

UKRAINIAN CATHOLIC UNIVERSITY

BACHELOR THESIS

---

# Large Language Model as a Lexicographer: Gloss Generation for Ukrainian WordNet

---

*Author:*

Hanna YERSHOVA

*Supervisor:*

MSc. Dmytro CHAPLYNSKIY,  
MSc. Mariana ROMANYSHYN

*A thesis submitted in fulfillment of the requirements  
for the degree of Bachelor of Science*

*in the*

Department of Computer Sciences and Information Technologies  
Faculty of Applied Sciences



Lviv 2024

*"I was so ahead of the curve, the curve became a sphere  
Fell behind all my classmates and I ended up here"*

Taylor Swift. Lyrics to "this is me trying"

UKRAINIAN CATHOLIC UNIVERSITY

Faculty of Applied Sciences

Bachelor of Science

**Large Language Model as a Lexicographer: Gloss Generation for Ukrainian  
WordNet**

by Hanna YERSHOVA

*Abstract*

WordNet serves as a valuable tool within the linguistic and natural language processing field, providing a structured lexical database of the language, organizing words into synsets (synonym sets), and establishing semantic relationships between them. As manual creation of WordNet requires lots of resources, automatic methods of building WordNets become popular. Unfortunately, work on developing Ukrainian WordNet is limited in terms of scope and accessibility for the general public, with a majority of the efforts concentrated on manual formation. This thesis addresses the need for an enhanced lexical resource for the Ukrainian language by extending the existing Ukrainian WordNet with comprehensive glosses. Leveraging automated methods and large language models, the research aims to enhance the completeness and utility of the Ukrainian WordNet. This study explores the effectiveness of a large language model in the role of lexicographer and supplements the Ukrainian WordNet basis that consists of 29,494 synsets with glosses, thereby making a step towards filling the gaps in linguistic resources for the Ukrainian language.

## *Acknowledgements*

I am insanely grateful to my supervisors, Dmytro Chaplynskyi and Mariana Romanyshyn, for letting me work with them, proposing such an amazing topic, always being ready to help me, and all the support I received from you. Also, I thank Natalia Romanyshyn, who made such an incredible basis for my work and was always there for me, providing great advice and support. I am not very good with words and I can not describe how happy I am to be able work with you three, your passion and dedication for developing resources for Ukrainian language is truly inspiring. Again, I want to thank you for all the work that you have done with me, you are the best supervisors I could ever have.

I want to say a huge thank you to my mom, who gave me nothing but support when she saw how much I struggled to get to this point. Also, I want to express my gratitude to Olya, my right hand, my throw-up-in-the-bathroom-but-still-love-them, my lifeline, my ride or die, my hate-you-but-you-know-that-that's-a-damn-lie, thank you for being my best friend, for all your unconditional support and just for being there for me. And, Nastya, although there are thousands of kilometers between us and we don't talk as often as we should, I always know that can count on you, that you will respond to my calls at 3 am and give support when I need it the most, I hope that we will be able to meet very soon.

This thesis would not exist without the Armed Forces of Ukraine, who sacrificed themselves and gave me the opportunity to continue living a peaceful life and pursue my studies during the ongoing Russo-Ukrainian war. Glory to Ukraine! Glory to the heroes!

I can not forget about the Faculty of Applied Sciences. At the same time, it feels like I just came to this university for the first time yesterday, but I have lived 127 different lives in my time here. Thank you to everyone on the university team who worked so hard to give me this education.

Lastly, I want to thank all my oomfs. This might sound silly, but all of you make me so happy, and I am so-so happy that I met so many incredible people who are always there to support my delusion. I am sending a big virtual hug to all of you.

# Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Contributions . . . . .	1
1.3 Structure of the Thesis . . . . .	2
<b>2 Theoretical Background</b>	<b>3</b>
2.1 WordNet . . . . .	3
2.1.1 WordNet Structure . . . . .	3
Nouns . . . . .	4
Adjectives and Adverbs . . . . .	5
Verbs . . . . .	5
2.2 Glosses in WordNet . . . . .	7
2.3 Definitions in Explanatory Dictionary . . . . .	7
2.3.1 Ukrainian Dictionaries . . . . .	8
<b>3 Related Work</b>	<b>10</b>
3.1 WordNet Development . . . . .	10
3.1.1 Ukrainian WordNet . . . . .	11
3.2 Generative Artificial Intelligence . . . . .	12
3.2.1 Large Language Models . . . . .	13
3.2.2 Prompt engineering . . . . .	14
<b>4 Research Gap</b>	<b>17</b>
<b>5 Proposed Solution</b>	<b>18</b>
5.1 Data Collection . . . . .	18
5.1.1 Ukrainian WordNet Basis . . . . .	18
5.1.2 Gloss Translation . . . . .	18
5.1.3 Dictionary Definitions . . . . .	18
5.2 Prompt Selection . . . . .	19
5.2.1 Test Dataset . . . . .	19
5.2.2 Prompts . . . . .	19
5.2.3 Evaluation . . . . .	19
5.3 Gloss Generation . . . . .	19
5.4 Evaluation . . . . .	19
<b>6 Experiments and Results</b>	<b>21</b>
6.1 WordNet Basis . . . . .	21
6.2 Prompt Evaluation . . . . .	21
6.3 Gloss Generation . . . . .	23

6.4	Results Evaluation . . . . .	23
<b>7</b>	<b>Conclusions</b>	<b>27</b>
7.1	Contribution . . . . .	27
7.2	Limitations . . . . .	27
7.3	Future Work . . . . .	27
<b>A</b>	<b>Propmts for Gloss Generation</b>	<b>29</b>
A.1	Zero-Shot Prompt . . . . .	29
A.2	One-Shot Prompt . . . . .	29
A.3	Zero-Shot Chain-of-Thought Prompt . . . . .	29
A.4	One-Shot Chain-of-Thought Prompt . . . . .	30
A.5	Guideline Prompt . . . . .	30
<b>B</b>	<b>Referenced Prompts</b>	<b>32</b>
B.1	Tree-of-Thought Prompting Sample[17] . . . . .	32
	<b>Bibliography</b>	<b>33</b>

# List of Figures

2.1	Example entry in Princeton WordNet 3.1. . . . .	4
2.2	Example of a WordNet noun tree. . . . .	5
2.3	Example of a WordNet adjective cluster. . . . .	6
2.4	Example of a WordNet verb tree. . . . .	7
2.5	Example of different senses and corresponding glosses for a word "lie" in PWN3.1. . . . .	8
2.6	Example of entry in SUM-20. . . . .	9
3.1	Comparison between the highest level of noun hierarchies of PWN3.0 and WordNet-like dictionary of the Ukrainian language. Adapted from Kulchytsky, Romaniuk, and Khariv[23] . . . . .	11
3.2	Example inputs and outputs of LLM for a few-shot prompting and few-shot prompting combined with chain-of-thought.[64, Figure 1] . . . . .	15
3.3	Example inputs and outputs of GPT-3 with (a) standard Few-shot (b) Few-shot-CoT, (c) standard Zero-shot, and (d) Zero-shot-CoT. [22, Figure 1] . . . . .	16
3.4	Representations of different approaches in prompt engineering. Each rectangle box represents a thought, a coherent language sequence that serves as an intermediate step toward problem-solving. [65, Figure 1] . . . . .	16
6.1	Comparison of glosses generated by different methods and a score of their cosine similarity to dictionary definitions. . . . .	24
6.2	Comparison of glosses generated by different methods and a score of their cosine similarity to dictionary definitions. . . . .	25
6.3	Histogram of semantic similarity scores after evaluation. . . . .	26

# List of Tables

6.1	General statistics related to the initial experiment developing Ukrainian WordNet basis[53]. . . . .	21
6.2	Results of experiment replication. . . . .	21
6.3	Example of generated gloss for word "КЛЮВАТИ"(peck) with different prompts by GPT3.5. . . . .	23
6.4	Example of generated gloss for a word "небо"(sky), original gloss from PWN3.1, different translations and dictionary definition. . . . .	23
6.5	General statistics of semantic similarity scores between different glosses and dictionary definitions. . . . .	24



# List of Abbreviations

<b>AI</b>	<b>Artificial Intelligence</b>
<b>GenAI</b>	<b>Generative Artificial Intelligence</b>
<b>LLM</b>	<b>Large Language Model</b>
<b>OEWN</b>	<b>Open English WordNet</b>
<b>PWN</b>	<b>Princeton WordNet</b>

## Chapter 1

# Introduction

### 1.1 Motivation

WordNet is a lexical database of semantic relations between words, structured as a network of interconnected concepts. WordNet is a valuable resource for various applications in computational linguistics and natural language processing (NLP), such as word sense disambiguation, information structuring and classification, machine translation, and automatic text summarization[37].

The development of WordNets for various languages worldwide was inspired by the original Princeton WordNet, which began as a psychological experiment on how the human mind organizes lexical knowledge [35]. However, the manual creation of WordNets is a time-consuming and resource-intensive task. Therefore, much research has focused on developing automatic methods for building and expanding WordNets.

Several studies showed the effectiveness of automated approaches in rapidly expanding lexical resources. Some of the many examples of WordNets built using automatic methods include Japanese[5], Uzbek[1], Estonian[46].

While significant progress has been made in developing WordNets for various languages, Ukrainian still lacks a comprehensive lexical database. Previously, a foundational Ukrainian WordNet was established[53]. The goal of this thesis is essentially a continuation and expansion of earlier research efforts by using large language models to enrich the existing Ukrainian WordNet with comprehensive glosses. This thesis aims to enhance the completeness and utility of the Ukrainian WordNet, thereby addressing the gaps in linguistic resources for the Ukrainian language.

### 1.2 Contributions

Research in this thesis was fueled by the lack of open WordNet for Ukrainian language. In recent years different steps have been taken in order to solve this problem. WordNet is a complex and powerful tool in the field of natural language processing, and requires a lot of resources to be built. The aim of this work is to expand current Ukrainian WordNet basis by adding glosses to synsets.

Another goal of the thesis is to explore capabilities of large language models and test them in the previously underexplored area. This study tests cutting-edge technology and uses it further develop an important lexical resource for the Ukrainian language.

### 1.3 Structure of the Thesis

The remaining work is divided into the following chapters.

In Chapter 2, we review the main concepts and terms used in this paper: what is a WordNet and its structure, what is gloss in a WordNet, and give primary information on Ukrainian explanatory dictionaries and word definitions in them.

Chapter 3 discusses approaches used in building WordNets for other languages, summarizes work done for creating Ukrainian WordNet, gives the overview of the most popular LLMs, explains prompt engineering, and describes some methods.

Chapter 4 outlines areas that are missing research.

In Chapter 5, we explain the problem solution pipeline, describe how the data for experiments is generated, and establish metrics and evaluation methods.

Chapter 6 provides an overview of the datasets generated following the methodology outlined in the previous section. It introduces the primary experiments conducted for the gloss generation for Ukrainian WordNet. Additionally, we discuss and analyze the results obtained from these experiments.

Chapter 7 summarizes the work done and suggests directions for future work.

## Chapter 2

# Theoretical Background

## 2.1 WordNet

WordNet is a lexical resource in natural language processing and computational linguistics. It represents a comprehensive lexical database of a language organized as a network of semantic relations between words. This network structure captures the hierarchical and associative relationships among words, providing valuable help for various natural language processing tasks like word-sense disambiguation[41], text classification and summarization, machine translation, etc.

WordNet organizes words into synsets, a set of synonymous words that represent a single concept or meaning. These synsets are interconnected through various semantic relations; they enable WordNet to encode the meanings of individual words and the intricate semantic connections between them, thus facilitating a deeper understanding of language semantics.

One of the critical strengths of WordNet lies in its ability to capture the different meanings of a word. WordNet provides a structured representation of word meanings by enumerating these senses and organizing them into distinct synsets, thereby aiding in tasks requiring precise semantic analysis and disambiguation.

The first and the most popular WordNet is a Princeton WordNet<sup>1</sup>. It started in the mid-1980s at a Cognitive Science Laboratory at Princeton University as a test on psycholinguistic theories on how humans use and understand words[11]. Today, Princeton WordNet stands as one of the most widely used resources in computational linguistics, providing valuable insights into the structure and organization of language. Its creation represents a milestone in lexical semantics, advancing the understanding of human language.

### 2.1.1 WordNet Structure

WordNet was designed as a network because this is how native speakers organize their mental lexicons[12]. The structure contains lexical units that are linked by various semantic relations. WordNet consists of four components: nouns, verbs, adjectives, and adverbs.

