TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Rasmus Vaik 142375IAPB

# DETECTING POLYSEMY STRUCTURES IN WORDNET

Bachelor's thesis

Supervisor: Ahti Lohk
PhD

Tallinn 2019

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Rasmus Vaik 142375IAPB

# POLÜSEEMIA STRUKTUURIDE TUVASTAMINE WORDNETIST

Bakalaureusetöö

Juhendaja: Ahti Lohk
PhD

Tallinn 2019

# Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Rasmus Vaik

22.5.2019

# Abstract

The purpose of this thesis is to visualise possible problematic spots in the lexical-semantic resource called wordnet.

The wordnet semantic hierarchies will be searched for polysemous lexical units (typically words with the same orthography or multi word expressions), between which both IS-A and vertical polysemous relationships are present and which form substructures of the hierarchies.

These polysemous structures will be presented by separately extracting the semantic hierarchy substructures of the wordnet for each polysemous word, for which IS-A relationships are simultaneously vertical polysemy relationships. Although in individual cases vertical polysemy relationships have previously been detected in wordnet, this approach has not been used to study polysemous structures where such relationships are involved.

Polysemy is one of the phenomena that is prone to producing errors and problematic spots in wordnet. As a result, polysemy is also one of the phenomena in wordnets that is the most studied. The result of this thesis will provide a new way to view polysemous spots in wordnet. The resulting polysemous structures will try to give lexicographers a bigger picture for each lexical unit that produces a polysemous structure, so that these structures may be verified and possibly corrected.

This thesis is written in English and is 28 pages long, including 6 chapters, 8 figures and 4 tables.

# Annotatsioon

## Polüseemia struktuuride tuvastamine wordnetist

Antud töös tegeletakse wordneti kui leksikaal-semantilise ressursi võimalike probleemkohtade visualiseerimisega.

Wordneti semantilistest hieararhiatest ekstraheeritakse polüseemseid leksikaalüksusi (tüüpiliselt sama ortograafiaga sõnad või mitmesõnalised väljendid), milliste vahel on nii hüperonüümia kui ka vertikaalse polüseemia seos ja mis moodustavad hierarhiates alamstruktuure. Üks sidus polüseemia struktuur hõlmab korraga vaid üht leksikaalset üksust.

Kuigi vertikaalse polüseemia seost on ka varem wordneti puhul tuvastatud, pole neid uuritud sidusate polüseemia-struktuuride kontekstis.

Leksikaalne polüseemia on wordnetis üks kõige kergemini vigu ja erinevaid probleemkohti esile toov nähtus. Seetõttu on polüseemia ka üks enim uuritud nähtus wordnetis. Praeguse töö tulemus annab uue vaate leksikaalse polüseemiaga seotud kohtadele wordneti semantilistes hierarhiates. Leitavad polüseemia struktuuride eesmärk on anda leksikograafidele suurem pilt iga polüseemiastruktuuri tekitava leksikaalüksuse kohta, et nende struktuuride korrektsust kontrollida ja vajadusel sisse viia muudatusi.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 28 leheküljel, 6 peatükki, 8 joonist, 4 tabelit.

# List of abbreviations and terms

| | |
|---|---|
| TalTech | Tallinn University of Technology |
| EstWn | Estonian Wordnet |
| NLP | Natural language processing |

**Synset** or **set of synonyms.** A group of cognitively similar synonyms.

**Lexical unit.** A member of a synset, typically word or multi word expression.

**Polysemy.** A phenomenon where a word or phrase has two or more meanings, and these meanings are interconnected.

# Table of contents

# List of figures

# List of tables

# 1 Introduction

As the fields of artificial intelligence and automatic text analysis have developed, the need for in-depth knowledge and descriptions of language have become increasingly apparent and for such applications, large lexical-semantic databases have been created. The widest spread of such databases are wordnet-type dictionaries or wordnets. These databases are mainly put together by expert linguists who possess a very good grasp of language.

During the course of natural development of wordnet-type dictionaries, they are constantly improved and corrected. New concepts and new or different semantic relations might be added as development progresses. When correcting the database some semantic relations might be replaced with new ones or an old incorrect relation might be removed.

However, one of the problems wordnet-type dictionaries have is that of being too specific in its distinction of polysemic words. For instance, a regular speaker of the Estonian language might distinguish between three to four meanings for the word “tee”, but the Estonian WordNet has twelve different meanings for it.

