# Automated Extraction of Hypo-Hypernym Relations for the Ukrainian WordNet

**Nataliia Romanyshyn, Dmytro Chaplynskyi, Mariana Romanyshyn**
Ukrainian Catholic University, lang-uk, Grammarly, Ukraine
romanyshyn.n@ucu.edu.ua, chaplinsky.dmitry@gmail.com, mariana.romanyshyn@grammarly.com

**Abstract**

WordNet is a crucial resource in linguistics and natural language processing, providing a detailed and expansive set of lexico-semantic relationships among words in a language. The trend toward automated construction and expansion of WordNets has become increasingly popular due to the high costs of manual development. This study aims to automate the development of the Ukrainian WordNet, explicitly concentrating on hypo-hypernym relations that are crucial building blocks of the hierarchical structure of WordNet. Utilizing the linking between Princeton WordNet, Wikidata, and multilingual resources from Wikipedia, the proposed approach successfully mapped 17% of Princeton WordNet (PWN) content to Ukrainian Wikipedia. Furthermore, the study introduces three innovative strategies for generating new entries to fill in the gaps of the Ukrainian WordNet: machine translation, the Hypernym Discovery model, and the Hypernym Instruction-Following LLaMA model. The latter model shows a high level of effectiveness, evidenced by a 41.61% performance on the Mean Overlap Coefficient (MOC) metric. With the proposed approach that combines automated techniques with expert human input, we provide a reliable basis for creating the Ukrainian WordNet.

**Keywords:** WordNet, Ukrainian, Large Language Models, Hypernym Discovery, Lexicography

## 1. Introduction

WordNet is an invaluable resource that offers a well-structured and comprehensive list of lexical and semantic relationships between words in a language. This highly versatile resource is widely used by experts in linguistics, psychology, and natural language processing (NLP). Unlike a conventional thesaurus, WordNet arranges concepts based on their semantic and lexical relations to other concepts. Its broad applications include word sense disambiguation, machine translation, information retrieval, automatic text classification and summarization (Morato et al., 2004).

In recent years, scholars studying languages other than English have tried to tackle the issue of the absence of digital lexical databases similar to the Princeton WordNet (Miller, 1994). Due to the high expenses associated with creating taxonomies manually, there has been a growing interest in automatic methods for building and enhancing WordNets. Various researches have demonstrated the effectiveness of this approach in producing and expanding WordNets for multiple languages, such as Chinese (Wang and Bond, 2013), Arabic (Elkateb et al., 2006), and Urdu (Adeeba and Hussain, 2011).

The main objective of this paper is to introduce a new approach that utilizes multilingual resources from Wikidata[1] and Wikipedia[2] to build the Ukrainian WordNet. The primary focus of this work is on hypo-hypernym relations, a fundamental type of semantic relation for nouns that reflects the hierarchical structure of WordNet. It links general terms to more specific ones. For example, *rose* is a hyponym of *flower*, which is a hypernym of *rose*.

By concentrating on hypo-hypernymy, we aim to create a strong foundation for the Ukrainian WordNet that can be further expanded with other semantic relations in the future.

This work presents contributions that include:

- Automated methods for constructing and extending the Ukrainian WordNet, specifically linking techniques between Princeton WordNet, Wikidata and multilingual resources from Wikipedia, which have enabled the mapping of 17% of PWN to Ukrainian Wiki.
- Three strategies for generating candidate words to fill gaps in the constructed WordNet basis: machine translation, the Hypernym Discovery model, and Hypernym Instruction-Following LLaMA. The latter achieved high-performance results on the MOC metric (41.61%).
- Established a scalable foundation for creating a comprehensive and reliable WordNet for the Ukrainian language and published the artifacts of this work, including code and data, in the GitHub repo[3].

[1] https://www.wikidata.org/wiki/Wikidata:Main_Page
[2] https://www.wikipedia.org
[3] https://github.com/lang-uk/wikidrill



*The Third Ukrainian Natural Language Processing Workshop (UNLP) @LREC-COLING-2024*, pages 51–60
25 May, 2024. 

The rest of the paper is organized as follows. Section 2 contains an overview of related work. Section 3 describes in detail the pipeline of our approach: compiling the basis for Ukrainian WordNet utilizing existing resources and methods for filling the gaps. We describe the statistics of the datasets obtained using the methodology described in the previous section and introduce the main experiments performed for the Hypernym Discovery task and instruction-tuned LLaMA in Section 4. We discuss the limitations of our approach, draw conclusions, and present future work in Section 5.

## 2. Related Work

The Princeton WordNet of the English language is widely regarded as the most comprehensive and established WordNet (Miller, 1994). With over 117,000 synonym sets and diverse relations, the PWN[4] has formed the benchmark for WordNets in other languages.

In the literature, two common approaches are used for building a WordNet for other languages: merge and expand (Vossen, 1997).