At the core of WordNet's structure are **synsets**, groups of words that share a common meaning or concept. Each synset represents a distinct sense or semantic interpretation of a word, capturing its various nuances and usages. For example, the word *bank* in WordNet is associated with multiple synsets representing different meanings, such as a financial institution, the land alongside a body of water, or the slope of a hill.

Additionally, each synset within WordNet contains a brief definition, a **gloss**, which describes the meaning or concept represented by the synset. Gloss gives

---

<sup>1</sup><https://wordnet.princeton.edu/>

information about the semantic interpretation associated with the synset, offering users a quick reference point for understanding its significance. Furthermore, many synsets include one or more example sentences that illustrate the use of the words in context.

In WordNet, the senses of words are primarily sorted based on their frequency of usage. The most common meaning of a word is typically listed first, followed by less common and more specialized senses. This arrangement helps users understand the most likely context in which a word is used, aligning with the principle that the most typical usage of a word should be the easiest to access (See Figure 2.5).

The general structure of Princeton WordNet entry is illustrated in Figure 2.1.

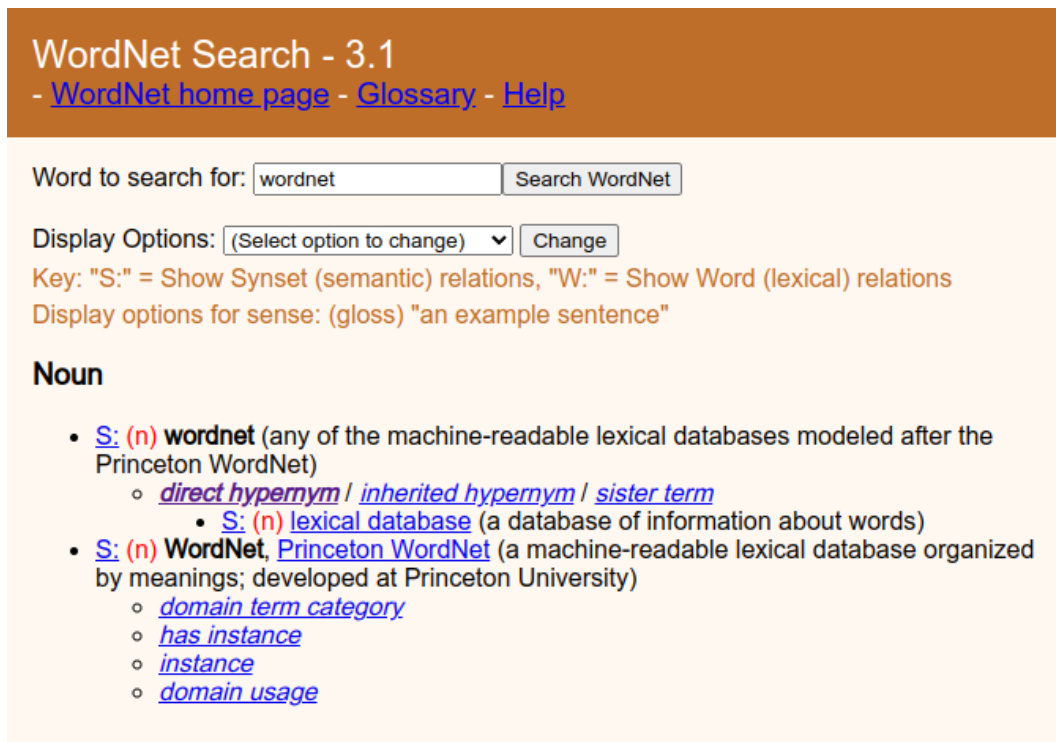


FIGURE 2.1: Example entry in Princeton WordNet 3.1.

## Nouns

Noun synsets in WordNet are interconnected through various relations that capture different aspects of lexical semantics. Some of the key semantic relations include:

- **Hypernymy and hyponymy (is-a relation):** Describes the relationship between a general concept (hypernym) and a specific instance of that concept (hyponym). For instance, *dog* is a hyponym of *canine*, and *canine* is a hypernym of *dog*.
- **Meronymy and holonymy (part-of relation):** Meronymy and holonymy (part-of relation): Indicates the part-whole relationship between lexical units. For example, *arm* is a meronym of *body*, and *body* is a holonym of a *arm*.

Hyponymy builds hierarchical trees with increasingly specific leaf concepts growing from an abstract root[10]. All noun hierarchies ultimately go up the root node *entity* (See Figure 2.2).

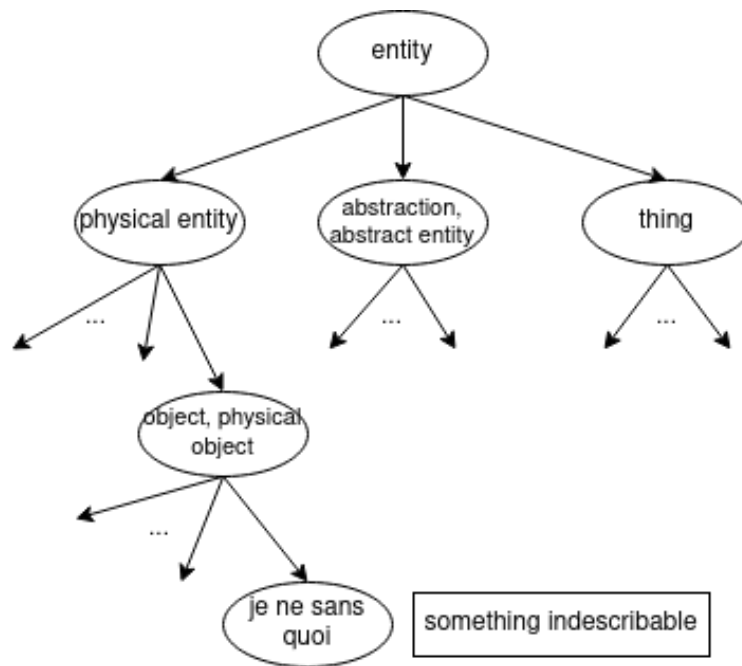


FIGURE 2.2: Example of a WordNet noun tree.

### Adjectives and Adverbs

WordNet separates adjectives into two groups: descriptive and relative. Descriptive adjectives are organized into direct antonym pairs, such as *wet-dry* and *cheap-expensive*[10]. Direct antonym pairs fill a small part of the WordNet lexicon. Many more adjectives are classified in WordNet as "semantically similar" to the members of the direct antonym pairs[11]. For example, words *sloppy*, *damp*, *misty*, *rainy* are semantically similar to a word *wet*, so they are indirect antonyms to a word *dry*. Likewise, words *rainless*, *arid*, *thirsty* are semantically similar to a word *dry*, and *wet* is their indirect antonym.

Descriptive adjectives in WordNet are structured in clusters, where each cluster centers around a direct antonym pair, and semantically similar adjectives encircle each adjective from a pair. These surrounding adjectives serve as the indirect antonyms of the word found on the opposite side of a pair (See Figure 2.3).

Relational adjectives like *atomic* and *industrial* can not be linked using antonymy. These adjectives are linked to nouns to which they correspond semantically and morphologically (*atomic* and *industry*, respectively)[10].

Adverbs in WordNet are primarily formed through the addition of the *-ly* suffix to their related adjectives. These adverbs are often associated with antonymous counterparts, mirroring the organization of their originating adjectives. However, for lexical adverbs such as *hard* and *even*, no specific organizational structure has been established[10].

### Verbs

Verbs are primarily organized by a troponymy relation. It connects two verbs such that one verb specifies a particular manner of carrying out the action referred to by the other verb[11]. For example, *orate*, *dogmatize*, *converse*, *chat up* are troponyms of a verb *talk*. Like hyponymy, troponymy builds hierarchies of several levels of

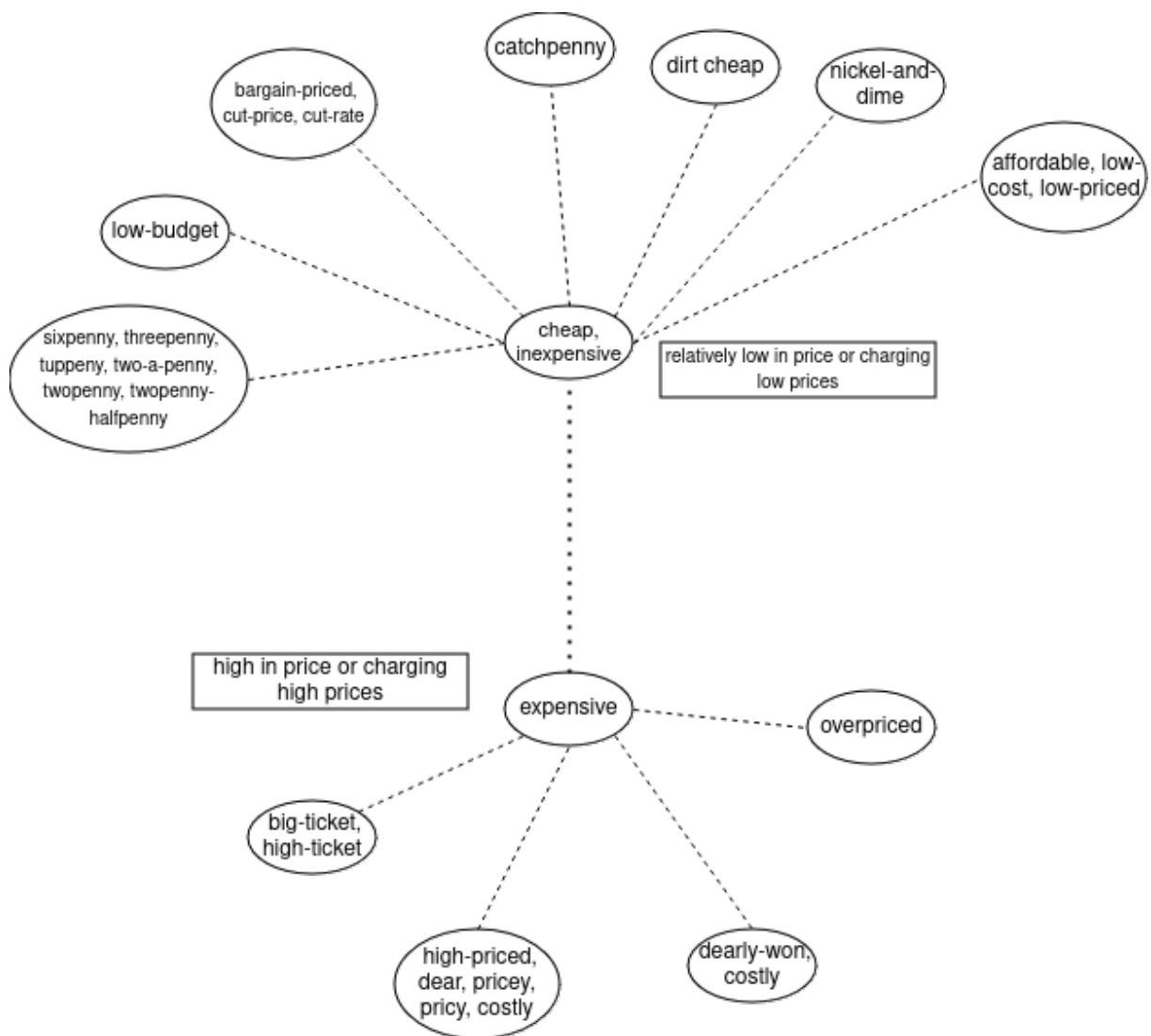


FIGURE 2.3: Example of a WordNet adjective cluster.

specificity[10]. But verb trees(See Figure 2.4) "grow" in breadth rather than in-depth like nouns.

