it can also be interpreted as the Princeton WordNet.

[2]Taken from `https://github.com/goodmami/wn/tree/v0.9.5` on 2024-10-11

introduce several methods in LLMs for automating the creation of high-quality WordNets. We demonstrate that the 'LLM as a judge' technique achieves a 93% success rate in predicting lemmas, highlighting its effectiveness in improving the accuracy and efficiency of WordNet construction.

The paper is structured as follows: Section 2 introduces and discusses the current challenges in automatic WordNet construction, along with related work in the field. Sections 3 and 4 elaborate on machine translation and hybrid approaches. In Section 5, we present the automated WordNet creation using large language models (LLMs). Section 6 provides a comprehensive evaluation of the proposed method, where the results generated by the LLM are compared against a ground truth.

2 WordNet Automation, the Current State

Vossen (1998) outlines two approaches for creating WordNets:

- The *merge* approach: all senses for an applicable word is compiled from scratch, with synsets that then contain all the words for a given sense.
- The *expansion* approach: synsets from an existing WordNet are used to create equivalent synsets in another WordNet.

Neale (2018) gives an outline of WordNet construction methods and best practices. The *merge* or *expansion* methods can be used independently to create WordNets, or they can be used in combination. In general, the *merge* approach is used to create higher quality WordNets, with more manual intervention, and takes longer. The expansion approach seems to be more suitable to create automated WordNets in a shorter time span, albeit at a lower quality. This approach also often uses OEWN as a base WordNet since it is of high quality, and with its more than 120,000 synsets also one of the most complete WordNets available. With the advent of the ILI, it also makes sense to work with OEWN as a base, so that the benefit of connecting equivalent synsets in different languages can be realized easily.

Machine translation plays a central role in creating automated WordNets quickly, but comes with some stumbling blocks, as reported by Siegel and Bergh (2023). Most notably, problems arise with machine translations when translating homographs (words with similar spelling but different meanings) and polysemes (words with similar spelling and closely related meanings). For example, the words 'washer' and 'bank' in English can have different meanings depending on the context, and if these words are converted into another language with machine translation, the result will not necessarily be correct, because of the missing context. Siegel and Bergh (2023) proposed a method for getting better machine translations, by providing additional context. All synsets in OEWN have a short, concise definition, and this can be used to provide the additional context for improving machine translations. This might be achieved by concatenating the synset lemma and definition before doing the machine translation. The translation of the word 'washer' from English to German is shown below:

- EWN ID: `ewn-10788571-n`
  ILI: `i94042`
  Lemma-Definition combination:
  `washer: someone who washes things for a living`
  Machine translation:
  `Wäscher: jemand, der beruflich Dinge wäscht`

- EWN ID: `ewn-04562157-n`
  ILI: `i60971`
  Lemma-Definition combination:
  `washer: seal consisting of a flat disk placed to prevent leakage`
  Machine translation:
  `Unterlegscheibe: Dichtung, die aus einer flachen Scheibe besteht, um ein Auslaufen zu verhindern`

- EWN ID: `ewn-04561970-n`
  ILI: `i60970`
  Lemma-Definition combination:
  `washer: a home appliance for washing clothes and linens automatically`
  Machine translation:
  `Waschmaschine: ein Haushaltsgerät zum automatischen Waschen von Kleidung und Wäsche`

Though the original intention with this method was to find the most suitable ILI from OEWN for OdeNet (Siegel and Bond, 2021), the German WordNet from OMW, it can also be used to create high-quality WordNets.

As an example, the Japanese WordNet (Kaji and Watanabe, 2006)(Isahara et al., 2008) was created through a non-context-aware translation of the synset lemmas from the English WordNet, and thereafter a word-sense disambiguation method was applied to resolve possible false translations due to the missing context. In these publications an explicit measurement of the success was not given, and in a consequent publication in 2009 (Bond et al., 2009) it is reported that the WordNet has only 51,000 synsets compared to the circa 120,000 found in OEWN. Romanyshyn et al. (2024) worked on methods for adding Hypo-Hypernym Relations for the recently created Ukrainian WordNet by Siegel et al. (2023). They also experimented with a form of context-aware machine translation as part of a process of identifying Hypo-Hypernym relations for the Ukrainian WordNet, but the focus of their work was not to create a complete WordNet by automated means. Recently, some attempts have been made to compile electronic dictionaries with LLMs in non-English languages, such as in the work of Chow et al. (2024). They worked on a Singlish dictionary that contains 1,783 entries. This work was not done within the context of WordNets though, and is still very limited in scope.