The merge approach involves developing a language-specific semantic network and integrating its synsets with those of the Princeton WordNet in the final stage of the project.

The expand approach involves mapping or translating local words to the synsets of an existing WordNet. While the expand approach is more efficient and requires less linguistic knowledge, it may result in less accurate representations of the semantic and lexical structure of the language.

Nevertheless, many WordNet developers opt for this approach due to the universal structure of lexical semantics that exists across languages (Youn et al., 2016).

The first published works on the construction of the Ukrainian WordNet were carried out in the 2010s.

Kulchytsky et al. (2010) conducted a study that focused on analyzing the relationships between nouns in the Princeton WordNet, selecting core nouns for the Ukrainian language, and organizing them into a hierarchical structure. The resulting WordNet-like dictionary includes 194 synsets, of which 183 are interconnected by hypo-hypernymy, 14 by antonymy, and 150 by meronymy/homonymy. The research in question was conducted manually using frequency dictionaries. Unfortunately, the project was not continued, and the results were not made publicly available.

Anisimov et al. (2013) described the development of a lexical semantic database for the Ukrainian language called UkrWordNet. The article focuses on the research and development of automated techniques for replenishing and extending UkrWordNet. The method developed for creating new nodes involved generating them from Ukrainian Wikipedia articles and binding them to the synsets of UkrWordNet. The paper also proposed a new measure of semantic similarity using latent semantic analysis (Deerwester et al., 1990) to improve the quality of the bindings. After manual post-processing, UkrWordNet contained over 82,000 synsets and approximately 145,000 nouns in the lexicon. Unfortunately, the work has never been publicly released.

In their article, Siegel et al. (2023) introduced Ukrajinet 1.0[5], a lexical database centered around physics terminology. The database contains 3,360 synonym sets of 8,700 words and shares a methodology similar to that used in creating OdeNet[6] for the German language (Siegel and Bond, 2021). However, Ukrajinet 1.0 does not include hypo-hypernym relations, essential for establishing a hierarchical structure of nouns within the WordNet framework.

Other developments in the field of the Ukrainian WordNet include materials[7] from theses of students of Lviv Polytechnic National University, but they are of a limited size.

Hence, developing an open-source WordNet for the Ukrainian language, with a representative number of relations, remains an ongoing area for research.

## 3. Proposed Approach

Our methodology for creating the basis of the Ukrainian WordNet builds on the expand approach. Figure 1 summarizes the proposed methodology. We propose utilizing the Princeton WordNet as a pivot structure, and linking it to Wikidata and Ukrainian Wikipedia. By mapping Ukrainian Wikipedia titles to synsets in the PWN and identifying hyponyms for each synset, a tree diagram is constructed using these resources. The resulting tree contains nodes that could not be linked to Ukrainian Wikipedia and thus lack a Ukrainian equivalent. We call them gap nodes and further propose the Gap Ranking algorithm to identify the best gap nodes for filling. To generate candidate words to fill these gaps, several strategies are proposed. The first strategy utilizes English lemmas translated into Ukrainian with Google Translate, Bing, and DeepL. The second strategy adapts the Hypernym Discovery task for Ukrainian and gener-