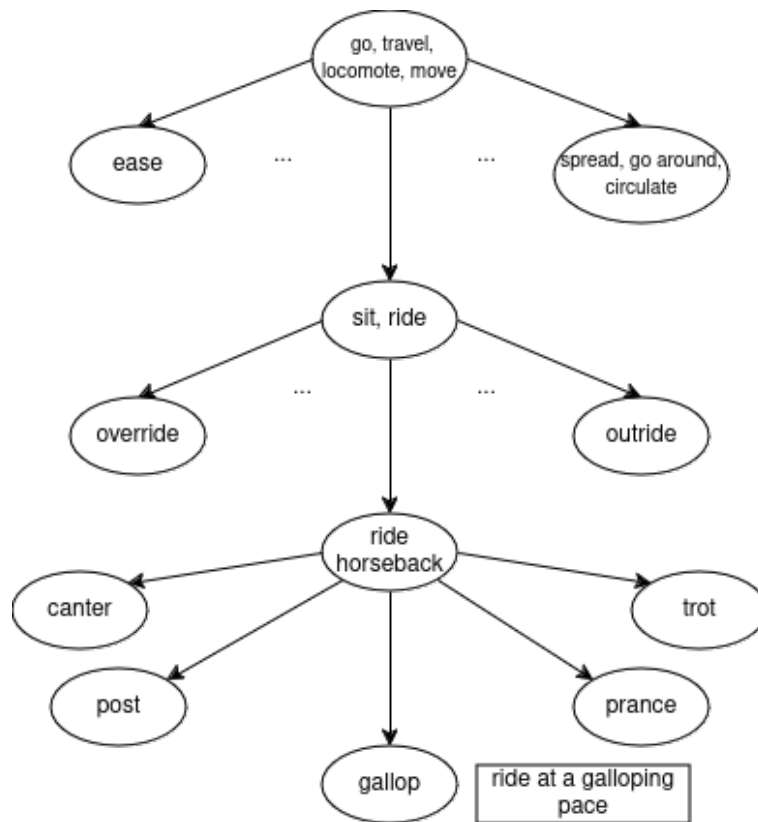


FIGURE 2.4: Example of a WordNet verb tree.

## 2.2 Glosses in WordNet

As mentioned in Section 2.1.1, WordNet glosses are brief descriptions of the meaning of lemmas found in the corresponding synset, aiming to provide concise explanations of the concept represented in the synset. Typically consisting of a single sentence, these glosses define or describe the basic sense of a word, following a standardized format within the database (See Figure 2.5).

Glosses were not initially included in the PWN, and were introduced around 1989 as number of words in WordNet increased and differentiate between different senses only by synonyms and semantic relations became complicated[51]. Glosses are essential for various natural language processing tasks, such as text mining and classification[42], information retrieval and sense disambiguation[32]. Glosses play a crucial role in understanding word meanings. They also serve as a resource for language learning, providing learners with clear and concise definitions of words, and helping in language acquisition and comprehension.

## 2.3 Definitions in Explanatory Dictionary

When trying to understand the meaning of a word, people often first consider looking up its definition in a dictionary. Explanatory dictionaries are the primary source of glosses for all possible word senses.



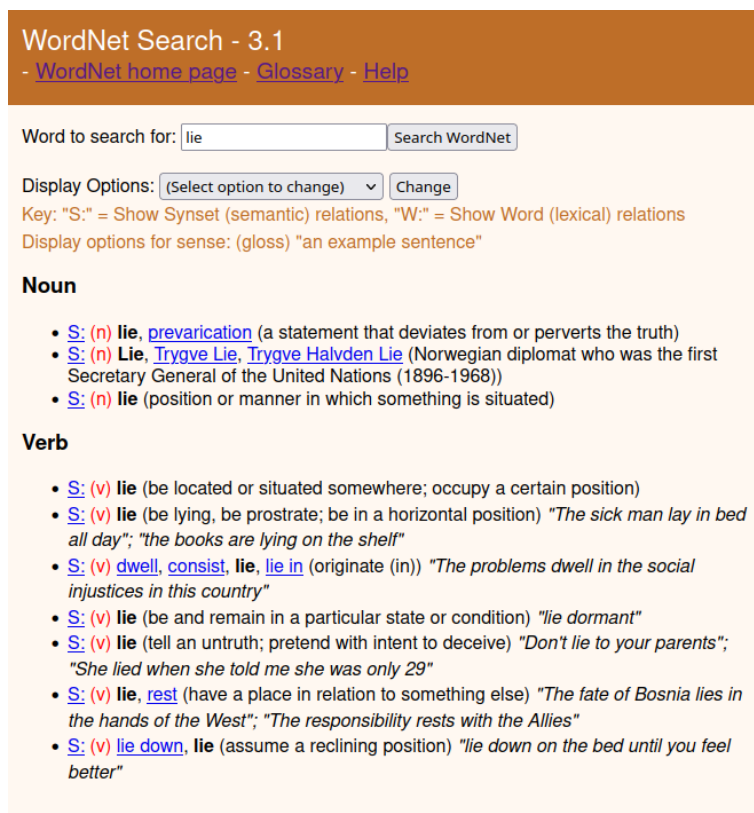


FIGURE 2.5: Example of different senses and corresponding glosses for a word "lie" in PWN3.1.

In an explanatory dictionary, each entry focuses on a specific word or phrase, starting with its definition and pronunciation. This is followed by a classification into parts of speech and detailed explanations of its meanings across different contexts. Definitions are often organized according to their frequency of usage (See Figure 2.6), similar to a gloss in WordNet. Additionally, usage examples are provided to demonstrate how the word or phrase is used in sentences, enhancing understanding.

Optionally, entries may include elements like etymology, synonyms, and antonyms. Additionally, there might be usage notes, spelling variations, and regional variations. In some dictionaries, entries feature illustrations or images to help the reader understand the word. Cross-references to related words or phrases are sometimes provided.

### 2.3.1 Ukrainian Dictionaries

"Словник української мови" (SUM)<sup>2</sup> is the first explanatory dictionary of Ukrainian language that consists of 11 volumes with the total of 134,058 entries. It was published in the 1970s and is now used as a base for most modern Ukrainian dictionaries and thesauri. The main criticism about SUM is that some entries are outdated, have incorrect spelling, or propagate political bias, originating from the historical time the dictionary was created.

<sup>2</sup><https://sum.in.ua/>

In 2010, The National Academy of Sciences of Ukraine started publishing "Тлумачний словник української мови у 20 томах"<sup>3</sup>(SUM-20)<sup>3</sup>, a successor of SUM. At the time of writing this thesis, 14 volumes out of the planned 20 have been published. The goals of SUM-20 creation are to release an updated Ukrainian explanatory dictionary with fixed mistakes of SUM, to replace examples from the Soviet times with more relevant ones, and to extend the vocabulary, including modern terms. The on-line versions of SUM and SUM-20 are made primarily for human usage and are not available in machine-readable format.

<p><b>БАЗАР</b>, <i>у, ч.</i> 1. Торгівля (перев. приватна) продуктами харчування або предметами широкого вжитку (часто в певні дні тижня) на площі або у спеціально збудованому приміщенні. <i>В неділю, в місто на базар Жінки возили свій товар</i> (І. Нехода); // Місце, де відбувається така торгівля. <i>Вже їй можна, коли захотіла, і по базару самій походити</i> (Г. Квітка-Основ'яненко); <i>Утоптала стежечку Через яр, Через гору, серденько, На базар</i> (Т. Шевченко); <i>Тепер базар був мертвий. Огидливо стирчали гнилими кроквами напівзруйновані ятки</i> (Б. Антоненко-Давидович); // Час такої торгівлі. – <i>У базар то й чоловіка попадеш, бо в базар людей найбільше</i> (Марко Вовчок); // <i>у знач. присл., кого, перен., розм. Дуже багато. Базар люду насходилося</i> (Т. Шевченко).</p> <p>2. Організована торгівля спеціальними товарами в певні сезони або періоди найбільшого попиту на них. <i>Книжковий базар; Шкільний базар.</i></p> <p>3. <i>перен., розм.</i> Шум, гам, голосні безладні розмови. <i>Хата з дітьми – базар</i></p>	<p>(прислів'я); <i>А я не люблю, коли на будівельному майданчику базар і смітник</i> (В. Кучер).</p> <p>(1) <b>Блошиний базар (рінок)</b> – місце, де продають старі, використані речі, а також валюту, зброю і т. ін. <i>Я люблю блошині базари, бо це, як на мене, унаочнене трепетання буттєвих субстанцій</i> (Ю. Андрухович); <i>На блошиному ринку можна придбати пап'є-маше, дерево, пластик, тканини і мережива для вбрань, нерідко антикваріат</i> (з газ.); (2) <b>Пташиний базар (рінок)</b>: а) <i>(тільки базар)</i> місце на морському березі, на острові, на скелях, де оселяються величезною масою птахи. <i>Хоча цей скелястий, похмурий, майже без дерев острів і звався пташиним базаром, проте там ніхто нічого не продавав і не купував ... Просто там завжди збиралася величезна кількість різних птахів</i> (О. Іваненко); б) місце, де продають тварин. <i>Якось я пішов на пташиний базар і купив своїй коханій маленького спанієля</i> (із журн.); <i>На пташиному ринку в ряд продавців кошенят та щенят затесався якийсь молодик, котрий продавав їжаченя</i> (з газ.).</p>
--	---

FIGURE 2.6: Example of entry in SUM-20.

Wiktionary<sup>4</sup> is a crowdsourcing project that aims to create dictionaries with descriptions of all words in all languages. It is free and allows to download data in a machine-readable format, making it very suitable for NLP tasks. Ukrainian Wiktionary<sup>5</sup> currently contains 56,057 entries, which is 2.4 times less than SUM. It is important to note that, unlike SUM and SUM-20, this dictionary was not compiled by lexicographers and thus might contain mistakes or inaccuracies.

"СЛОВНИК.ua"<sup>6</sup> is another dictionary that combines entries from SUM (over 130,000) and manually added by users (over 21,000). The data is not available in the machine-readable format. As in the case of Wiktionary, the collaborative part still needs to be reviewed by experts.

<sup>3</sup><https://sum20ua.com>

<sup>4</sup><https://www.wiktionary.org/>

<sup>5</sup>[https://uk.wiktionary.org/wiki/БЗД;ДЗД;ДЗД;ДЗ\\_НАЊЊД;НАЊЊД;ДЗД](https://uk.wiktionary.org/wiki/БЗД;ДЗД;ДЗД;ДЗ_НАЊЊД;НАЊЊД;ДЗД)

<sup>6</sup><https://slovnyk.ua/index.php>

## Chapter 3

# Related Work

### 3.1 WordNet Development

Princeton WordNet(PWN) stands as the first and by far most extensively developed WordNet for the English language[34]. This structured resource is used for various natural language processing tasks, such as word sense disambiguation[41], metaphor detection and interpretation[28], natural language generation[19]. The WordNet structure enables the exploration of semantic relationships, enhancing the ability of computational algorithms to process textual data effectively. It is also integrated into popular NLP libraries like NLTK and spaCy, which help use it in linguistics, computer science research, and problem solutions.

PWN contains over 117,000 synonym sets and various relations and serves as a standard reference for other WordNets. Its success has inspired the creation of WordNets for other languages. There are two main approaches to building a WordNet from scratch: merge and expand[63]. Merge approach involves collecting words from monolingual dictionaries, corpus, etc., and creating a lexical network that is specific for a language that is being studied, and in the finishing stages of research linking it to the existing resource, such as PWN[38]. The expand approach uses a translation of PWN into a target language, and as a result, new WordNet gets a structure similar to PWN[6].

The expand approach is more accessible because it is less time-consuming and does not require as much linguistic knowledge as the merge method. But WordNet created by the expand approach might lack terms unique to that specific language and not present in the PWN. WordNet created by merge approach will be more accurate as it is not influenced by the English language structure[49].

There are multiple examples of WordNets created by each method. Hungarian WordNet was created using an expand approach based on PWN 2.0[62]. Thai WordNet was developed by a semi-automatic method based on PWN[57]. Batsuren et al. used expert, monosemy, and hypernym-based translations to build a Mongolian WordNet from PWN[4]. The EuroWordNet project<sup>1</sup> contains WordNets for several European languages (Dutch, Italian, Spanish, German, French, Czech and Estonian)[63], merge approach was used for the most languages in it[2]. Polish WordNet<sup>2</sup> was also built by merge approach[48].

Open English WordNet(OEWN)<sup>3</sup>[29] is another valuable resource for the English language; it is a fork of a PWN and is developed as an open-source project

<sup>1</sup><https://archive.illc.uva.nl/EuroWordNet/>

<sup>2</sup><http://plwordnet.pwr.edu.pl/wordnet>

<sup>3</sup><https://en-word.net/>

on GitHub<sup>4</sup> and can take contributions from the community of users. The aim behind OEWN creation is to add neologisms, fix spelling mistakes, and change language use cases that PWN currently lacks and to produce the highest quality and most complete wordnet for English and to do so in an open manner[29]. The project and its open methodology have been successful so far, and in the 2020 year, OEWN contained over 120,000 synsets, and it is still growing[30].