3 Automated Creation of a WordNet with Context-Aware Machine Translation

To create a non-English WordNet by automated means from scratch using OEWN as a base, the only requirement is a machine translation API from English to the target language of your choice. Similar to the process described by Siegel and Bergh (2023), Figure 1 depicts the WordNet creation process. The first step would be to translate all the lemma-definition combinations in OEWN to the target language and save it to a database.

- For each lemma in each synset in OEWN:
  - combine the lemma and the synset definition with a colon
  - then translate it with translation API to obtain the corresponding translated lemma-definition pair
  - save the original as well as translated pair with the corresponding ILI and WordNet ID (of the original in OEWN) in database table

As an example, let us take the OEWN synset ewn-05933552-n:

- lemmas: `criterion` and `standard`
- definition: `the ideal in terms of which something can be judged`

The target language that we are translating into is Afrikaans (a minority Germanic language of South Africa with Dutch roots), and we use the Google Translate API [3]. After applying the algorithm, an extract of the database for this synset is displayed in Table 2. The version of OEWN we used had about 211,000 lemmas, meaning that the complete database will also have 211,000 entries after the algorithm and machine translation have been applied to all the OEWN synsets. We now have all the data necessary to construct the complete Afrikaans WordNet. WordNets are published in various file formats (McCrae et al., 2021), including the XML-based WN-LF format[4] and RDF, but we will take the WN-LF WordNet file for OEWN which can be downloaded from the OEWN website[5] or through the wn python package (Goodman and Bond, 2021) on pypi[6]. We will refer to the WordNet we are about to create as an *inferred WordNet*. Inferred WordNets are always created using the *expansion* approach. We now have to go through a series of steps to replace the data contained in the OEWN WN-LF file with the data contained in our database table with the machine translations:

- Each synset in OEWN has an ILI and a synset ID. We do not want to change the ILI, since it should stay the same across WordNets so that synsets with similar meanings in different languages can stay connected to each other. We do want to change the synset ID though, since this is language specific. Synset IDs in OEWN start with `ewn-`. We replace this with the prefix `in` with stands for inferred WordNet plus the language code; for the Afrikaans WordNet, `ewn-` would be replaced with `inaf-`. For example, `ewn-05933552-n` becomes `inaf-05933552-n`.

- Lemmas or `Lexical Entries` in WordNets are connected to synsets via `Senses`. They also have IDs starting with a prefix; in OEWN it is also `ewn-`. These IDs also needs to be replaced with `inaf-`.

- Now that we have created suitable IDs for all the WordNet elements (Lemmas, Senses and Synsets) in our target language, we also need to copy and replace the machine translated values for the Lemmas and the synset definitions to the correct places in the file. This process is a simple find and replace operation that is done by looking up the correct values in the database table with the machine translations.

- If all the above steps have been followed, we will have a valid WordNet that can be used, but one final clean-up step is still required to make it optimal. There will be some synsets in the target language that have duplicate lemmas. This is because two lemmas that form part of a synset in OEWN, might be translated into the target language as the same lemma. As an example, the OEWN Synset `ewn-05623283-n` which means ‘knowing how to avoid embarrassment or distress’ has two lemmas, prudence and circumspection, which are both translated into ‘versigtigheid’ in Afrikaans. To clean this up, the duplicate lemmas must be removed, and the `Senses` that connect the lemmas with the synsets must be restructured, so that the duplicate lemmas are not connected to the synsets any more.

The inferred Afrikaans WordNet is available online[7]. Similarly, inferred WordNets can also be created for other languages. As part of this project, Deepl[8] was used to create inferred WordNets for the most important European languages[9], and Google Translate was

[3] `https://translate.google.com`
[4] `https://globalWordNet.github.io/schemas/`
[5] `https://en-word.net/`
[6] `https://pypi.org/project/wn/`
[7] `https://github.com/pssvlng/open-afrikaans-wordnet`
[8] `https://www.deepl.com`
[9] `https://github.com/pssvlng/open-european-wordnets-inferred`