For the purpose of solving this problem multiple methods for reducing polysemy have been proposed, but they tend to be limited in their scope of polysemy. To allow for the further development of methods for polysemy reduction, a way of distinguishing large polysemous structures is needed. This circumstance is one of the incentives for studying the occurrence of polysemy. However, the motivating factors in the context of this thesis are the following:

1. The developers of wordnet lack an overview of the different ways of manifestation for polysemy.
2. Polysemy is a phenomenon, of which the manifestation needs to be scrutinised. Earlier works of detecting polysemy have always resulted in exposing unexpected mistakes.
3. Natural language processing is directly affected by the quality of a wordnet.

However, this work will not try to solve all the problems relating to polysemy. Instead the goal is to find specific structures, from the wordnet semantic hierarchies, that should demonstrate how lexical units with large amounts of polysemy form such structures. The expected structures consist of substructures of semantic hierarchies. The nodes of such a structure should all hold lexical units with the same orthography and between which simultaneously are relations of *hyperonymy/hyponymy* as well as vertical polysemy. The resultant substructures will be provided to the group of researchers working with wordnet at University of Tartu where all such occurrences will be validated.

In this thesis the 70th version of the Estonian Wordnet is used as a basis for a wordnet-type dictionaries. The Estonian Wordnet is based on wordnet theory and was built following principles adopted from the EuroWordNet and Princeton WordNet projects. In October 2018 the Estonian Wordnet held a total of about 139 000 words for 86 000 different concepts and between these 239 000 relations are noted. The types of words include adjectives, adverbs, nouns and verbs, but a number of multiword units are also present [1].

The goal of this thesis is to create an algorithm, that can detect the fore mentioned polysemous structures in a wordnet-type dictionary. The results will be visualised, validated and an overview will be given.

This Bachelor thesis is composed of the following parts: in the second chapter theoretical background is presented. In this chapter an overview of concepts relating to polysemy and wordnet will be given, as well as discussing the used technologies. The third chapter is comprised of chapters describing the algorithm proposed in this work. In the fourth chapter the results are described as well as how they were validated.

# 2 Theoretical background

In this chapter the theoretical background, necessary for understanding the following chapters, will be given. To start with, the essence, structure and fields of use of wordnet will be described. Secondly, we will define polysemy and related concepts. The phenomenon of vertical polysemy will also be made clear. Thirdly, we will describe the different possible forms of manifestation of polysemy in wordnet hierarchies. Finally, an overview of used technologies will be given.

## 2.1 Wordnet

A wordnet is a large lexical database of a language. Groups of cognitive synonyms, each expressing a distinct concept, called synsets are formed of adjectives, adverbs, nouns and verbs [2].

Between synsets (concepts) different types of semantic relationships occur. Some of these create a hierarchy of concepts (E.g. *hyperonymy* or *IS-A* relationship, *meronymy*) and some do not (E.g. *role of*, *type of*). The most attractive relationship for NLP tasks is the IS-A relationship, that occur between concepts of nouns and verbs creating hierarchies of noun and verb type. In this work as well, we will exclusively be dealing with the IS-A also known as the *hyperonymy* relationship. The opposite relationship to *hyperonymy* is the *hyponymy*. Therefore, we can say that the *hyperonymy/hyponymy* relationship links general and more specific concepts and, in a wordnet, is considered the most important semantic relationship. Such relationships link more general *hyperonyms* to *hyponyms*, that have a more specific meaning. The more specific concept inherits all the information present in the linked more general concept [3].

Wordnet can be used as a synonyms dictionary by searching the wordnet for synonyms to a word that is currently of interest. More often though are wordnets used in natural language processing tasks as a background knowledge base.

## 2.2 Concepts related to polysemy

As the topic of this work is to do with polysemous structures in a wordnet it is important to understand the different types of relations, types of polysemy and other related concepts that can be observed in a wordnet.

**Homonymy** is one of the two forms of lexical ambiguity (the other being Polysemy) and refers to the phenomenon of "One of two or more words spelled and pronounced alike but different in meaning (such as the noun quail and the verb quail)" [4].

**Metonymy** is the use of a word or phrase in a figurative sense based on chronological, spatial, causal or other relation [5].

According to Freihat a word is **systematic polysemous** when the "meanings of this word are not homonyms and they describe different aspects of the same term" [6].