### 3.1.1 Ukrainian WordNet

Research on constructing the Ukrainian WordNet began in the 2010s as Kulchytsky, Romaniuk, and Khariv[23] initiated this effort by analyzing noun relationships in the Princeton WordNet. The initial part of the research involved identifying core nouns in the Ukrainian language. To solve this problem, authors used frequency dictionaries to identify the most frequent nouns in texts from different literature genres and time periods. Core nouns later were structured hierarchically (See Figure 3.1), resulting in a WordNet-like dictionary comprising 194 synsets. Synsets were interconnected predominantly through hypo-hypernymy, antonymy, and meronymy/holonymy. As the process could not be automated, the research was not continued beyond this initial stage, and the results were inaccessible to the public.

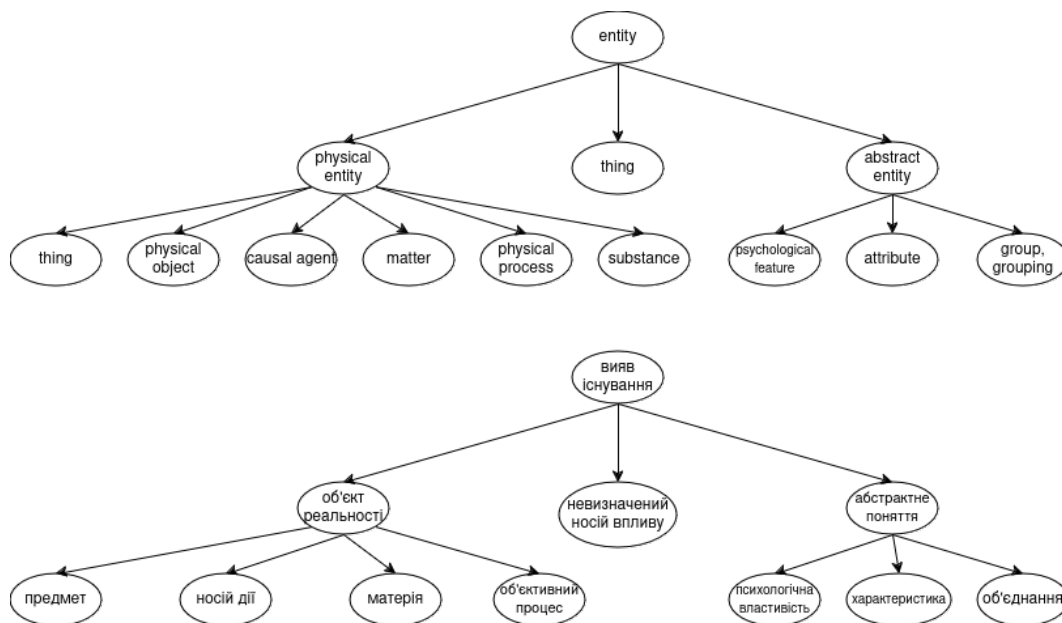


FIGURE 3.1: Comparison between the highest level of noun hierarchies of PWN3.0 and WordNet-like dictionary of the Ukrainian language. Adapted from Kulchytsky, Romaniuk, and Khariv[23]

Anisimov et al.[3] contributed by developing UkrWordNet, a lexical semantic database for the Ukrainian language. Their approach involved automated techniques to expand UkrWordNet by generating nodes from Ukrainian Wikipedia articles and associating them with synsets. They introduced a novel semantic similarity measure to enhance associations' quality. Despite growing UkrWordNet to over 82,000 synsets and approximately 145,000 nouns, its unavailability to the public remains a significant limitation.

<sup>4</sup><https://github.com/globalwordnet/english-wordnet>

There have been efforts, in the form of student theses from Lviv Polytechnic National University<sup>5</sup>, but their contributions were limited.

Additionally, in 2023, Siegel et al.[56] attempted to create Ukrainian WordNet covering the physics domain. They developed and released Ukrajinet 1.0<sup>6</sup> that contains 3,360 synsets and 8,700 nouns. However, as the authors stated, this is not the final stage, and the project is considered a work in progress.

The next significant work towards the creation of Ukrainian WordNet was made by Romanyshyn[53] with a novel data-driven methodology for automating the construction of hypernym hierarchies within the Ukrainian WordNet, contributing significantly to the field of Ukrainian natural language processing. The approach establishes a robust framework for creating this new WordNet resource by leveraging data from PWN, Wikidata, and Wikipedia. However, not all PWN nodes can be mapped into entries of Ukrainian words, creating gaps between nodes in Ukrainian WordNet. Therefore, a Gap Ranking algorithm is proposed to identify optimal nodes for gap filling. Various techniques are used to generate potential candidates to fill the gaps, including methods utilizing information about missing nodes and their respective children within the hierarchy.

Also, the work explores the efficiency of LLMs in the Hypernym Discovery task. This is demonstrated by showcasing the process of constructing a comprehensive set of instructions from initially limited data and fine-tuning LLMs for specialized hypernym suggestion tasks, similar to a chatbot assistant.

A user-friendly tool has been developed to manually annotate identified candidates for gap-filling within the Ukrainian WordNet. This tool will undergo further refinement to cater to the specific needs of lexicographers.

## 3.2 Generative Artificial Intelligence

The development of generative artificial intelligence(GenAI) models, such as ChatGPT[45], has revolutionized the NLP field, offering novel capabilities in problems that require generating text and understanding human language. Among the most prominent developments in this domain are Large Language Models (LLMs), which have garnered significant attention for their remarkable ability to understand, generate, and manipulate human-like text.

GenAI models can produce original content, distinguishing themselves from previous expert systems that primarily analyze or utilize existing data[14]. GenAI models operate through user interaction, utilizing techniques like prompting, where users engage with the model through natural language to create desired outputs. Their outputs are probabilistic and vary with each prompt, sometimes requiring users to iterate on prompts to achieve desired outcomes.

GenAI users continuously specify tasks as input prompts until they get the desired output. Unlike discriminative AI, which analyzes data to make decisions, GenAI leverages large datasets to produce diverse content based on user input. Thus, the primary difference lies in the role of data, with GenAI focusing on content generation and discriminative AI in decision-making. GenAI is emerging as a transformative tool with various applications across different industries. It enables the creation of unique content, going beyond mere assistance. GenAI operates across different styles, generating text, images, videos, code, etc. Its architecture consists of

---

<sup>5</sup><https://github.com/lang-uk/wordnet/tree/main/resources>

<sup>6</sup><https://github.com/hdaSprachtechnologie/ukrajinet>

model, connection, and application layers, enabling the integration of GenAI into information systems[27].

### 3.2.1 Large Language Models

Language modeling(LM) has been researched since the 1950s, initially explored by Shannon in the context of n-grams and information theory[55], and now is one of the primary methods for enhancing machine language intelligence. The goal of LM is to determine the probability of a given sequence of words occurring in a sentence and be able to predict future or missing words[67]. Studies on LM represent decades of work and research. **Statistical language models(SLMs)** relied on statistical methods that were popular in 1990s to analyze the structure and patterns of natural language[67]. They operated by calculating the probability of a word or sequence of words occurring within a given context, often utilizing techniques such as n-grams or Markov models. **Neural language models(NLMs)**: address the challenge of data sparsity by introducing embedding vectors[33], and using neural networks models were able to predict the next word by aggregating embedding vectors of previous words. **Pre-trained language models(PLMs)** were trained on unlabeled text corpora and then fine-tuned on smaller datasets of labeled data for solving specific tasks.

The three generations of language models have demonstrated varying problem-solving capacity levels. Initially, SLMs primarily aided specific tasks, such as retrieval or speech tasks[18], by leveraging predicted or estimated probabilities to improve task-specific approaches. Subsequently, NLMs prioritized learning task-agnostic representations, thereby aiming to reduce the need for human feature engineering. Furthermore, PLMs acquired context-aware representations optimized for downstream tasks [67].

The latest generation, **Large language models(LLMs)** became general-purpose task solvers. When deep neural network architecture, called transformer[60] was introduced in 2017, it received much attention and became a powerful tool for NLP researchers. LLMs are usually transformer-based models pre-trained on massive text data and exhibit superior language understanding and generation abilities, along with emergent abilities like in-context learning, instruction following, and multi-step reasoning[36].

The three most popular LLM families are GPT, LLaMA, and PaLM. As this research focuses on work with the GPT model, we will not perform any deep analysis of models from the last two families, but here is a quick overview of both of them:

- **LLaMA** is a collection of LLMs of different sizes(7B, 13B, 33B, and 65B) released by Meta in the February 2023[58]. LLaMA was trained on data from publicly available datasets, which are open-source and free for research. Because of this, LLaMA has grown rapidly as numerous research groups have utilized these models extensively. They use them to enhance existing open-source language models, to rival closed-source alternatives, or to tailor language models for specific tasks in different applications[36]. This family's latest and most effective model to this day, LLaMA-3[31], was launched in April 2024.
- **PaLM(Pathways Language Model)** is an LLM developed by Google and has 540B parameters. For the first time, it was introduced[40] in April of 2022 and became publically available through API in March of 2023[61].

The **GPT(Generative Pre-trained Transformers)** family consists of several LLMs developed by OpenAI. Early models like GPT-1[52] and GPT-2[43] that were launched



in June 2018 and February 2019, respectively, are open-source, newer ones like GPT-3[7] and GPT-4[45] are close-source and are accessible only via APIs. GPT-3, with 175 billion parameters, shows outstanding abilities such as in-context learning, enabling strong performance across various NLP tasks. CODEX[66], a programming model released in March 2023, generates code from natural language and powers GitHub Copilot. WebGPT[39] is a fine-tuned GPT-3 that answers open-ended questions using a text-based web browser. InstructGPT[47] aligns models with user intent through fine-tuning with human feedback. ChatGPT, launched in November 2022, is a versatile chatbot powered by GPT-3.5 (and GPT-4 for the paid version), adept at tasks like question answering and text summarization. GPT-4, launched in March 2023, is currently the most powerful model from the whole family and shows outstanding results on the evaluation benchmarks[36].

### 3.2.2 Prompt engineering

Although LLMs are extremely powerful, they need human directions on what to do. They are competent, but they need clear instructions to do well. It is vital to instruct LLM properly, and this is part where prompt engineering becomes essential. Prompt engineering is a process of structuring text to be understood and interpreted correctly by an LLM[9].

According to guidelines<sup>7</sup> written in collaboration with OpenAI, one should follow two main principles to write a good prompt: to write clear and specific instructions, and to give the model time to "think."

Different tactics help to get a prompt that satisfies the first principle. One of them is to use delimiters like quotations, XML tags, section titles, etc., that help LLM understand what part of the prompt is a request and what part of a text it has to work with. Delimiters can also help with a prompt injection, which can happen when a user adds some input to the prompt that can create conflicting instructions with the actual request. Another tactic is to ask for a structured output, such as HTML or JSON.

The third tactic is called "few-shot" prompting. Contrary to a standard zero-shot, where LLM receives only a task that must be solved, a few-shots give a model example of a completed task before asking it to solve a similar one. The number of given examples can vary depending on the specific task. The one-shot is identical to the few-shots except that the model gets only one demonstration alongside a task to complete[7].

The essence of the second principle of prompt engineering is to avoid mistakes the model makes while rushing to incorrect solutions. When the problem the model is trying to solve is too complex, it will likely produce inaccurate results. The tactic is to specify the required steps for reaching the optimal solution. If instead of one big problem that LLM has to solve, there are a few more minor steps, then the model is less likely to take a wrong turn and will produce the correct result.

One technique used to divide one big task into a bunch of smaller tasks is called chain-of-thought prompting[64], which mirrors the human cognitive process of breaking down complex reasoning tasks into intermediate steps to get to the final solution. The effect of chain-of-thought can be achieved by adding the phrase "Think carefully and explain your answer" or a similar one to a prompt[17]. Chain-of-thought also can be combined with a few-shot prompting technique, especially for problems that require reasoning. According to Wei et al., chain-of-thought prompting presents several advantages. It enables models to break down multi-step problems, allowing for

<sup>7</sup><https://learn.deeplearning.ai/courses/chatgpt-prompt-eng/lesson/2/guidelines>

the allocation of additional computational resources for complex reasoning tasks. This method offers interpretable insights into the model's behavior, understanding how it arrived at a particular answer and identifying potential errors in the reasoning process. Moreover, it is versatile and applicable to commonsense reasoning tasks, mirroring human problem-solving abilities[64]. Figure 3.2 shows how chain-of-thought makes a difference to a prompt created by a simple few-shot prompting technique.