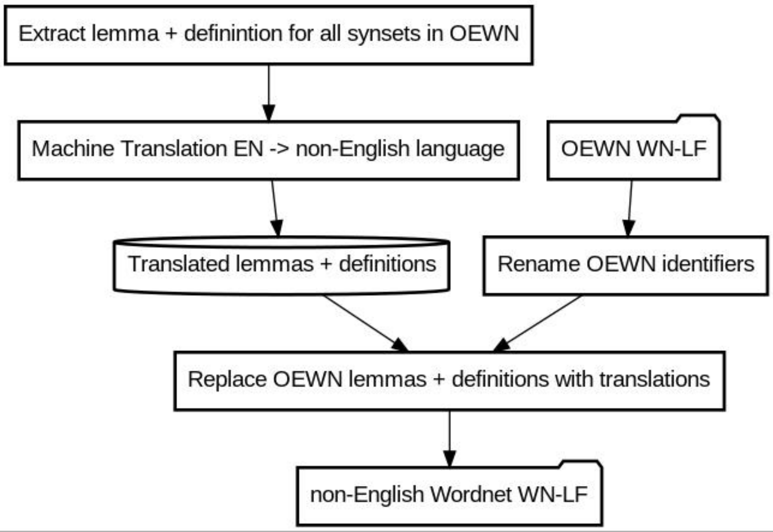

Figure 1: Creation of an Inferred WordNet

ILI	WordNet ID	Source Lemma-Definition	Target Lemma-Definition
i67968	ewn-05933552-n	criterion: the ideal in terms of which something can be judged	kriterium: die ideaal op grond waarvan iets beoordeel kan word
i67968	ewn-05933552-n	standard: the ideal in terms of which something can be judged	norm: die ideaal waarteen iets getoets kan word

Table 2: Context Aware Machine Translation for `ewn-05933552-n` into Afrikaans

used to create WordNets for some of the most important indigenous South African languages[10] and Ukrainian[11]. In theory, with this method, it would be possible to create a WordNet in any language that has machine translation support.

4 Automated Creation of Hybrid WordNets

As mentioned in the introduction, there are many existing non-English WordNets that form part of the OMW project, but have significantly fewer synsets than OEWN. Though lacking in quantity, these existing WordNets are mostly hand-curated, and therefore are of high quality. Consequently, the idea is to merge an inferred WordNet, as described in Section 3, with one of the existing WordNets in the OMW project, thereby creating a complete bilingual resource (i.e. the resulting WordNet in the target language has just as many synsets as OEWN). An inferred WordNet can also be said to be a complete bilingual resource, in the sense that it has the same number of synsets than OEWN, but merging a WordNet from OMW into the inferred WordNet will result in higher quality. A WordNet, formed as a result of merging an inferred WordNet with a WordNet from OMW, will be referred to as a *hybrid WordNet*.

As an example, we take OdeNet, and then the process of creating a hybrid WordNet can be described as follows:

- create an inferred German WordNet $inferred_1$ using the methods described in Section 3
- get a list of ILIs $list_1$ for all the synsets in OdeNet
- remove all the synsets in $inferred_1$ with ILIs in $list_1$, resulting in a modified WordNet $inferred_2$ which will have fewer synsets than $inferred_1$
- insert all synsets from OdeNet into $inferred_2$, resulting in a new WordNet $hybrid_1$ which has the same number of synsets as OEWN and $inferred_1$

Hybrid WordNets were created for the most important European languages, and are available online[12], as is an interface for browsing the Hybrid German WordNet[13].

Hybrid WordNets is a good way to retain the high quality of hand-curated WordNets, but also supplementing it with additional synsets from inferred WordNets in order to create a complete bilingual resource. Since hand-curated synsets are of higher quality, the idea is that the synsets taken from inferred WordNets to supplement hybrid WordNets will become fewer over time as more synsets are added to the hand-curated, non-English OMW WordNet; consequently improving the quality of hybrid WordNets as time progresses.