**Specialization polysemy** is a word or phrase used to refer [to] a more general meaning and a more specific meaning [7].

A **metaphor** is "a word or phrase for one thing to refer to another thing in order to show or suggest that they are similar" [8].

**Vertical polysemy** - A. Koskela in "Metonymy, category broadening and narrowing, and vertical polysemy" defines vertical polysemy as the multiplicity of meaning that results from semantic broadening and narrowing [9]. Therefore, in the case of vertical polysemy, an IS-A relationship is also always present. In addition, it should be mentioned that vertical polysemy might also be present in the case where the relationship between the polysemous word is not that of a superordinate-subordinate or a father-son, but that of a grandfather-grandchild. However, in this work we will only be focusing on the father-child relationship.

## 2.3 Overview of different polysemous structures

In the article "The structure of Polysemy: A study of multi-sense words based on WordNet" Jen-Yi Lin, Chang-Hua Yang, Shu-Chuan Tseng and Chu-Ren Huang define multiple ways senses can be clustered based on synset and features like entailment and polysemy.

The following examples are from the Princeton WordNet version 3.0 [10] and were first presented in a PhD thesis by Ahti Lohk [7].

The word senses are called **Sisters** if two or more-word senses share the same hypernym (or parent) (Figure 1).

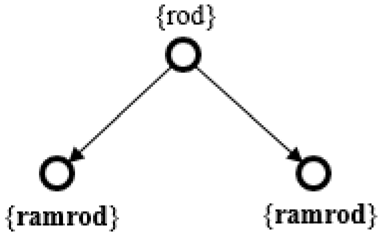

Figure 1. Polysemy structure - sisters

**Cousins**, synsets with one word in common and also at least one pair of direct or indirect related hypernyms, based on a predefined list [11].

**Twins** are synsets have one or more identical members they are called twins (Figure 2).

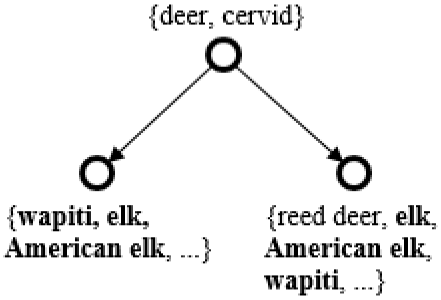
{deer, cervid}
{wapiti, elk, American elk, ...}
{reed deer, elk, American elk, wapiti, ...}

Figure 2. Polysemy structure - twins

**Child** is a sense cluster, where a superordinate synset might have the same lexicon entry as its subordinate synset, it is possible the *hyperonymy/troponomy* relation might link the multiple senses of a polysemous verb (Figure 3).

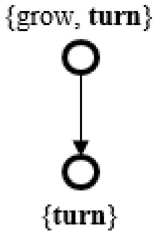

Figure 3. Polysemy structure - child

There are also cases where more than two senses can share a hypernymic/troponymic **Chain** and all of them share the same word form (Figure 4).

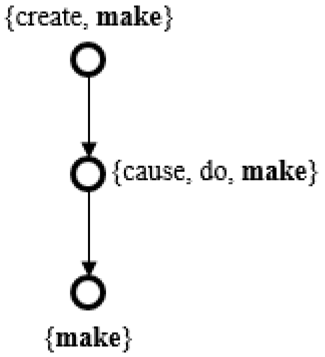

Figure 4. Polysemy structure - chain

A **Triangle** can be observed in the case where sister senses both have the same hypernym which also shares the same word form as the sisters (Figure 5).

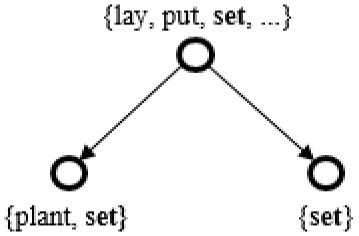

Figure 5. Polysemy structure - triangle

It is also possible that the sense tree might be more complex than the above described patterns.

The resultant structures are expected to contain child, chain and triangle shape structures, due to the conditions set in this thesis being fully satisfied for those shapes - lexical units with matching orthography (or same spelling) are connected by an IS-A (*hyperonymy*) as well as a vertical polysemous relationship.

## 2.4 Overview of used technologies

Due to the complex nature of the topic of this thesis, a lot of different libraries are used to help simplify the resultant algorithm and to handle different aspects of the algorithm that would otherwise be strenuous to implement without the use of such libraries. In this chapter, a list of these technologies will be given and each of the choices will be explained.