[4] https://wordnet.princeton.edu

[5] https://github.com/hdaSprachtechnologie/ukrajinet

[6] https://github.com/hdaSprachtechnologie/odenet

[7] https://github.com/lang-uk/wordnet/tree/main/resources

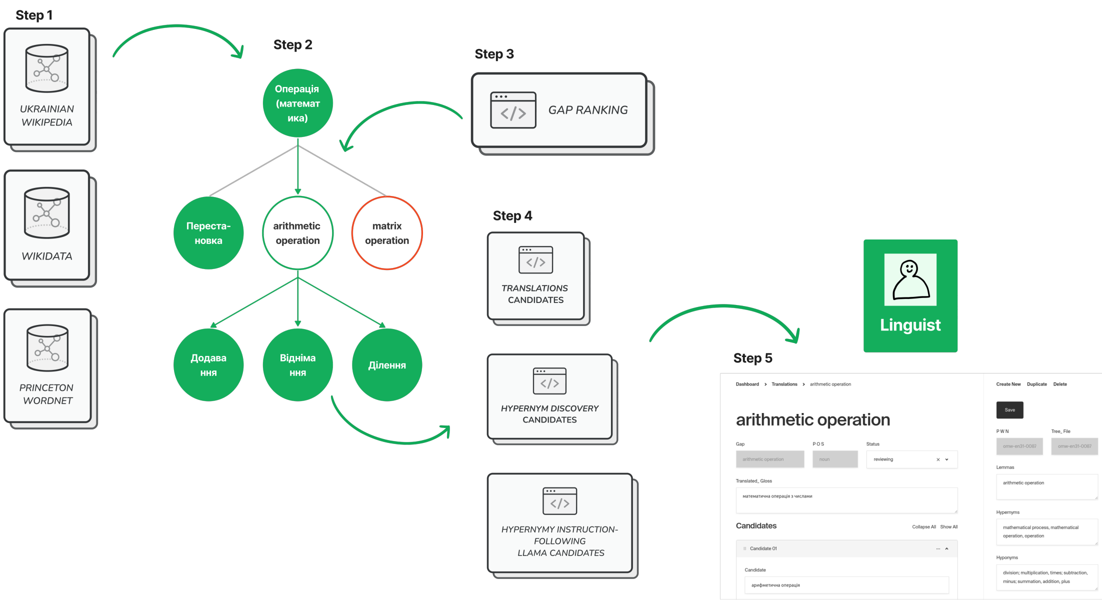


Figure 1: Overview of the proposed methodology for developing the Ukrainian WordNet foundation through integration with Princeton WordNet, Wikidata, and Wikipedia.

ates candidates given the gap hyponym. The third strategy generates hypernym candidates with the Instruction-Following LLaMA model. The hypernym candidates generated via the three strategies are then aggregated in a MongoDB and surfaced in an annotation tool built with Payload CMS to streamline further human annotation. Overall, the proposed approach combines automated techniques with expert human input to create a comprehensive and reliable resource for the Ukrainian language.

### 3.1. PWN and Wikidata

Our methodology leverages the linking between Princeton WordNet and Wikidata, as proposed by McCrae and Cillessen (2021). First, we utilize the synset's ID to identify the PWN synset linked with Wikidata. This link allows us to search for the corresponding Ukrainian Wikipedia article using the Wikidata Q identifier. At this point, we encounter two possible scenarios. If the search yields a result, we acquire a word that can populate a node in our lexical tree. However, if the search does not provide any results, we temporarily store the English lemma from PWN at this node, intending to address this issue later. The techniques for filling these gaps are elaborated upon in subsequent sections. Once we have identified a linked synset, we proceed to discover hyponyms associated with the given synset ID.

### 3.2. Gap Ranking

The Gap Ranking algorithm aims to identify the most suitable gap nodes for filling, specifically those with the highest number of non-gap children in the given tree. We consider these most suitable because filling them creates the highest number of links. The algorithm utilizes a depth-first search (DFS) tree traversal method to determine the ideal path. Beginning from the root node, it recursively navigates the tree, viewing each node as a potential gap node. For each gap node, the algorithm computes the number of valid pairs of nodes in its subtree by considering its non-gap children. The algorithm then ranks the gap nodes based on the number of identified valid pairs.

This metric is instrumental in identifying the gap nodes with the greatest potential for enhancing the quality of the Ukrainian WordNet. With this algorithm, we found that completing 793 gaps would result in 5403 new hyper-hyponym pairs in the Ukrainian WordNet.

### 3.3. Candidate Generation

We used two methods to generate candidates to fill gaps in our lexical resource. The first method involved automatic translation from English to Ukrainian. The second method used the hyponym of the gap to generate hypernyms with the help of the Hypernym Discovery model and Instruction-Following LLaMA.

| Gap | DeepL Direct | DeepL Contextualized | Translated PWN3.1 |
|---|---|---|---|
| performance | продуктивність<br>produktyvnist | вистава<br>vystava | вистава, спектакль<br>vystava, spektakl |
| head cabbage | качанна капуста<br>kachanna kapusta | качанна капуста<br>kachanna kapusta | головна капуста<br>holovna kapusta |
| agency | агентство<br>ahentstvo | агентство<br>ahentstvo | офіс, орган<br>ofis, orhan |

Table 1: Comparison examples of gap translations obtained using machine translation methods. All terms are nouns. The gap is identified as the most optimal for filling using the algorithm described in Section 3.2

#### 3.3.1. Machine Translations

To run automatic translation, we utilized three distinct methods. Initially, we accessed the existing Ukrainian translation[8] of Princeton WordNet 3.1, which was developed with Google Translate and Bing. Subsequently, we relied on the neural machine translation capabilities of DeepL (Ronzon, 2018). This process entailed directly translating individual lemmas and creating contextual sentences in the format of *"<Synset lemmas> is a <PWN gloss>."*, from which we extracted the first lemma and recorded it as a candidate for the gap.

Ultimately, this approach enabled us to promptly produce a list of potential translations for the gap nodes, although due to the lack of specialized training or fine-tuning of the machine translation models for the Ukrainian language their accuracy remains arguable. For example, machine translation can generate Russianism[9], such as "kachanna kapusta" seen in row 2 of Table 1, or false concepts like "holovna kapusta," which do not exist in Ukrainian. Furthermore, the issue of ambiguity, demonstrated in rows 1 and 3, presented challenges by offering multiple possible senses. Although employing specialized Word Sense Disambiguation systems, as suggested by Laba et al. (2023), could mitigate this issue, exploring such solutions falls beyond the scope of this paper.