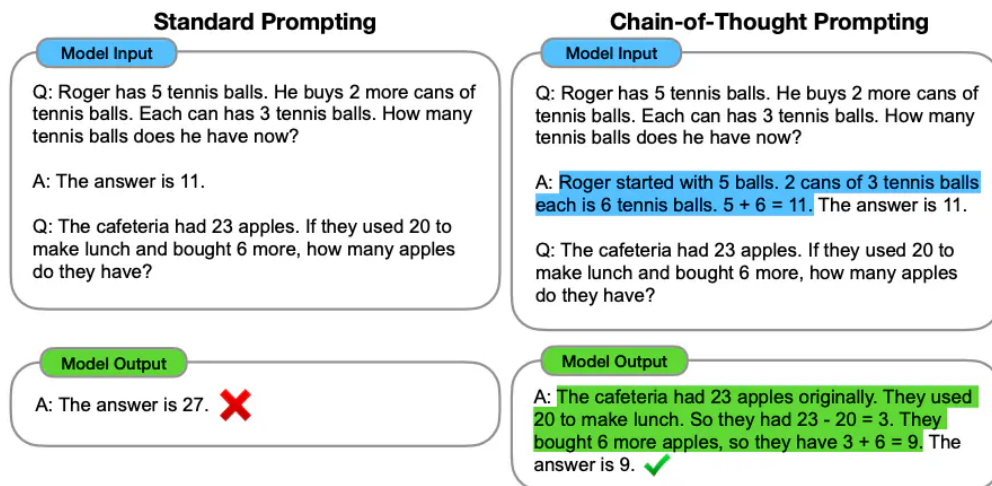


FIGURE 3.2: Example inputs and outputs of LLM for a few-shot prompting and few-shot prompting combined with chain-of-thought.[64, Figure 1]

Kojima et al. proposed an alternative zero-shot chain-of-thought. The main idea of this practice is to add "Let's think step by step" or another similar phrase to make a model to return a step-by-step solution to a given problem[22]. Comparison between zero-shot-CoT LLM's inputs and outputs to different prompt engineering strategies can be seen in Figure 3.3.

Another concept introduced by Yao et al., similar to chain-of-thought, is tree-of-thought; it also offers a problem-solving approach for LLMs inspired by human problem-solving processes. It frames problem-solving as a search over a tree structure where nodes represent partial solutions, allowing exploration of multiple reasoning paths[65]. To compare tree-of-thought to other methods and to see its structure and path to a final answer, refer to Figure 3.4.

Based on the tree-of-thought framework, Hulbert described a tree-of-thought prompting, a technique that incorporates ideas behind the tree-of-thought but uses it in one single prompt. The author proposed the prompt (See Appendix B.1) and claims that it has promising results and is more effective than chain-of-thoughts prompting[17].





FIGURE 3.3: Example inputs and outputs of GPT-3 with (a) standard Few-shot (b) Few-shot-CoT, (c) standard Zero-shot, and (d) Zero-shot-CoT. [22, Figure 1]

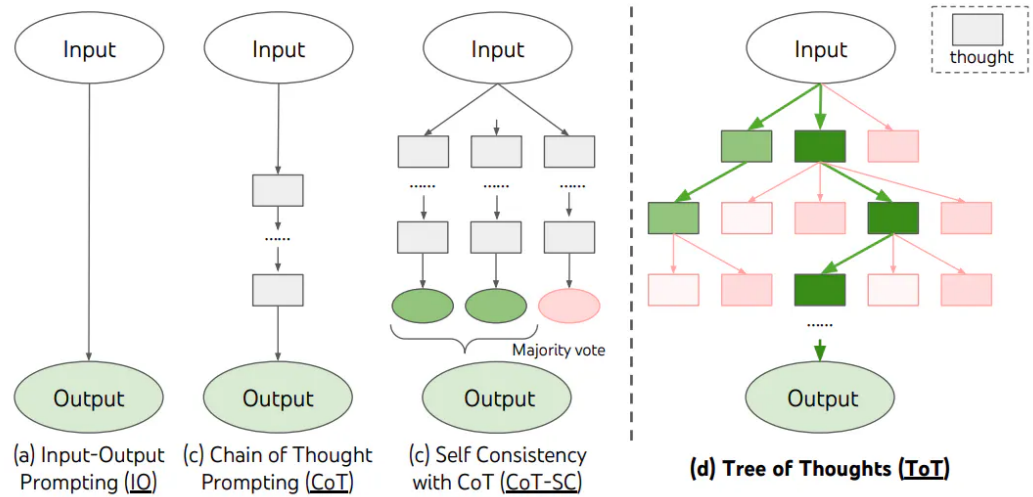


FIGURE 3.4: Representations of different approaches in prompt engineering. Each rectangle box represents a thought, a coherent language sequence that serves as an intermediate step toward problem-solving. [65, Figure 1]

## Chapter 4

# Research Gap

Despite significant advancements in automated methods for WordNet generation, described in Chapter 3, there is a notable gap in utilizing GenAI models for this purpose. While previous studies have explored various automated techniques, there is limited research on using LLMs for building or expanding WordNets. This presents an opportunity to explore the potential of LLMs in enhancing the efficiency and accuracy of WordNet construction.

Moreover, the absence of an open Ukrainian WordNet further underscores the need for novel approaches in lexical resource development. Existing WordNet generation methods may not fully address the linguistic nuances and complexities unique to the Ukrainian language. By employing LLMs, which can capture unique linguistic patterns and semantics, it becomes possible to bridge this gap and create a comprehensive Ukrainian WordNet.

Furthermore, it is essential to recognize that each language encapsulates its distinct cultural and historical context within its lexicon. Traditional approaches to WordNet generation often overlook these cultural nuances, resulting in linguistic resources that may not fully reflect the richness of a language. By leveraging LLMs, which have been trained on vast corpora of text data, it becomes feasible to capture and incorporate cultural and historical elements into the WordNet generation process, thus enriching the quality and depth of the resulting lexical resource.

In summary, utilizing LLMs in WordNet generation represents a novel and underexplored area of research. By addressing this research gap, we can advance the field of computational linguistics and contribute to developing important linguistic resource for the Ukrainian language.

## Chapter 5

# Proposed Solution

The gloss generation process for the Ukrainian WordNet involves utilizing the GPT model with specifically designed prompts. This chapter details the steps involved in the proposed solution.

## 5.1 Data Collection

### 5.1.1 Ukrainian WordNet Basis

The first step in the process is to acquire the Ukrainian WordNet basis by replicating the experiment from the “Data-Driven Approach to Automated Hypernym Hierarchy Construction for the Ukrainian WordNet”[53] paper. The objective is to re-run the code<sup>1</sup> responsible for linking data from PWN, Wikidata, and Ukrainian Wikipedia and subsequently compare the obtained results to results mentioned in the original paper.

In referenced paper[53] the author proposes Python[59] scraper that uses web-crawling framework Scrapy[16], wtf\_wikipedia[21], a Python package for extraction and parsing data from Wikipedia articles, wn library[13] that provides access to the WordNet lexical database, and query language SPARQL[50].

The result of experiment replication is a dataset of Ukrainian synsets, each with its corresponding match from PWN.

### 5.1.2 Gloss Translation

After acquiring the Ukrainian WordNet basis, the next step is to collect machine translations of the original English glosses into Ukrainian. For this we use previously translated PWN3.1 into Ukrainian<sup>2</sup> with Google Translate[20] and Bing[26]. Using its Python API, we also implement gloss translation with DeepL[54].

### 5.1.3 Dictionary Definitions

As previously mentioned in Section 2.3, dictionary definitions can be considered analogs for glosses in the WordNet. In the the proposed experiment pipeline, we use definitions as a benchmark to compare generated Ukrainian glosses and translations of English glosses.

We use definitions from SUM-20. However, as this dictionary is not available in machine-readable format, its data must be parsed beforehand.

---

<sup>1</sup><https://github.com/lang-uk/wikidrill>

<sup>2</sup>[https://github.com/lang-uk/wordnet/tree/main/pwn\\_translated\\_basic](https://github.com/lang-uk/wordnet/tree/main/pwn_translated_basic)

## 5.2 Prompt Selection

Designing effective prompts for the GPT model to generate high-quality glosses is essential. Prompt engineering techniques are employed to create multiple prompts, which are then evaluated to determine the most effective one. There are several ways to build a prompt for generating glosses for Ukrainian WordNet. All prompts constructed with different prompt engineering techniques are run on a smaller dataset. The results are humanly evaluated to choose the best prompt that will be later used on all data.

### 5.2.1 Test Dataset

A test dataset consisting of 100 entries is created for prompt evaluation. Each entry in the dataset includes:

- Ukrainian lemma
- Corresponding lemma from PWN
- English gloss of the lemma
- Different translations of the gloss into Ukrainian

### 5.2.2 Prompts

Several prompts are constructed using strategies described in Section 3.2.2 to determine the best prompting technique.

### 5.2.3 Evaluation

The prompt evaluation aims to select the most effective prompt for gloss generation. In this process, the outputs of all prompts are compared to each other. For each lemma in the dataset, the best-generated gloss is selected. Points are assigned to prompts based on the quality of the generated glosses. If multiple prompts generate the same best gloss for a lemma, all those prompts receive a point. The prompt with the highest overall score, indicating the most consistently high-quality gloss generation across the dataset, is selected as the best-performing prompt.

## 5.3 Gloss Generation

Using the selected prompt, glosses are generated for all lemmas on the Ukrainian WordNet basis.

## 5.4 Evaluation

After generating glosses using a selected prompt, the final step is to evaluate the quality of the generated glosses. This evaluation involves comparing the generated glosses with dictionary definitions and translations.

In this stage, lemmas available in a dictionary are selected for evaluation. Python code is then developed to vectorize the generated gloss, dictionary definition, and translation. Semantic similarity is calculated between the generated gloss, translation and dictionary definition.

The semantic similarity scores provide a quantitative measure of the semantic similarity between the generated glosses and the dictionary definitions and between the translations and the dictionary definitions. Higher similarity scores indicate a better semantic match between the glosses and definitions. By comparing these scores, we can assess the linguistic appropriateness and accuracy of the generated glosses, providing valuable insights into the effectiveness of the gloss generation approach for the Ukrainian WordNet.

## Chapter 6

# Experiments and Results

### 6.1 WordNet Basis

As a result of Romanyshyn’s experiment[53], 17% of the PWN, incorporating 21,015 synsets, were linked to the Ukrainian WordNet framework. From the total number of 127,020 PWN3.1 synsets, 23% of them had established connections to Wikidata. Table 6.1 provides general statistics for the initial experiment.

TABLE 6.1: General statistics related to the initial experiment developing Ukrainian WordNet basis[53].

	PWN3.1	Linked to Wikidata	Linked to Ukrainian Wiki
# of synsets	127,020	29,730	21,015
% of synsets	100%	23%	17%

We were unable to replicate the experiment the same way it was initially conducted because scraping tool<sup>1</sup> used for PWN base used website<sup>2</sup> that is no longer supported now. To solve this problem, we switched to web pages under Open English WordNet domain<sup>3</sup>. As described in Section 3.1, OWN is based on PWN and, therefore, is a suitable substitution for this experiment. After the experiment replication, the number of linked nodes has increased (See Table 6.2).

TABLE 6.2: Results of experiment replication.

	PWN3.1	Linked to Wikidata	Linked to Ukrainian Wiki
# of synsets	126,990	38,776	29,494
% of synsets	100%	30%	23%

### 6.2 Prompt Evaluation

The experiment of prompt selection was conducted on six prompts designed with different techniques mentioned in Section 3.2.2. Prompts were implemented using langchain python library[15](See Appendix A) and openai python library[44] that

<sup>1</sup><https://github.com/lang-uk/wikidrill>

<sup>2</sup><http://wordnet-rdf.princeton.edu/>

<sup>3</sup><https://en-word.net/>

gives access to the API. Here is a description of the prompts that were used in the experiment:

1. **Zero-Shot**: instructs the model to generate a gloss in Ukrainian WordNet for a given lemma, based on the English gloss provided and its translations into Ukrainian. The prompt has the basic structure for generating the output without any additional context or guidance.
2. **One-Shot**: is similar to the zero-shot prompt but includes an example question posed by a researcher. The model is instructed to answer the question consistently without additional output formatting.
3. **Zero-Shot Chain-of-Thought**: is similar to the Zero-Shot prompt, but with the added instruction to think carefully and logically and provide reasoning along with the answer. It also specifies the format for the model output: JSON with keys "reasoning" and "gloss."
4. **One-Shot Chain-of-Thought**: combines two previous prompts, and adds to Zero-Shot Chain-of-Thought prompt an example of successfully generated gloss.
5. **Guideline Prompt**: provides detailed instructions on proposing a variant of Ukrainian gloss for a lemma in Ukrainian WordNet. It guides the model through analyzing the English gloss and its translations, evaluating variants, and creating a grammatically correct gloss in Ukrainian. The model is instructed to provide reasoning and the generated gloss in JSON format.