### 2.4.1 Python

Python is a powerful high-level multipurpose programming language, that is designed to be quick, easily readable and portable between platforms. Python was chosen for its support of a wide variety of different libraries for very specific applications.

### 2.4.2 NetworkX

NetworkX is an Open source and well tested Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks [12].

NetworkX was used for its simple yet powerful data structures for graphs, in which nodes can have multiple fields that can hold a wide array of values and the graphs edges can hold arbitrary data, which allows us to hold the specific word of the synset for which the polysemy was detected.

The provided graph data structure has a built-in function for extracting all connected component subgraphs which is very helpful for getting all of the polysemous structures that were detected and graphed by our algorithm.

Additionally, the graph data structure has the ability to ignore duplicate edges, which is useful for simplifying the algorithm by eliminating the need for additional checks for redundant edges for our graph, that might be present in a directed tree like the WordNet.

### 2.4.3 Matplotlib

Matplotlib is a 2d plotting library that is supported by Python scripts, Python and IPython shells. It was mainly used for its Pyplot module and also its multiple backends feature which allows the user to easily change if the created plots are displayed in an interactive plotting window or are saved into image files without opening a plotting window.

**Pyplot** is a lightweight MATLAB like plotting module provided by Matplotlib. Pyplot was chosen because of its simple implementation and good documentation.

The module provides the state-machine interface for the underlying plotting library [13]. Pyplot automatically creates figures and axes to achieve the desired plot and later re-uses the current axes. It also allows to set titles and labels for nodes and axis.

### 2.4.4 Pandas

Pandas is an open source library for the Python programming language, that provides lightweight, easily usable data structures and tools for data analysis. Pandas was used for its DataFrame data structure and also for its ability of reading a multiple table excel file into DataFrames.

The DataFrame is a tabular, two-dimensional data structure with labeled axes and mutable size. It is usually the most commonly used object in Pandas library. Pandas DataFrame was also useful for its inbuilt function of *DataFrame.loc[]* which allows to access rows or columns by label(s) or a boolean array.  This inbuilt function eliminated the need for a search algorithm for finding rows with certain parameters, as the needed row could easily be accessed *DataFrame.loc[]*.

# 3 Detecting polysemy structures

The goal of this work was to develop a program that could extract all of the vertical polysemy relationships between lexical units of synsets in the wordnet. Such relationships occur when lexical units in synsets, tied by *hyponymy/hyperonymy* relationship, share the same orthography but different definitions. These relations are added to a graph and visualised as a plot. In this chapter the techniques to achieve this will be discussed.

Firstly, the structure of the input data will be described. Then we will give an overview of the algorithm for searching for polysemous structures. Thirdly, a solution for eliminating the mixing of polysemous structures will be discussed. Lastly, the technique we use for visualising the resultant structures will be outlined.

## 3.1 Description of input data

Before a more in-depth description of the algorithm for finding polysemous structures, that relate to vertical polysemy, it is necessary to give an overview of the given input data structure.

The input in the context of this work is the EstWN that comes in the form of a .xlsx file. In the .xlsx file there are three tables:

1. **SS** for synsets, holds all of the synsets and their associated id's as seen in Table 1.
2. **REL** for relationships, holds pairs of synset ids and the relationship between them as seen in Table 2.
3. **DEF** for definitions, has the definitions of the associated synsets

For this algorithm we will only use tables "**SS**" and "**REL**".

The "SS" table holds the rows "id" and "synset". As seen in Table 1. the synset id in this version of the EstWN is comprised of three parts:

1. The wordnet version (EST70)
2. The six-digit synset id
3. Either a "v" or a "n", showing if the concept is that of a verb or a noun

Lexical units, that comprise the synsets, have a similar structure. First there is the lexical unit, then a letter to signify if the word is a noun, verb, adjective or adverb and then the sense index.

Table 1. Example of the “SS” table

| **id** | **synset** |
|---|---|
| EST70-000001-v | {korraldama_v_7, korda seadma_v_3, korrastama_v_5, korda tegema_v_3} |
| EST70-000002-n | {korraldamine_n_3} |
| ... | ... |
| EST70-000720-v | {seadma_v_2, korrastama_v_4, kohendama_v_2, sättima_v_3, korda tegema_v_2, korda seadma_v_5} |

In the “REL" table we have one row for each semantic relationship in the EstWN. We’ll be focusing on the *has_hyponym* relationship which signifies that synset with “id1” has a *hyponym* in the form of the synset with “id2”.