[10] `https://github.com/pssvlng/open-african-wordnets`
[11] `https://github.com/pssvlng/ua-wordnet`
[12] `https://github.com/pssvlng/open-european-wordnets-hybrid`
[13] `https://edu.yovisto.com/wordnet/`

5 Automated WordNet Creation with LLMs

Though the quality of WordNets created with context-aware machine translation is good for certain languages, there are also a few challenges. Firstly, the quality depends on the language and the machine translation API used. The quality for widely-spoken languages such as German, French and Spanish are much better than those of minority languages. Minority languages done with Google Translate such as Afrikaans proved to be of lower quality. In many instances, the machine translations are correct but use inefficient grammatical structures and repetitions. For example, the OEWN synset `ewn-05622811-n` has the following Lemma-Definition combination in English, followed by the translated Afrikaans and German:

- logic: reasoned and reasonable judgment
- logika: redelike en redelike oordeel
- Logik: begründetes und vernünftiges Urteil

In the Afrikaans translation we see the unnecessary duplication of the word `redelike`, since the machine translation was unable to find suitable words for both `reasoned` and `reasonable`. The German translation fares better and has two unique words in this case, namely `begründetes` and `vernünftiges`. By crafting suitable prompts with LLMs, better results can be obtained for the non-optimal Afrikaans translation. For example, consider the following prompt and result obtained by using GPT4[14]:

- Prompt: The word 'logic' is a noun in English, with the definition 'reasoned and reasonable judgment'. Translate the definition into Afrikaans, making sure that the correct meaning in context is conveyed. Give only the translation in your answer.
- Result: beredeneerde en redelike oordeel

Here we see that GPT4 fared much better than the machine translation and was able to find two suitable words for `reasoned` and `reasonable`, namely `beredeneerde` and `redelike`.

As mentioned previously, there is also the issue that the inferred WordNet constructed from the context-aware machine translations will always have fewer lemmas than OEWN (even though the number of synsets will be the same). This is because two lemmas that form part of a synset in OEWN, might be translated into the target language as the same lemma. Well-crafted LLM prompts can be used to 'recover' some of the lost lemmas as a result of the machine translation. Consider the example of the Afrikaans word 'versigtigheid', mentioned previously, that has the meaning 'prudence' with the definition 'knowing how to avoid embarrassment or distress'. The following prompt with GPT4 helps us to achieve our goal:

- Prompt: The Afrikaans word 'versigtigheid' is a noun that has the meaning 'prudence', with the definition being 'knowing how to avoid embarrassment or distress'. Give synonyms for the word 'versigtigheid'. Your answer should only contain a list of comma-separated words in Afrikaans.
- Result: omsigtigheid, bedagsaamheid, behoedsaamheid, waaksaamheid, sorgsaamheid, verskrokkenheid, voorbedagtheid

We also observed that machine translations do not seem to do well with slang and neologisms. For example, the English neologism 'staycation', referring to a vacation spent at home, is merely translated into 'Urlaub' in German, meaning 'vacation' in English, when using Deepl. GPT4, on the other hand, does a better job. Observe the following prompt and result:

- Prompt: Translate the English word 'staycation' into German, making sure to convey the correct meaning in context. Give only the correct word in your answer.
- Result: Heimurlaub

Here 'Heimurlaub' correctly encoded the semantic meaning in German of a vacation spent at home.

Some languages, especially Germanic languages, are very rich in compound words and other deducible linguistic patterns. LLMs can be used to 'harvest' the most significant compound words in these languages. Consider the following prompt and result in GPT4 for the German word 'Bank', which, as in English, means a financial institution:

- Prompt: The German word 'Bank' refers to a financial institution. In your answer, give only a comma-separated list of the most import compound German words for 'Bank'
- Result: Bankkonto, Bankangestellter, Bankfiliale, Banküberweisung, Bankgeheimnis, Bankkarte, Bankomat, Bankeinzug, Bankkredit, Bankkunde, Banknote, Bankverbindung, Bankwesen, Bankenkrise, Bankenaufsicht

Follow-up prompts can also be crafted to determine if one of these compound words have a connection to 'Bank' in the context of a WordNet. For example, some of these words could either be hypernyms, hyponyms, meronyms or holonyms of 'Bank'.

Finally, a word goes to the non-deterministic nature of LLMs. Non-determinism in LLMs is often criticized, since results are not always reproducible. We can use this to our advantage though, to optimize quality. To illustrate, let us take another example from the Afrikaans WordNet. The OEWN synset `ewn-05625839-n` has a lemma 'brainpower' with a definition of 'mental ability'. The machine translation into Afrikaans looks as follows:

[14] https://openai.com/index/gpt-4/

- breinkrag: geestelike krag

The translation of 'mental ability' into 'geestelike krag' is very clumsy and even partially incorrect, as it rather conveys the meaning of 'spiritual power' instead of mental ability. Now consider the following prompt with GPT4 that was run 3 times, obtaining different results:

- Prompt: The word 'brainpower' is a noun in English with the definition of 'mental ability.' Translate the definition into Afrikaans, making sure that the correct meaning in context is conveyed. Give only the translation in your answer.
- Result 1: mentale vermoë.
- Result 2: geestelike vermoë.
- Result 3: verstandelike vermoë.