As was mentioned in Section 5.1, the available data for this stage of the experiment is WordNet basis created by replication of the experiment described in Section 6.1, every entry of this generated dataset contained original English word, its gloss and id from PWN, and corresponding Ukrainian word. Another dataset was translated PWN, for every gloss were three alternative translation generated with Google Translator, Bing, and DeepL. Before running the experiments, this data has to be pre-processed in such a way that the resulting dataset will contain all 29,494 Ukrainian lemmas, and for every lemma, there are its corresponding lemma and gloss from PWN and three translations of the gloss into Ukrainian.

The pipeline of the experiment is following:

1. Choose 100 random entries from the generated dataset.
2. For every prompt, run a loop, and for each dataset entry, generate gloss using OpenAI Python API. Save outputs.
3. Compare generated glosses and choose the best-performing prompt.

This experiment was conducted two times: using GPT-4 and GPT-3.5. In the first case, the best prompt was **One-Shot Chain-of-Thought**, and glosses generated with it were the most accurate and grammatically correct. But when switched to GPT-3.5 in some cases this prompt generated output the same as in the example that it was given, therefore this prompt became unsuitable for the main task. With GPT-3.5, the best result showed **Zero-Shot Chain-of-Thought**, and because this is the model that we use for the main experiment, we chose this prompt.

TABLE 6.3: Example of generated gloss for word "клювати"(peck) with different prompts by GPT3.5.

prompt	generated gloss
zero_shot	\ "клювати\" українському WordNet означає їсти наподоби птаха
zero_shot_cot_reasoning	дія, коли птах їсть їжу невеликими порціями
one_shot	їсти так, як птах
one_shot_cot	їсти подібно до птаха
guideline_prompt_with_reasoning	дія, що характеризується тим, як птахи їдять

### 6.3 Gloss Generation

The gloss generation for all words in the Ukrainian WordNet basis is very similar to the experiment described in Section 6.2. So for the entire dataset of 29,494 words, we generate glosses with GPT-3.5 using **Zero-Shot Chain-of-Thought** prompt. After the experiment, the generated outputs were saved for evaluation analysis.

### 6.4 Results Evaluation

To evaluate LLM's performance, we compared generated glosses and different translations of PWN glosses into Ukrainian. Evaluation is performed on a sample of words with one gloss present in SUM-20 with unambiguous dictionary definitions. After filtering 29,494 synsets that form the base of Ukrainian WordNet, we get approximately 4,000 words that fit the criteria. To measure semantic similarity between glosses and dictionary definitions, we used a sentence-transformers model fine-tuned for the Ukrainian language[24].

TABLE 6.4: Example of generated gloss for a word "небо"(sky), original gloss from PWN3.1, different translations and dictionary definition.

<b>pwn</b>	the atmosphere and outer space as viewed from the earth
<b>llm</b>	атмосфера землі та зовнішній простір, які спостерігаються з планети
<b>google</b>	атмосфера та космічний простір, якщо дивитися із землі
<b>bing</b>	атмосфера і космічний простір, як видно з Землі
<b>deepl</b>	атмосфера і космічний простір, якщо дивитися з Землі
<b>sum_20</b>	Видимий над поверхнею землі повітряний простір у формі купола

Results of gloss evaluation for every method are very close, the general statistics for every method are shown in the Table 6.5. Figures 6.1, 6.2 and 6.3 show visualizations of evaluation results.

After comparing the results of evaluation, it becomes clear that LLM performed gloss generation better than Google Translate and Bing and slightly worse than DeepL, however differences between all methods are very insignificant. However, it is important to mention that we compared a subset of almost 4,000 words that have only one definition. So, results still have to be manually checked by professional lexicographers.



TABLE 6.5: General statistics of semantic similarity scores between different glosses and dictionary definitions.

	mean	max	min	median	1st quartile	3rd quartile
llm	0.70	0.99	-0.02	0.76	0.61	0.85
google	0.69	0.99	-0.03	0.76	0.60	0.85
bing	0.68	0.99	-0.03	0.75	0.59	0.84
deepl	0.70	0.99	-0.02	0.76	0.62	0.85

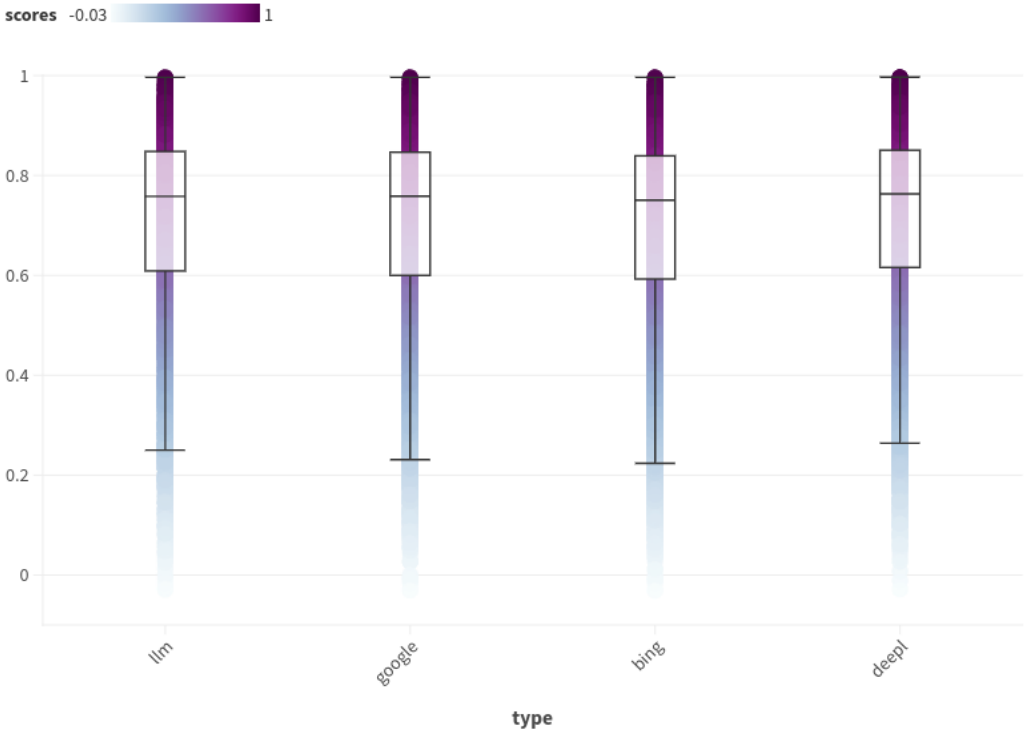


FIGURE 6.1: Comparison of glosses generated by different methods and a score of their cosine similarity to dictionary definitions.

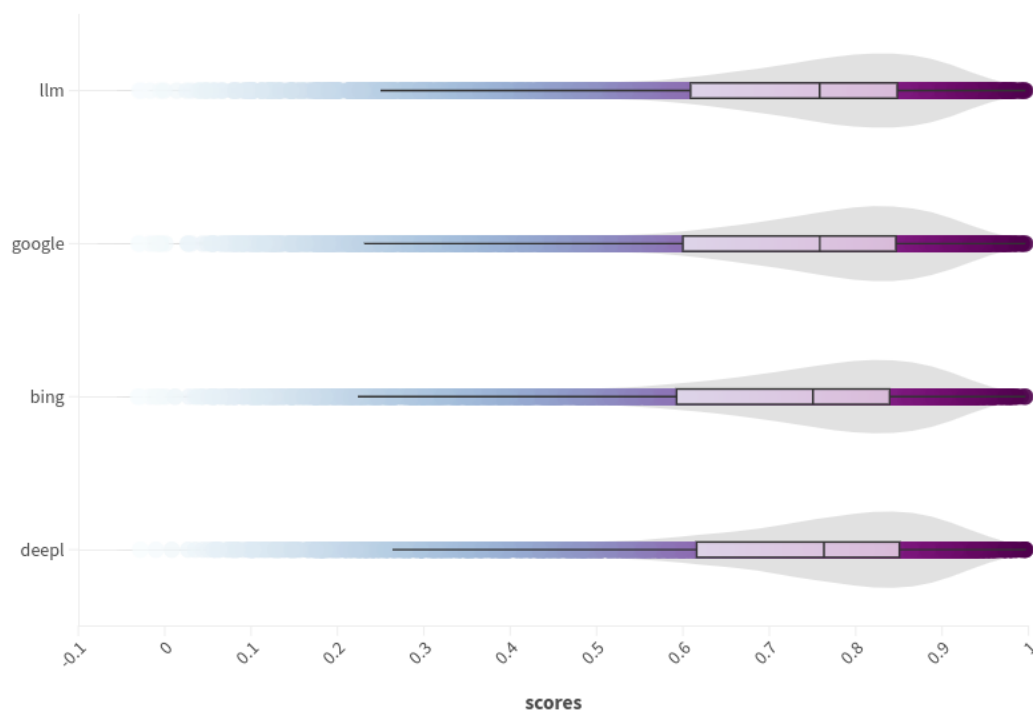


FIGURE 6.2: Comparison of glosses generated by different methods and a score of their cosine similarity to dictionary definitions.

Overall, taking into account the lack of research in the area of gloss generation with LLMs, the results are promising and show potential in this area of WordNet development.

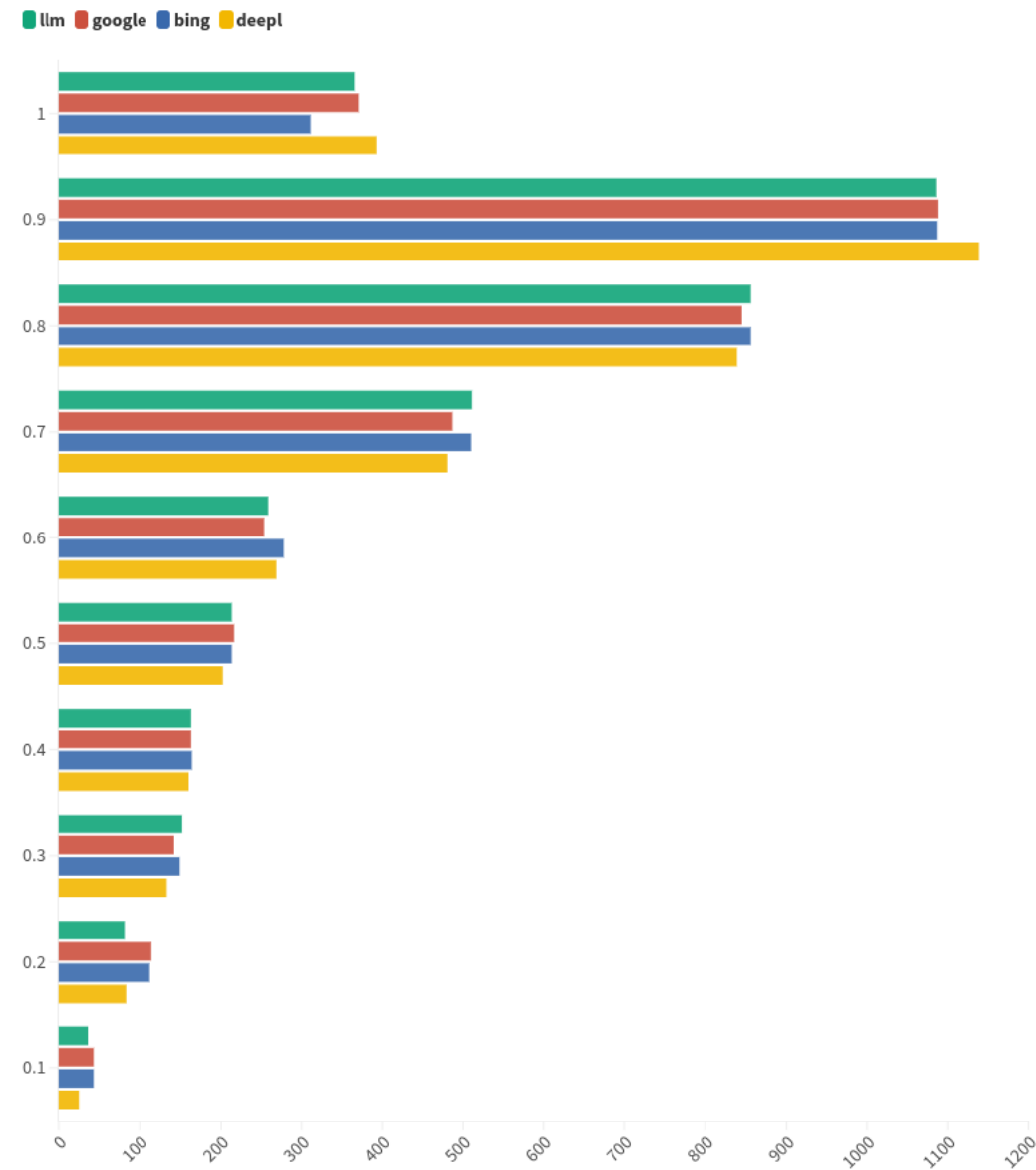


FIGURE 6.3: Histogram of semantic similarity scores after evaluation.