#### 3.3.2. Hypernym Discovery and LLaMA

To perform Hypernym Discovery in the Ukrainian language, we adopted the setting provided for this task by Camacho-Collados et al. (2018). We utilized the supervised part of the model proposed by Bernier-Colborne and Barrière (2018), the SemEval-2018 Task 9 winners. Their approach uses pre-trained word embeddings and projection learning to discover the hypernyms of a given query (hyponym).

Pretrained large language models (LLMs) have showcased remarkable results in various natural language processing (NLP) tasks, leading us to explore their potential for Hypernym Discovery. A previous study conducted by Hanna and Mareček (2021) utilized a prompting methodology to investigate BERT's (Devlin et al., 2019) understanding of hypernymy. Our research focused on the potential of another advanced LLM, multilingual LLaMA (Touvron et al., 2023), which has exhibited exceptional performance on various NLP benchmarks. Instead of prompting, we opted to fine-tune LLaMA by providing hypernym instructions to determine if it can suggest hypernyms.

### 3.4. Evaluation Metrics

Camacho-Collados et al. (2018) proposed evaluating the Hypernym Discovery systems as a soft ranking problem. This involved utilizing the top $N$[10] hypernyms generated by the model and evaluating performance using Information Retrieval (IR) metrics:

1. **Mean Reciprocal Rank** (MRR) measures how well a system is able to rank the relevant hypernyms by rewarding the position of the first correct result in the ranked list of outcomes.
2. **Mean Average Precision** (MAP) measures the average correctness of retrieved hypernyms for each query word and averages these across all dataset queries.
3. **Precision at k** (P@k) measures the number of correctly retrieved hypernyms at different cut-off thresholds.

To better understand the model's ability to predict relevant hypernyms regardless of their order, we propose the Mean Overlap Coefficient (MOC) as an additional evaluation criterion:

$$MOC = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{|GT_i \cap P_i|}{|GT_i|}, \tag{1}$$

where Q represents the number of queries, GT represents the set of ground truth hypernyms, and

[8] https://github.com/lang-uk/wordnet/tree/main/pwn_translated_basic

[9] https://en.wikipedia.org/wiki/Russianism

[10] Set the value to 6 for our experiments.

P represents the predictions for a given input term. The numerator calculates the number of common hypernyms between the ground truth and predicted sets, while the denominator ensures that the metric is normalized by the size of the ground truth set.

For our specific task of generating candidates for professional annotators, we found the MOC score to be the most helpful metric as it indicates the proportion of relevant values predicted regardless of their order.

We used the same metrics to measure the performance of the Instruction-Following LLaMA.

# 4. Experimental Results

## 4.1. WordNet Basis

To link the data from PWN, Wikidata, and Ukrainian Wikipedia, we implemented a Python scraper using the web-crawling framework Scrapy (Hoffman et al., 2008), wtf_wikipedia library (Kelly, 2017) for Wikipedia parsing, wn package (Goodman and Bond, 2021), which provides an interface to WordNet data, and an RDF (Resource Description Framework) query language SPARQL (Prud'hommeaux and Seaborne, 2008).

We managed to link 17% of the Princeton WordNet, resulting in **21,015** synsets forming the foundation of the Ukrainian WordNet. Out of the 127,020 PWN3.1 synsets, we could link 23% to Wikidata; subsequently, 17% of those synsets were connected to the Ukrainian Wikipedia. These results reflect the linking percentage as of April 2023. Due to the dynamic nature of Wikidata, with its continuous updates and expansions, subsequent iterations of this experiment could yield an even higher proportion of linked synsets. Table 2 provides an overview of the general statistics.

| | **# synsets** | **% synsets** |
|---|---|---|
| **PWN3.1** | 127,020 | 100% |
| **-> Wikidata** | 29,730 | 23% |
| **-> Ukrainian Wiki** | 21,015 | 17% |

Table 2: General statistics related to the development of the Ukrainian WordNet basis, including the total number of synsets in the PWN3.1, the number and percentage of synsets linked to Wikidata and the Ukrainian Wikipedia.

In addition, we developed a dataset of Ukrainian Hypernymy Pairs consisting of noun pairs that express hypernymy relationships between words. Please, refer to Table 3 for the detailed dataset statistics. We maintained the partition of hypernyms and hyponyms with their instances that is offered by PWN in our dataset. In this split, the instance hypernym denotes a reflexive type, while an instance hyponym represents a specific instance of something. For example, the instance hypernym of *Dnipro River* is *river*. We identified a few data samples where the word on the left is the same as the word on the right (e.g., <river, river> pair), resulting from multiple WordNet IDs linking to the same Wikidata page. To improve the quality of the dataset, we removed such entries. The dataset[11] is available for public use through the Hugging Face platform and can be particularly useful for the Hypernym Detection task, which involves presenting a model with pairs of words and asking it to determine whether a specific relationship exists between them.

| **Relation Type** | **% pairs** |
|---|---|
| Hypernym-Hyponym | 6,906 |
| Co-Hyponyms | 42,860 |
| Hypernym-Instance | 2,971 |
| Co-Instances | 22,927 |

Table 3: Ukrainian Hypernymy Pairs dataset statistics. This table presents the number of word pairs obtained for each type of relationship.