All three of the results correctly translated the English word 'ability' as 'vermoë' instead of 'krag'. The correct translation of 'mental' is in `Result 3`, namely 'verstandelike'. In `Result 1` 'mentale' is a suboptimal anglicization, and 'geestelike' in `Result 2` is incorrect. As the second step in this process, we now craft another prompt to try and extract the correct result:

- Prompt: Which one of the following three Afrikaans phrases refering to 'Brainpower' is grammatically the most correct: 1) metale vermoë, 2) geestelike vermoë, 3) verstandelike vermoë. Give only the most correct phrase in your answer.
- Result: verstandelike vermoë

With this additional prompting, the LLM acts as a judge of the before generated information. To verify and compare the effectiveness of the different methods, we conducted an evaluation.

6 Evaluation

For the above-mentioned implementations, an evaluation was done to verify the success rate for German and Afrikaans. A selection of 697 synsets was taken from OdeNet. Hand-curated synsets in OdeNet, where the quality has been verified manually, are marked with a confidence score of 1.0 in the metadata. For each of these synsets we took the ILIs and extracted the first lemma and definition of the corresponding OEWN synset. We then obtained a translated lemma and definition in German using context-aware machine translation, and also using LLM prompts in GPT4. For example, the prompt for getting the translated English adjective 'drunk' (ILI: i5040) in German, looks as follows:

- Prompt: The word 'drunk' is an adjective in English with the definition 'as if under the influence of alcohol'. Translate the word 'drunk' into German, making sure that the correct meaning in context is conveyed. Give only the translated word in your answer.

In addition, a non-context-aware machine translation has been done, using only the lemma.

OdeNet is quite synonym-rich (with many lemmas per synset on average), and as the next step we now try to find out if our translated lemmas for LLM prompts, context-aware machine translation and non-context-aware machine translation are found in the lemma list of the corresponding OdeNet synset. For example, let us take the result we get from the prompt example above (English adjective 'drunk' - ILI: i5040) and compare it with the lemma list of the corresponding OdeNet synset:

- Prompt result: **betrunken**
- OdeNet lemma list (ILI: i5040): ['voll', 'zu', 'berauscht', 'dicht', 'voll wie eine Haubitze', 'alkoholisiert', 'breit', 'strunz', 'unter Alkohol', 'besoffen', 'blau', 'hackevoll', 'trunken', 'im Rausch', 'abgefüllt', **'betrunken'**, 'stoned', 'bezecht', 'hacke', 'strack']

Here we can see that the prompt result is found in the OdeNet lemma list and therefore verified as correct. The results in Table 3 show the number of matches (lemma found in the OdeNet synset) for each of the above-mentioned options for our selected dataset of 697 synsets. Here we can clearly see that the LLM-prompt approach fares the best with 488 matches (or 70%), meaning that these translations are verified as correct because they were found in the lemma list of the corresponding OdeNet synset. As expected, the context-aware machine translations does better than the non-context-aware machine translation, with 437 (63%) vs. 398 (57%) matches in the lemma list of the corresponding OdeNet synset. The remaining entries, which were not found in the lemma list of the corresponding OdeNet synset, are not necessarily incorrect, so these were evaluated manually. For the LLM prompts, 142 of its remaining entries were judged to be correct, while there still remained 67 errors. The context-aware translations had 140 correct results of its remaining entries, with 120 errors. Finally, for the non-context aware translations, 135 of its remaining entries were correct, with 164 errors. Therefore, for the LLM-prompt approach a total of 630 entries were correct, translating to an overall success rate of 90%. The context-aware translations has 577 correct entries with an overall success rate of 83%. Finally, the success rate for the non-context aware translations was 76%, with 533 correct entries.