## Chapter 7

# Conclusions

### 7.1 Contribution

The research presented in this thesis contributes to a natural language processing field by introducing a novel approach to gloss generation using large language models. This work also expanded the foundation for the new Ukrainian WordNet resource. We used different prompt engineering techniques to test their performance on gloss generation.

We replicated the experiment of linking PWN with Ukrainian Wikipedia and increased the number of synsets from 21,015 to 29,494. For all synsets, we generated glosses that complement and add understanding to every word in the Ukrainian WordNet.

This thesis continues the work of building a comprehensive and reliable WordNet for the Ukrainian language. The future work that is required to develop a better resource is discussed in Section 7.3. The code for gloss generation experiment<sup>1</sup> is available on GitHub.

### 7.2 Limitations

The scope of this thesis is limited to automatically generating glosses for established WordNet basis of 29,494 words connected by hypernym relations. Generated data has only been evaluated automatically, and further professional verification and input from linguists are necessary.

Additionally, the whole Ukrainian WordNet project requires a lot of work and development. It is necessary to include other lexico-semantic relationships. Also, as Ukrainian WordNet has been developed by expand method, it has the underlying structure of PWN. Therefore, many linguistic and cultural nuances of the Ukrainian language are not presented fully.

### 7.3 Future Work

WordNet creation is a long and complex process, so there are many directions for future research.

Here are some possible steps for the future work:

1. Current generated glosses must be evaluated manually by professional linguists and lexicographers. A guideline must also be created so that the evaluation process can be unified.

---

<sup>1</sup>[https://github.com/hannusia/wordnet\\_thesis](https://github.com/hannusia/wordnet_thesis)

2. Try Auto Prompt[25] or a similar framework on the current prompt to increase its effectiveness.
3. Some other prompt engineering techniques were not discussed in Section 3.2.2 and were not implemented, but are worth trying, for example, using emotional bias by adding phrases like "Consider your response carefully. A man's life is at stake."
4. Experiment on gloss generation with different LLMs from GPT and other families.
5. As Ukrainian Wikipedia is a resource that continues to grow, it is essential to periodically replicate experiments of building WordNet basis and generating glosses for new synsets.
6. With the help of professional lexicographers, construct guidelines for manual evaluation that will help annotators upgrade the current Ukrainian WordNet.
7. WordNet is an important resource for the Ukrainian language; it has great potential and should be open to the general public

We don't know what is next. *But we know who is next...* [8]

## Appendix A

# Propmts for Gloss Generation

### A.1 Zero-Shot Prompt

```
from langchain.prompts import PromptTemplate
zero_shot = PromptTemplate.from_template(
    """
    Generate gloss in Ukrainian WordNet for lemma "{lemma_ukr}"
    which is lemma "{lemma_eng}" in English WordNet,
    based on gloss from English Wordnet: "{gloss}"
    and its translations in Ukrainian, each translation in separate
    angle brackets: <{gloss_trans1}>, <{gloss_trans2}>, <{gloss_trans3}>.
    """
)
```

### A.2 One-Shot Prompt

```
from langchain.prompts import PromptTemplate
one_shot = PromptTemplate.from_template(
    """
    Your task is to answer in a consistent style.
    {example}

    <reseacher>: Generate gloss in Ukrainian WordNet for lemma "{lemma_ukr}"
    which is lemma "{lemma_eng}" in English WordNet,
    based on gloss from English Wordnet: "{gloss}"
    and its translations in Ukrainian, each translation in separate
    angle brackets: <{gloss_trans1}>, <{gloss_trans2}>, <{gloss_trans3}>.
    """
)
```

### A.3 Zero-Shot Chain-of-Thought Prompt

```
from langchain.prompts import PromptTemplate
zero_shot_cot_reasoning = PromptTemplate.from_template(
    """
    Generate gloss in Ukrainian WordNet for lemma "{lemma_ukr}"
    which is lemma "{lemma_eng}" in English WordNet,
    based on gloss from English Wordnet: "{gloss}"
    and its translations in Ukrainian, each translation
    in separate angle brackets:
    <{gloss_trans1}>, <{gloss_trans2}>,<{gloss_trans3}>.
    """
)
```

Think carefully and logically, provide reasoning to your answer.

Provide your answer in JSON format with  
the following keys: reasoning, gloss.  
"""

)

## A.4 One-Shot Chain-of-Thought Prompt

```
from langchain.prompts import PromptTemplate
one_shot_cot = PromptTemplate.from_template(
    """
    Your task is to answer in a consistent style.
    {example}

    <researcher>: Generate gloss in Ukrainian WordNet for lemma "{lemma_ukr}"
    which is lemma "{lemma_eng}" in English WordNet,
    based on gloss from English Wordnet: "{gloss}"
    and its translations in Ukrainian, each translation in separate
    angle brackets: <{gloss_trans1}>, <{gloss_trans2}>, <{gloss_trans3}>.

    Think carefully and logically, provide reasoning to your answer.

    Provide your answer in JSON format with the following keys: reasoning, gloss.
    """
)
```

## A.5 Guideline Prompt

```
from langchain.prompts import PromptTemplate
guideline_prompt_with_reasoning = PromptTemplate.from_template(
    """
    Your task is to propose your variant of Ukrainian gloss
    for a lemma in Ukrainian WordNet.
    To solve the problem do the following:
    – First, analyze given English gloss, and its translations into
      Ukrainian
    – Then evaluate each of the variants, and based on given information
      create your own gloss
    – Make the answer grammatically correct, pay special attention
      to the conjugation

    Give output in following format:
    """
    gloss: gloss that you generated
    reasoning: explain why you choose this gloss
    """

    Use the following format:
    Ukrainian lemma:
    """
    ukrainian lemma here
    """
)
```

English lemma:  
'''

english lemma here  
'''

Gloss from English WordNet:  
'''

gloss from english wordnet here  
'''

Gloss translations:  
'''

one or multiple translation of english gloss into ukrainian delimited by <>  
'''

Your answer:  
'''

Ukrainian gloss that you propose here  
'''

Ukrainian lemma:  
'''

{lemma\_ukr}  
'''

English lemma:  
'''

{lemma\_eng}  
'''

Gloss from English WordNet:  
'''

{gloss}  
'''

Gloss translations:  
'''

<{gloss\_trans1}>

<{gloss\_trans2}>

<{gloss\_trans3}>  
'''

Provide your answer in JSON format with  
the following keys: reasoning, gloss.  
"""

)



## Appendix B

# Referenced Prompts

### B.1 Tree-of-Thought Prompting Sample<sup>[17]</sup>

Imagine three different experts are answering this question.  
All experts will write down 1 step of their thinking,  
then share it with the group.  
Then all experts will go on to the next step, etc.  
If any expert realises they're wrong at any point then they leave.  
The question is...

# Bibliography

- [1] Alessandro Agostini et al. “UZWORDNET: A Lexical-Semantic Database for the Uzbek Language”. In: *Proceedings of the 11th Global Wordnet Conference*. Ed. by Piek Vossen and Christiane Fellbaum. University of South Africa (UNISA): Global Wordnet Association, Jan. 2021, pp. 8–19. URL: <https://aclanthology.org/2021.gwc-1.2>.
- [2] Antonietta Alonge et al. “The Linguistic Design of the EuroWordNet Database”. In: *Computers and the Humanities* 32 (1998), pp. 91–115. URL: <https://api.semanticscholar.org/CorpusID:41577228>.
- [3] Anatoly Anisimov et al. “Ukrainian WordNet: Creation and Filling”. In: *Proceedings of the 10th International Conference on Flexible Query Answering Systems - Volume 8132*. FQAS 2013. Granada, Spain: Springer-Verlag, 2013, 649–660. ISBN: 9783642407680. DOI: 10.1007/978-3-642-40769-7\_56. URL: [https://doi.org/10.1007/978-3-642-40769-7\\_56](https://doi.org/10.1007/978-3-642-40769-7_56).
- [4] Khuyagbaatar Batsuren et al. “Building the Mongolian WordNet”. In: *Proceedings of the 10th Global Wordnet Conference*. Ed. by Piek Vossen and Christiane Fellbaum. Wroclaw, Poland: Global Wordnet Association, July 2019, pp. 238–244. URL: <https://aclanthology.org/2019.gwc-1.30>.
- [5] Francis Bond and Takayuki Kuribayashi. “The Japanese Wordnet 2.0”. In: *Proceedings of the 12th Global Wordnet Conference*. Ed. by German Rigau, Francis Bond, and Alexandre Rademaker. University of the Basque Country, Donostia - San Sebastian, Basque Country: Global Wordnet Association, Jan. 2023, pp. 179–186. URL: <https://aclanthology.org/2023.gwc-1.22>.
- [6] Sonja E. Bosch and Marissa Griesel. “Strategies for building wordnets for under-resourced languages: The case of African languages”. In: 2017. URL: <https://api.semanticscholar.org/CorpusID:43940087>.
- [7] Tom B. Brown et al. *Language Models are Few-Shot Learners*. 2020. arXiv: 2005.14165 [cs.CL].
- [8] BUDDY et al. WIN. Performed by Ateez. 2019. URL: <https://open.spotify.com/track/6iaVgWHD1Yxjjvj3PuKSQX?si=fbd2ac2350b5427b>.
- [9] Mohamad Diab et al. *Stable Diffusion Prompt Book*. OpenArt, 2022. URL: <https://openart.ai/promptbook>.
- [10] Christiane Fellbaum. “WordNet and wordnets”. In: *Encyclopedia of Language and Linguistics*. Ed. by Keith Brown. Oxford: Elsevier, 2005, pp. 665–670. URL: <http://wordnet.princeton.edu/>.
- [11] Christiane D. Fellbaum. “A Semantic Network of English: The Mother of All WordNets”. In: *Computers and the Humanities* 32 (1998), pp. 209–220. URL: <https://api.semanticscholar.org/CorpusID:37534063>.
- [12] Christiane D. Fellbaum. “Wordnet: An Electronic Lexical Database”. In: *The Library Quarterly* 69.3 (1999), pp. 406–408. DOI: 10.1086/603115. eprint: <https://doi.org/10.1086/603115>. URL: <https://doi.org/10.1086/603115>.