## 4.2. Hypernym Discovery

To advance research in the field of hypernym discovery, *SemEval-2018 Task 9*[12] was organized (Camacho-Collados et al., 2018). The participants were asked to build a system that discover suitable hypernyms from a target corpus given an input term. The organizers (of the task) provided a reliable framework for evaluating proposed models with the IR metrics described in Section 3.4. To perform Hypernym Discovery in the Ukrainian language, we adopted the setting provided for this task.

### 4.2.1. Dataset Creation

Following the approach of Camacho-Collados et al. (2018) in SemEval, our data gathering process involved a series of sequential steps, beginning with the compilation of a vocabulary. Our objective was to establish an all-encompassing list of prospective hypernyms by identifying words that appeared at least five times within the chosen corpus. To do so, we utilized UberText 2.0[13] (Chaplynskyi, 2023), a corpus that boasts 31GB of data and around 2.5 billion tokens, which accurately represents the variety and abundance of the Ukrainian language.

The original Hypernym Discovery dataset consisted of two main components: input hyponym

[11] https://huggingface.co/datasets/lang-uk/hypernymy_pairs

[12] https://competitions.codalab.org/competitions/17119

[13] https://lang.org.ua/en/ubertext/

along with its type and gold hypernyms. The type is either a concept (hyponym) or a named entity (instance). Utilizing the created Ukrainian WordNet basis, we automated the extraction of these terms, including direct and indirect hypernyms up to five nodes deep to mirror the original setup. The refinement process involved:

- Excluding overly broad terms from the upper levels of the WordNet hierarchy;
- Normalizing entries by removing bracketed information that comes from the Wikipedia titles;
- Discarding non-unigram terms;
- Eliminating entries composed of Latin characters, which usually denote animal species, plants, etc.;
- Excluding terms without a direct hypernym relation.

We maintained a frequency threshold, requiring terms to appear at least five times in the UberText corpus. The classification of input terms, as instances or hyponyms was determined automatically via synset relation parameters.

The resulting dataset[14], consisting of 4,890 samples, offers a balanced split for training and test sets alongside a smaller trial set for developmental evaluation.

#### 4.2.2. Model Setup

This work employs the supervised part of the Hybrid Approach to Hypernym Discovery, developed by Bernier-Colborne and Barrière (2018) and accessible on GitHub[15].

To establish a baseline, we utilized 200-dimensional word2vec embeddings with a skip-gram model, trained according to the specifications outlined in the abovementioned research (HD_Baseline). As the next step, we chose to explore the fasttext embeddings, which are advantageous for Ukrainians because of their ability to capture subword information (HD_Fasttext). Hyperparameters were based on previous studies (Romanyshyn et al., 2023), and the vector size was increased to 300 dimensions.

#### 4.2.3. Results

Table 4 summarizes the model's performance by each metric. Overall, we can see that the HD_Baseline model performed the best overall, but HD_Fasttext achieved the highest score in terms of the MOC metric.

[14] https://github.com/lang-uk/wikidrill/tree/main/hypernymy_discovery/hd_dataset

[15] https://github.com/gbcolborne/hypernym_discovery

### 4.3. Hypernym Instruction-Following LLaMA

We utilized a parameter-efficient tuning technique called low-rank adaptation (LoRA) to fine-tune LLaMA-7B on hypernymy instructions (Hu et al., 2021). This approach involves freezing the pretrained model's weights and adding trainable rank decomposition matrices into each layer of the transformer architecture, reducing the number of trainable parameters for downstream tasks (Maurya, 2023).

#### 4.3.1. Intructions Dataset

We developed instruction datasets of three different types and ran experiments on them. The data for Hypernym Discovery was used as a basis. The main difference is that we merged training and trial (dev) sets into one.

**Lean Approach.** Our initial method involved generating simple prompts that instructed the model to provide a specific number of hypernyms for a given term. For example, we would ask the model to *"Generate six hypernyms for 'lavender'."* While we could generate **2,490** instructions, the model's performance was poor.

**Full Setup.** We improved the instruction set by creating 19 distinct patterns for each query, using ChatGPT for initial generation, and manually validating the results. This approach greatly enhanced model performance, resulting in **47,310** input prompts. We utilized various query formats, including *"What are broader terms for 'lavender'?"* to broaden the model's comprehension across similar phrasings.

**Multiple Relations.** Building on our enhanced approach, we introduced instructions for hypernyms, hyponyms, and co-hyponyms, maintaining 19 hypernym patterns while adding 13 for co-hyponyms and 14 for hyponyms. This resulted in **78,149** samples, but we noticed a dip in performance, suggesting potential overgeneralization. Further research is needed to balance instruction diversity and specificity effectively.