As a final test, we also took all the 697 synsets, and used the approach as described in the final part of Section 5, i.e. run the prompt 3 times and then ask for the best result in a follow-up prompt (LLM as a Judge). Of the 697 synsets, 430 results were verified as correct, because they were found in the lemma list of the corresponding OdeNet synset. We also found another 86 correct entries from the manual evaluation that has already been done. In the same way, we were also able to

	LLM Prompts	Context-aware Trans.	Non-context Aware Trans.	LLM as a Judge
Automated Matches (exists in OdeNet)	488	437	398	430
Manually Verified Matches	142	140	135	215 (86 + 129)
Manually Verified Errors	67	120	164	52 (32 + 20)
Success Rate	630 (90%)	577 (83%)	533 (76%)	645 (93%)

Table 3: Matches for German (of 697 synsets)

	LLM Prompts	Context-aware Trans.	LLM as a Judge
Matches for Overlapping Subset	348	348	339 (260 + 79)
Errors for Overlapping Subset	4	4	14
Matches for Non-overlapping Subset	272	197	291 (136 + 155)
Errors for Non-overlapping Subset	73	148	53 (19 + 34)
Success Rate	620 (89%)	545 (78%)	630 (90%)

Table 4: Matches for Afrikaans (of 697 synsets)

confirm 32 errors. The remaining 149 entries were also evaluated manually, and 129 were judged to be correct, with a further 20 errors. The overall success rate for LLM as a Judge amounts to 93%, and can be seen in the rightmost column of Table 3.

For the evaluation of Afrikaans, we followed a different approach, since we do not have a hand-curated, manually verified WordNet for Afrikaans, as was the case for German. We used the same set of 697 ILIs as for the German. If the context-aware machine translation and the LLM prompts produced the same result, we suspected that the probability of correctness would be quite high, seeing that it has been confirmed by two different approaches. The 354 overlapping entries, and the remaining 343 entries were evaluated manually as two subsets. The results are presented in Table 4. In the subset with the 354 overlapping entries, 348 (or 99%) were correct, thereby confirming our hypothesis. Of the 343 entries in the second subset, 272 LLM prompt results were correct, and 197 correct for the context-aware translations. In summary, the LLM prompts did much better than the context-aware machine translations, with an overall success rate of 89% vs. 78%. Interesting to note is that of the 148 incorrect context-aware machine translations, 24 were close to correct, but had part of speech inconsistencies. This means that the translation was often correct, but, for example, was given in the noun form instead of the verb form. This problem did not occur with the LLM prompt results, and it also makes sense, since the prompts give more detailed instructions about the part of speech to the LLM, as can be seen from the example prompt shown earlier in this section.

Similar to the German, we also did a LLM as a Judge evaluation on all the synsets for Afrikaans to do a comparison. In the overlapping subset, we were able to confirm 260 correct entries from the manual evaluation already done. The remaining 93 entries were also evaluated manually, with 79 correct results and 14 errors. Similarly, for the non-overlapping subset, 136 entries were confirmed to be correct, and 19 were confirmed to be errors. Of the remaining 189 entries, a manual evaluation showed that 155 were correct, with 34 errors. The overall success rate translates to 90%; making it slightly better than the LLM-prompt approach, as can be seen in the final column of Table 4. The result sets are available online[15].

7 Conclusion

In this paper, we discussed methods for improving the quality and quantity of synsets in WordNets created by automatic means. The ILI allows us to link synsets with similar meaning in different languages to each other. Consequently, we used the OEWN WordNet from the OMW project as a base WordNet from which WordNets in other languages could be created, allowing us to also retain the link between the synsets with the ILI; therefore the resulting WordNet also is a complete bilingual resource.

Inferred WordNets created for languages other than English are realized by doing context-aware machine translations of OEWN. We referred to the usage of machine translation API for getting context-aware machine translations to construct inferred WordNets. In practice, though, we chunked WordNet lemma-definition combinations into a set of documents, which were translated and decoded, to get the results quicker. It enabled us to create inferred WordNets for any language in about 2–3 days. The lemma-definition combinations of the synsets in OEWN were chunked into 26 files and passed through machine translation, which

[15]`https://github.com/pssvlng/gwc-2025-results`

takes less than a day. The remaining part of the process is managed by automated scripts, with minor manual intervention required. This method was used to create inferred WordNets for Afrikaans, Dutch, French, German, Italian, Portuguese, Spanish, Romanian, Ukrainian, as well as some indigenous South African languages, namely, Northern Sotho, Sesotho, Tsonga, Xhosa and Zulu.