- [13] Michael Wayne Goodman and Francis Bond. "Intrinsically Interlingual: The Wn Python Library for Wordnets". In: *Proceedings of the 11th Global Wordnet Conference*. Ed. by Piek Vossen and Christiane Fellbaum. University of South Africa (UNISA): Global Wordnet Association, Jan. 2021, pp. 100–107. URL: <https://aclanthology.org/2021.gwc-1.12>.
- [14] Roberto Gozalo-Brizuela and Eduardo C. Garrido-Merchán. *A survey of Generative AI Applications*. 2023. arXiv: 2306.02781 [cs.LG].
- [15] Chase Harrison. *LangChain*. 2022. URL: <https://github.com/langchain-ai/langchain>.
- [16] Pablo Hoffman, Daniel Graña, and Martin Olveyra. *A fast and powerful scraping and web crawling framework*. 2008. URL: <https://scrapy.org/>.
- [17] Dave Hulbert. *Using Tree-of-Thought Prompting to boost ChatGPT's reasoning*. <https://github.com/dave1010/tree-of-thought-prompting>. May 2023. DOI: 10.5281/ZENODO.10323452. URL: <https://doi.org/10.5281/zenodo.10323452>.
- [18] Frederick Jelinek. *Statistical Methods for Speech Recognition (Language, Speech, and Communication)*. The MIT Press, 1998. ISBN: 0262100665. URL: <http://www.amazon.fr/exec/obidos/ASIN/0262100665/citeulike04-21>.
- [19] Hongyan Jing. "Usage of WordNet in Natural Language Generation". In: *Usage of WordNet in Natural Language Processing Systems*. 1998. URL: <https://aclanthology.org/W98-0718>.
- [20] Gregory Johnson. "Google Translate <http://translate.google.com>". In: *Technical Services Quarterly* 29 (2012), p. 165. ISSN: 0731-7131. URL: <http://www.tandfonline.com/doi/abs/10.1080/07317131.2012.650971>.
- [21] Spencer Kelly. *WTF\_WIKIPEDIA: A pretty-committed Wikipedia markup parser*. 2017. URL: [https://github.com/spencermountain/wtf\\_wikipedia](https://github.com/spencermountain/wtf_wikipedia).
- [22] Takeshi Kojima et al. *Large Language Models are Zero-Shot Reasoners*. 2023. arXiv: 2205.11916 [cs.CL].
- [23] I. Kulchytsky, A. Romaniuk, and Kh. Khariv. "Rozroblennia Wordnet-podibnoho slovnyka ukrainskoi movy [Developing a WordNet-like Dictionary of Ukrainian]". In: *Bulletin of Lviv Polytechnic National University*. 2010, pp. 306–318. URL: <https://ena.lpnu.ua/handle/ntb/6774>.
- [24] Yurii Laba et al. "Contextual Embeddings for Ukrainian: A Large Language Model Approach to Word Sense Disambiguation". In: *Proceedings of the Second Ukrainian Natural Language Processing Workshop (UNLP)*. Ed. by Mariana Romanushyn. Dubrovnik, Croatia: Association for Computational Linguistics, May 2023, pp. 11–19. DOI: 10.18653/v1/2023.unlp-1.2. URL: <https://aclanthology.org/2023.unlp-1.2>.
- [25] Elad Levi, Eli Brosh, and Matan Friedmann. *Intent-based Prompt Calibration: Enhancing prompt optimization with synthetic boundary cases*. 2024. eprint: arXiv: 2402.03099.
- [26] Bing Liu. *Web data mining exploring hyperlinks, contents, and usage data*. 2011. URL: <http://dx.doi.org/10.1007/978-3-642-19460-3>.
- [27] Alexander Loth, Martin Kappes, and Marc-Oliver Pahl. *Blessing or curse? A survey on the Impact of Generative AI on Fake News*. 2024. arXiv: 2404.03021 [cs.CL].

- [28] Rui Mao, Chenghua Lin, and Frank Guerin. "Word Embedding and WordNet Based Metaphor Identification and Interpretation". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Ed. by Iryna Gurevych and Yusuke Miyao. Melbourne, Australia: Association for Computational Linguistics, July 2018, pp. 1222–1231. DOI: [10.18653/v1/P18-1113](https://doi.org/10.18653/v1/P18-1113). URL: <https://aclanthology.org/P18-1113>.
- [29] John P. McCrae et al. "English WordNet 2019 – An Open-Source WordNet for English". In: *Proceedings of the 10th Global Wordnet Conference*. Ed. by Piek Vossen and Christiane Fellbaum. Wroclaw, Poland: Global Wordnet Association, July 2019, pp. 245–252. URL: <https://aclanthology.org/2019.gwc-1.31>.
- [30] John Philip McCrae et al. "English WordNet 2020: Improving and Extending a WordNet for English using an Open-Source Methodology". English. In: *Proceedings of the LREC 2020 Workshop on Multimodal Wordnets (MMW2020)*. Ed. by Thierry Declerk, Itziar Gonzalez-Dios, and German Rigau. Marseille, France: The European Language Resources Association (ELRA), May 2020, pp. 14–19. ISBN: 979-10-95546-41-2. URL: <https://aclanthology.org/2020.mmw-1.3>.
- [31] Meta. *Introducing Meta Llama 3: The most capable openly available LLM to date*. Accessed: 2024-04-28. 2024. URL: <https://ai.meta.com/blog/meta-llama-3/>.
- [32] Rada Mihalcea and Dan Moldovan. "AutoASC - A System for Automatic Acquisition of Sense Tagged Corpora." In: *IJPRAI 14* (Feb. 2000), pp. 3–18. DOI: [10.1142/S0218001400000039](https://doi.org/10.1142/S0218001400000039).
- [33] Tomas Mikolov et al. *Efficient Estimation of Word Representations in Vector Space*. 2013. arXiv: [1301.3781](https://arxiv.org/abs/1301.3781) [cs.CL].
- [34] George A. Miller. "WordNet: A Lexical Database for English". In: *Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994*. 1994. URL: <https://aclanthology.org/H94-1111>.
- [35] George A. Miller et al. "Introduction to WordNet: An On-line Lexical Database\*". In: *International Journal of Lexicography* 3.4 (Dec. 1990), pp. 235–244. ISSN: 0950-3846. DOI: [10.1093/ijl/3.4.235](https://doi.org/10.1093/ijl/3.4.235). eprint: <https://academic.oup.com/ijl/article-pdf/3/4/235/9820417/235.pdf>. URL: <https://doi.org/10.1093/ijl/3.4.235>.
- [36] Shervin Minaee et al. "Large Language Models: A Survey". In: (2024). URL: <https://arxiv.org/pdf/2402.06196.pdf>.
- [37] Jorge Morato et al. "WordNet Applications". In: 2004, pp. 270–278. URL: <http://www.fi.muni.cz/gwc2004/proc/105.pdf>.
- [38] Laxmi Nadageri and Y.V. Haribhakta. "Building Wordnet: a Survey". In: *International Journal of Engineering Science and Innovative Technology* 4.6 (2017), pp. 17–27.
- [39] Reiichiro Nakano et al. *WebGPT: Browser-assisted question-answering with human feedback*. 2022. arXiv: [2112.09332](https://arxiv.org/abs/2112.09332) [cs.CL].
- [40] Sharan Narang and Aakanksha Chowdhery. *Pathways Language Model (PaLM): Scaling to 540 Billion Parameters for Breakthrough Performance*. Accessed: 2024-04-28. 2022. URL: <https://research.google/blog/pathways-language-model-palm-scaling-to-540-billion-parameters-for-breakthrough-performance/>.

- [41] Roberto Navigli. "Word Sense Disambiguation: A Survey". In: *ACM Comput. Surv.* 41 (Feb. 2009). DOI: [10.1145/1459352.1459355](https://doi.org/10.1145/1459352.1459355).
- [42] Jan Nemrava. "Refining search queries using WordNet glosses". In: (Jan. 2006).
- [43] OpenAI. *GPT-2: 1.5B release*. Accessed: 2024-04-28. 2019. URL: <https://openai.com/research/gpt-2-1-5b-release>.
- [44] OpenAI. *OpenAI Python Library*. URL: <https://github.com/openai/openai-python>.
- [45] OpenAI et al. *GPT-4 Technical Report*. 2023. arXiv: [2303.08774](https://arxiv.org/abs/2303.08774) [cs.CL].
- [46] Heili Orav, Kadri Vare, and Sirli Zupping. "Estonian Wordnet: Current State and Future Prospects". In: *Proceedings of the 9th Global Wordnet Conference*. Ed. by Francis Bond, Piek Vossen, and Christiane Fellbaum. Nanyang Technological University (NTU), Singapore: Global Wordnet Association, Jan. 2018, pp. 347–351. URL: <https://aclanthology.org/2018.gwc-1.42>.
- [47] Long Ouyang et al. *Training language models to follow instructions with human feedback*. 2022. arXiv: [2203.02155](https://arxiv.org/abs/2203.02155) [cs.CL].
- [48] Maciej Piasecki, Stanisław Szpakowicz, and Bartosz Broda. *A Wordnet from the Ground Up*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 2009. URL: [http://www.plwordnet.pwr.wroc.pl/main/content/files/publications/A\\_Wordnet\\_from\\_the\\_Ground\\_Up.pdf](http://www.plwordnet.pwr.wroc.pl/main/content/files/publications/A_Wordnet_from_the_Ground_Up.pdf).
- [49] Venkatesh Prabhu et al. "An Efficient Database Design for IndoWordNet Development Using Hybrid Approach". In: *Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing*. Ed. by Virach Sormlertlamvanich and Abbas Malik. Mumbai, India: The COLING 2012 Organizing Committee, Dec. 2012, pp. 229–236. URL: <https://aclanthology.org/W12-5021>.
- [50] Eric Prud'hommeaux and Andy Seaborne. *SPARQL Query Language for RDF*. W3C Recommendation. <http://www.w3.org/TR/rdf-sparql-query/>. 2008. URL: <http://www.w3.org/TR/rdf-sparql-query/>.
- [51] Alexandre Rademaker, Abhishek Basu, and Rajkiran Veluri. "Semantic Parsing and Sense Tagging the Princeton WordNet Gloss Corpus". In: *Proceedings of the 12th Global Wordnet Conference*. Ed. by German Rigau, Francis Bond, and Alexandre Rademaker. University of the Basque Country, Donostia - San Sebastian, Basque Country: Global Wordnet Association, Jan. 2023, pp. 243–253. URL: <https://aclanthology.org/2023.gwc-1.30>.
- [52] Alec Radford and Karthik Narasimhan. "Improving Language Understanding by Generative Pre-Training". In: 2018. URL: <https://api.semanticscholar.org/CorpusID:49313245>.
- [53] Nataliia Romanyshyn. *Data-Driven Approach to Automated Hypernym Hierarchy Construction for the Ukrainian WordNet*. June 2023. DOI: [10.13140/RG.2.2.31621.32486](https://doi.org/10.13140/RG.2.2.31621.32486).
- [54] Thomas Ronzon. "DeepL - überraschend anders". Deutschland. In: *Javaspektrum* 2 (2018), p. 69.
- [55] Claude E. Shannon. "Prediction and entropy of printed English". In: *Bell System Technical Journal* 30 (1951), pp. 50–64. URL: <https://api.semanticscholar.org/CorpusID:9101213>.

- [56] Melanie Siegel, Maksym Vakulenko, and Jonathan Baum. "Towards UkrainianWordNet: Incorporation of an Existing Thesaurus in the Domain of Physics". In: *Conference on Natural Language Processing*. 2023. URL: <https://api.semanticscholar.org/CorpusID:267770749>.
- [57] Sareewan Thoongsup et al. "Thai WordNet Construction". In: *Proceedings of the 7th Workshop on Asian Language Resources (ALR7)*. Ed. by Hammam Riza and Virach Sornlertlamvanich. Suntec, Singapore: Association for Computational Linguistics, Aug. 2009, pp. 139–144. URL: <https://aclanthology.org/W09-3420>.
- [58] Hugo Touvron et al. *LLaMA: Open and Efficient Foundation Language Models*. 2023. arXiv: 2302.13971 [cs.CL].
- [59] G. Van Rossum and F.L. Drake. *Python 3 Reference Manual: (Python Documentation Manual Part 2)*. Documentation for Python. CreateSpace Independent Publishing Platform, 2009. ISBN: 9781441412690. URL: <https://books.google.com.ua/books?id=KIybQQAACAAJ>.
- [60] Ashish Vaswani et al. *Attention Is All You Need*. 2023. arXiv: 1706.03762 [cs.CL].
- [61] James Vincent. *Google opens up its AI language model PaLM to challenge OpenAI and GPT-3*. Accessed: 2024-04-28. 2023. URL: <https://www.theverge.com/2023/3/14/23639313/google-ai-language-model-palm-api-challenge-openai>.
- [62] Veronika Vincze and Attila Almási. "Non-Lexicalized Concepts in Wordnets: A Case Study of English and Hungarian". In: *Proceedings of the Seventh Global Wordnet Conference*. Ed. by Heili Orav, Christiane Fellbaum, and Piek Vossen. Tartu, Estonia: University of Tartu Press, Jan. 2014, pp. 118–126. URL: <https://aclanthology.org/W14-0116>.
- [63] P.J.T.M. Vossen. "EuroWordNet: a multilingual database for information retrieval". English. In: *Proceedings of the DELOS workshop on Cross-language Information Retrieval, March 5-7, 1997 Zurich*. Proceedings of the DELOS workshop on Cross-language Information Retrieval, March 5-7, 1997, Zurich. Vrije Universiteit, 1997. URL: <https://dl.acm.org/doi/10.5555/314515>.
- [64] Jason Wei et al. *Chain-of-Thought Prompting Elicits Reasoning in Large Language Models*. 2023. arXiv: 2201.11903 [cs.CL].
- [65] Shunyu Yao et al. *Tree of Thoughts: Deliberate Problem Solving with Large Language Models*. 2023. arXiv: 2305.10601 [cs.CL].
- [66] Wojciech Zaremba and Greg OpenAI Brockman. *OpenAI Codex*. Accessed: 2024-04-28. 2021. URL: <https://openai.com/blog/openai-codex>.
- [67] Wayne Xin Zhao et al. *A Survey of Large Language Models*. 2023. arXiv: 2303.18223 [cs.CL].