#### 4.3.2. Results

Our testing across the Lean, Full, and Multiple models utilized identical input queries and gold hypernyms from the Hypernym Discovery dataset, with tailored strategies to mitigate specific challenges encountered in each setup.

For the Lean model, given its simplicity, we applied a heuristic of repeating each instruction three

| | MOC | MRR | MAP | P@1 | P@3 | P@6 |
|---|---|---|---|---|---|---|
| **HD_Baseline** | 26.55 | **29.23** | **20.84** | **25.25** | **20.22** | **19.3** |
| **HD_Fasttext** | **27.63** | 28.7 | 19.87 | 22.42 | 19.53 | 18.76 |

Table 4: Our Hypernym Discovery systems performance on the test set. HD_Baseline refers to the model with word2vec embeddings, HD_Fasttext to the one using fasttext. The best score for each model is marked in **bold**.

| | MOC | MRR | MAP | P@1 | P@3 | P@6 |
|---|---|---|---|---|---|---|
| **LLaMA_Hypernymy_Lean** | 6.38 | 4.54 | 2.92 | 3.08 | 2.88 | 2.8 |
| **LLaMA_Hypernymy_Full** | **41.61** | **42.6** | **36.74** | **39.0** | **36.27** | **35.93** |
| **LLaMA_Hypernymy_Multiple** | 37.07 | 35.48 | 31.19 | 30.42 | 31.72 | 30.8 |

Table 5: The LLaMA fine-tuning results with hypernymy instructions using different setups. The LLaMA_Hypernymy_Lean setup only uses the most basic hypernymy instructions, while LLaMA_Hypernymy_Full includes 19 instruction patterns for a single input query. In the Multiple setup, three relation types were used in addition to diverse patterns.

times and aggregating unique hypernym candidates to counteract issues of non-responses or repetitive outputs. This approach aimed to enhance result reliability.

In contrast, while not facing duplication issues, the Full and Multiple setups sometimes produced no candidates. To address this, we diversified the testing instruction set, employing four varied prompts to elicit hypernyms, thus balancing output richness and relevance. This method prioritized candidate frequency and maintained the model's original proposal order for equally frequent terms, aligning closely with the evaluative framework of the Hypernym Discovery task. The prompts were as follows:

1. Надай мені декілька гіперонімів до слова "input_term". (Give me some hypernyms to the word "input_term".)

2. Надай мені шість гіперонімів до слова "input_term". (Give me six hypernyms to the word "input_term".)

3. Які слова є гіперонімами поняття "input_term"? (Which words are hypernyms of the term "input_term"?)

4. Які загальні поняття описують слово "input_term"? (What general concepts describe the word "input_term"?)

Table 5 showcases the superior performance of the LLaMA_Hypernymy_Full model across all metrics, reflecting the effectiveness of our comprehensive and nuanced instruction and evaluation methodology.

## 4.4. Error Analysis

In addition to analyzing quantitative results, we also performed a qualitative evaluation of the outputs produced by our top-performing models based on MOC scores from our experiments. Figure 2 presents a metrics comparison of HD_Fasttext and LLaMA_Hypernymy_Full.

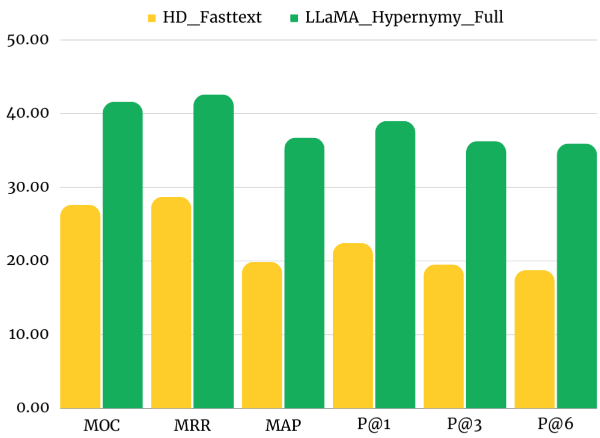


Figure 2: Metrics comparison of the two top-performing models based on MOC score for all entity types.

We randomly sampled several examples from our testing dataset to further investigate the models' predictions.