A hybrid WordNet takes synsets from a hand-curated, non-English WordNet in OMW, and synsets from an inferred WordNet of the same language, merging them into one WordNet. The high quality, hand-curated synsets from OMW are used, and the supplementation of additional synsets from an inferred WordNet results in a complete bilingual resource. In the context of this project, hybrid WordNets were created for Dutch, French, German, Italian, Portuguese, Spanish. Recently, Siegel et al. (2023) also started working on a Ukrainian WordNet (using the *merge* approach) in the context of the OMW project, using more conventional sources, such as electronic dictionaries. This development opens up the possibility of creating a hybrid Ukrainian WordNet in the foreseeable future.

We introduced some techniques that can be used with LLMs to improve the quality of existing WordNets, but indeed also to construct WordNets from scratch. The commercial LLM models, such as GPT4, provide the best quality for minority languages such as Afrikaans, but it comes at a cost. Constructing a non-English WordNet from scratch with LLMs will require well over a million prompts, and therefore might not be viable for everyone. Open Source LLM models found on platforms such as hugging-face[16] are rapidly improving in quality, while also providing more affordable pricing options. Therefore, the creation of high-quality WordNets for minority languages with LLMs is a definite possibility in the near future.

Acknowledgement

This work has partially been funded by German Federal Ministry of Education and Research (BMBF) under FKZ 16INBI001.

References

Francis Bond, Hitoshi Isahara, Sanae Fujita, Kiyotaka Uchimoto, Takayuki Kuribayashi, and Kyoko Kanzaki. 2009. Enhancing the Japanese WordNet. In *Proceedings of the 7th Workshop on Asian Language Resources (ALR7)*, pages 1–8, Suntec, Singapore. Association for Computational Linguistics.

Francis Bond, Piek Vossen, John McCrae, and Christiane Fellbaum. 2016. CILI: the collaborative interlingual index. In *Proceedings of the 8th Global WordNet Conference (GWC)*, pages 50–57, Bucharest, Romania. Global Wordnet Association.

[16] https://huggingface.co/

Siew Yeng Chow, Chang-Uk Shin, and Francis Bond. 2024. This word mean what: Constructing a Singlish dictionary with ChatGPT. In *Proceedings of the 2nd Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia (EURALI) @ LREC-COLING 2024*, pages 41–50, Torino, Italia. ELRA and ICCL.

Michael Wayne Goodman and Francis Bond. 2021. Intrinsically interlingual: The wn python library for wordnets. In *Proceedings of the 11th Global Wordnet Conference*, pages 100–107, University of South Africa (UNISA). Global Wordnet Association.

Hitoshi Isahara, Francis Bond, Kiyotaka Uchimoto, Masao Utiyama, and Kyoko Kanzaki. 2008. Development of the Japanese WordNet. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Hiroyuki Kaji and Mariko Watanabe. 2006. Automatic construction of Japanese WordNet. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

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.

Steven Neale. 2018. A survey on automatically-constructed WordNets and their evaluation: Lexical and word embedding-based approaches. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Hugo Gonçalo Oliveira. 2023. On the acquisition of wordnet relations in portuguese from pretrained masked language models. In *Proceedings of the 12th Global Wordnet Conference*, pages 41–49.

Nataliia Romanyshyn, Dmytro Chaplynskyi, and Mariana Romanyshyn. 2024. Automated extraction of hypo-hypernym relations for the Ukrainian WordNet. In *Proceedings of the Third Ukrainian Natural Language Processing Workshop (UNLP) @ LREC-COLING 2024*, pages 51–60, Torino, Italia. ELRA and ICCL.

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.

Melanie Siegel and Francis Bond. 2021. OdeNet: Compiling a GermanWordNet from other resources.

In *Proceedings of the 11th Global Wordnet Conference*, pages 192–198, University of South Africa (UNISA). Global Wordnet Association.

Melanie Siegel, Maksym Vakulenko, and Jonathan Baum. 2023. Towards UkrainianWordNet: Incorporation of an existing thesaurus in the domain of physics. In *Proceedings of the 19th Conference on Natural Language Processing (KONVENS 2023)*, pages 121–126, Ingolstadt, Germany. Association for Computational Lingustics.

Piek Vossen, editor. 1998. *Euro WordNet*. Kluwer.

Konrad Wojtasik, Arkadiusz Janz, and Maciej Piasecki. 2023. Wordnet for definition augmentation with encoder-decoder architecture. In *Proceedings of the 12th Global Wordnet Conference*, pages 50–59.