As we can see from Table 6, the model tends to overfit to frequently occurring hypernyms such as метрополіс (metropolis), муніципалітет (municipality), дистрикт (district), and організм (organism), resulting in incorrect predictions. However, when these candidates are true hypernyms, the model generally ranks them as the top predictions. Moreover, the MOC metric proves helpful in cases where the ground truth contains only one hypernym, such as the річка (river) hypernym for the Ориноко (Orinoco) query and the model ranks it as the last candidate. Notably, the model can also suggest

| Query | OC | HD_Fasttext Predictions |
|---|---|---|
| | | **High OC examples** |
| молюски | 67 | организм, їжа, безхребетні, дистрикт, артефакт, рослини |
| shellfish | | organism, food, invertebrates, district, artifact, plants |
| Ориноко | 100 | организм, дистрикт, муніципалітет, метрополіс, артефакт, річка |
| Orinoco | | organism, district, municipality, metropolis, artifact, river |
| | | **Low OC examples** |
| Гіпатія | 17 | дистрикт, организм, муніципалітет, артефакт, їжа, метрополіс |
| Hypatia | | district, organism, municipality, artifact, food, metropolis |
| Сапфо | 0 | метрополіс, артефакт, организм, дистрикт, муніципалітет, їжа |
| Sappho | | metropolis, artifact, organism, district, municipality, food |
| **Query** | **OC** | **LLaMA_Hypernymy_Full Predictions** |
| | | **High OC examples** |
| холангіт | 100 | симптом, запалення, хвороба |
| cholangitis | | symptom, inflammation, disease |
| Неккар | 100 | річка |
| Neckar | | river |
| | | **Low OC examples** |
| метамфетамін | 0 | опіати, наркотик, анальгетики |
| methamphetamine | | opiates, narcotic, analgesics |
| Сент-Джонс | 0 | озеро, річка |
| St. John's | | lake, river |

Table 6: Examples of predictions made by the HD_Fasttext and LLaMA_Hypernymy_Full models, showing input queries, overlap coefficients (OC), and top predicted hypernyms. High OC values indicate accurate predictions, while low values reflect mismatches. Correct predictions are underlined.

relevant candidates absent in the ground truth, as observed in the Low OC Entity examples, where it proposed организм (organism) as a hypernym for Сапфо (Sappho), which is not the direct hypernym but still relevant as it is the same case as for query hyponym Гіпатія (Hypatia), where the организм (organism) was present in gold hypernyms.

The instruction-following LLaMA model appears to be confident in its predictions, often providing the same answer for four instructions, resulting in fewer variants of predictions. For instance, it predicts the single hypernym річка (river) for the input term Неккар (Neckar). Furthermore, in this scenario, the memorization problem of frequent hypernyms is less noticeable.

In addition, the model can predict relevant hypernyms that are not present in the ground truth set, such as хвороба (disease) for the input word холангіт (cholangitis) and наркотик (narcotic) for метамфетамін (methamphetamine).

Another challenge the model faces is the ambiguity of some hyponyms. For instance, by providing the hypernym річка (river) for the entity Сент-Джонс (St. John's), the model may have referred to an actual river in Florida, United States, while our data referred to a city in Canada.

## 5. Discussion and Conclusion

This paper reports on the ongoing efforts in building the Ukrainian WordNet. We proposed a data-driven approach for automated hypernym hierarchy construction. By mapping PWN, Wikidata, and Wikipedia, we have created a robust foundation for this new WordNet resource. Additionally, we have developed a simple Gap Ranking algorithm to determine the best gap nodes for filling.

To generate candidates for filling the gaps, we have explored various techniques, including machine translation that uses the current missing node in the tree and two others that use information about its children — Hypernym Discovery and Instruction-Following LLaMA.

To adapt SemEval 2018 Task 9: Hypernym Discovery to the Ukrainian language, we have created Hypernym Discovery datasets and utilized an existing large language corpus Ubertext2.0.

Furthermore, we have investigated the capabilities of state-of-the-art LLMs for solving the Hypernym Discovery task. We have demonstrated how to construct a sufficiently large set of instructions from an initial small dataset and how LLMs can be fine-tuned to create a chatbot-like assistant specializing in a particular hypernym suggestion task.

## 5.1. Limitations

Please be aware that our work is subject to certain limitations.

To establish a WordNet basis, we initially mapped the Ukrainian language to English, which may not fully capture all linguistic nuances and cultural phenomena and could contain errors. Hence, it is crucial to have further professional verification and input from linguists.

Another restriction is that, according to our approach, each obtained synset is represented by only one lemma due to Wikipedia articles being primarily represented by one word and linking is on the synset level. As a result, additional effort is required to add synonyms to the obtained lemma-synsets.

Overall, our approach is limited to only creating hypo-hypernym relations. Further research is needed to include other lexico-semantic relations. Nevertheless, it is important to note that the proposed method has the potential to be adapted for other languages as long as comprehensive Wikipedia data is available.

## 5.2. Future Work

As creating WordNet is a complex and lengthy process, there is ample opportunity for future research to improve its coverage and quality. To this end, we have identified critical areas for improvement that we hope to focus on going forward:

1. One priority is to leverage Wikipedia as a constantly updated resource by rerunning the linking algorithm of Wikidata and Ukrainian Wiki to obtain more initial pairs. Additionally, we can independently add links to Wikidata using annotated gaps, thereby enhancing this resource.
2. Exploring larger LLaMA or other open-source language models is another promising direction that can significantly boost performance on our task.
3. An essential next step is to create a high-quality and comprehensive manual for annotators, which will take the WordNet development pipeline to a new level.
4. Ultimately, WordNet should have a user-friendly interface accessible to the general public.

# 6. Bibliographical References


Farah Adeeba and Sarmad Hussain. 2011. Experiences in building Urdu WordNet. In *Proceedings of the 9th Workshop on Asian Language Resources*, pages 31–35, Chiang Mai, Thailand. Asian Federation of Natural Language Processing.

Anatoly Anisimov, Oleksandr Marchenko, Andrey Nikonenko, Elena Porkhun, and Volodymyr Taranukha. 2013. Ukrainian wordnet: Creation and filling. In *Proceedings of the 10th International Conference on Flexible Query Answering Systems - Volume 8132*, FQAS 2013, page 649–660. Springer-Verlag.

Gabriel Bernier-Colborne and Caroline Barrière. 2018. CRIM at SemEval-2018 task 9: A hybrid approach to hypernym discovery. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 725–731, New Orleans, Louisiana. Association for Computational Linguistics.

Jose Camacho-Collados, Claudio Delli Bovi, Luis Espinosa-Anke, Sergio Oramas, Tommaso Pasini, Enrico Santus, Vered Shwartz, Roberto Navigli, and Horacio Saggion. 2018. SemEval-2018 task 9: Hypernym discovery. In *Proceedings of the 12th International Workshop on Semantic Evaluation*, pages 712–724, New Orleans, Louisiana. Association for Computational Linguistics.

Dmytro Chaplynskyi. 2023. Introducing UberText 2.0: A corpus of modern Ukrainian at scale. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop*, pages 1–10, Dubrovnik, Croatia. Association for Computational Linguistics.

Scott C. Deerwester, Susan T. Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. 1990. Indexing by latent semantic analysis. *J. Am. Soc. Inf. Sci.*, 41:391–407.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Sabri Elkateb, William Black, Piek Vossen, David Farwell, Horacio Rodríguez, Adam Pease, Musa Alkhalifa, and Christiane Fellbaum. 2006. Arabic WordNet and the challenges of Arabic. In *Proceedings of the International Conference on the Challenge of Arabic for NLP/MT*, pages 15–24, London, UK.

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.

Michael Hanna and David Mareček. 2021. Analyzing BERT's knowledge of hypernymy via prompting. In *Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, pages 275–282, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Pablo Hoffman, Daniel Graña, and Martin Olveyra. 2008. A fast and powerful scraping and web crawling framework.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *CoRR*, abs/2106.09685.

Spencer Kelly. 2017. Wtf_wikipedia: A pretty-committed wikipedia markup parser.

I. Kulchytsky, A. Romaniuk, and Kh. Khariv. 2010. Rozroblennia wordnet-podibnoho slovnyka ukrainskoi movy [developing a wordnet-like dictionary of ukrainian]. In *Bulletin of Lviv Polytechnic National University*, pages 306–318.

Yurii Laba, Volodymyr Mudryi, Dmytro Chaplynskyi, Mariana Romanyshyn, and Oles Dobosevych. 2023. Contextual embeddings for Ukrainian: A large language model approach to word sense disambiguation. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*, pages 11–19, Dubrovnik, Croatia. Association for Computational Linguistics.

Aniket Maurya. 2023. Accelerating llama with fabric: A comprehensive guide to training and fine-tuning llama.

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.

Jorge Morato, Miguel Ángel Marzal, Juan Lloréns, and José Moreiro. 2004. WordNet Applications. pages 270–278.

Eric Prud'hommeaux and Andy Seaborne. 2008. SPARQL Query Language for RDF. W3C Recommendation. `http://www.w3.org/TR/rdf-sparql-query/`.

Nataliia Romanyshyn, Dmytro Chaplynskyi, and Kyrylo Zakharov. 2023. Learning word embeddings for Ukrainian: A comparative study of fastText hyperparameters. In *Proceedings of the Second Ukrainian Natural Language Processing Workshop*, pages 20–31, Dubrovnik, Croatia. Association for Computational Linguistics.

Thomas Ronzon. 2018. Deepl - überraschend anders. *Javaspektrum*, (2):69.

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.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models.

Shan Wang and Francis Bond. 2013. Building the Chinese open Wordnet (COW): Starting from core synsets. In *Proceedings of the 11th Workshop on Asian Language Resources*, pages 10–18, Nagoya, Japan. Asian Federation of Natural Language Processing.

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


## 7. Language Resource References


Miller, George A. 1994. *WordNet: A Lexical Database for English*. ISLRN 379-473-059-273-1.

P.J.T.M. Vossen. 1997. *EuroWordNet: a multilingual database for information retrieval*. Vrije Universiteit, ISLRN 129-018-230-332-2. Proceedings of the DELOS workshop on Cross-language Information Retrieval, March 5-7, 1997, Zurich.