Table 2. Example of the “REL” table

| **id1** | **rel** | **id2** |
|---|---|---|
| EST70-000001-v | near_synonym | EST70-000720-v |
| EST70-000001-v | has_hyperonym | EST70-000195-v |
| EST70-000001-v | has_hyponym | EST70-005748-v |

## 3.2 Searching for polysemy

To find if a relationship is also a polysemous one we find both of the synsets, from the previous *has_hyponym* table row, by their id’s and split them up into lexical units. We then compare each of the lexical units from the first synset to each of the lexical units from the second synset. In case a matching word is found from both synsets, we add the synsets to a NetworkX graph data structure, with an edge between both nodes. The respective lexical units, for which a polysemous relation was detected, are saved as parameters for both these nodes. We do this for all of the rows in the “REL” table, that have *has_hyponym* in their rel column.

The simplicity of this approach is that most of the graph creation is handled by the NetworkX library. One NetworkX graph structure is populated with pairs of nodes and the connecting edge that polysemy was detected for. The graph structure automatically ties the nodes together based on node/synset ids, this creates the full final graph.

When all of the rows of the relationships table have been looked through, in the manner previously described, a inbuilt function of NetworkX graph data structure, for extracting all connected subgraphs is used. This produces a list with graph generators for all of the subgraphs.

## 3.3 Multiple words in a graph

Due to using NetworkX graph structure and adding nodes in pairs with edge between them, a new problem arises. The graph is populated with pairs of nodes with an edge between them and the actual formulation of graph is left up to the NetworkX library. As a result, graphs with multiple different polysemous words present are created. This is caused by one synset having two or more lexical units that are polysemous, and thus the the edges representing polysemy are for the same node and are displayed on the same graph as shown in Figure 1.

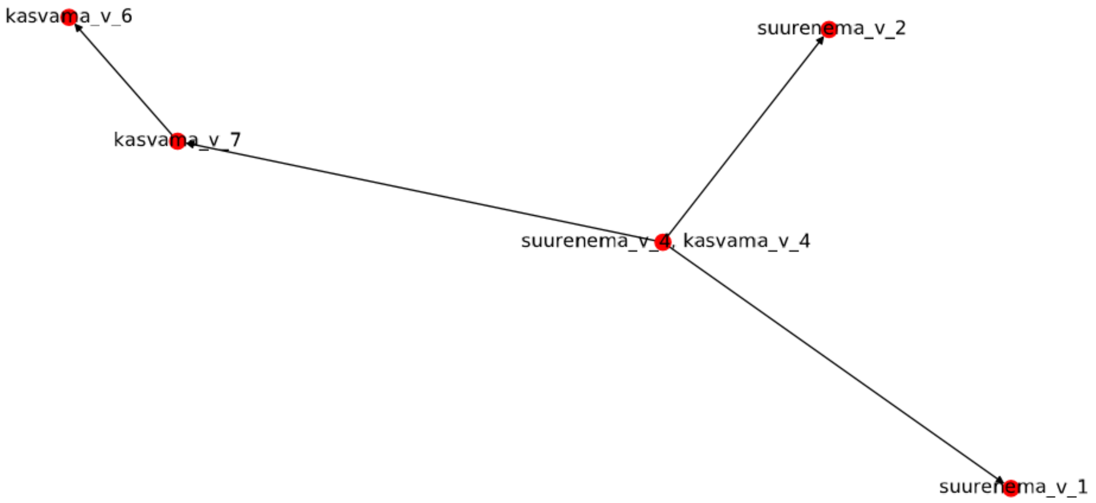

Figure 6. Example of graph with multiple polysemous words

To counteract this a helper function is made, that checks if a node with a polysemous edge is already present in the graph and if it contains a different word than the one currently used. If both conditions are true, then a new node with a slightly different id is made. For instance, the synset id is multiplied by 100000 so that a conflict is avoided with any of the original six-digit synset ids. Then a number is added to the new synset that signifies how many nodes have been made for this synset. This is done because one synset can have as many polysemous connections as many lexical units are present in that synset.

This approach splits the previous graph (Figure 6.) into the two following graphs seen on Figures 7. and 8., with the node holding "suurenema_v_4, kasvama_v_4" being split into two nodes.

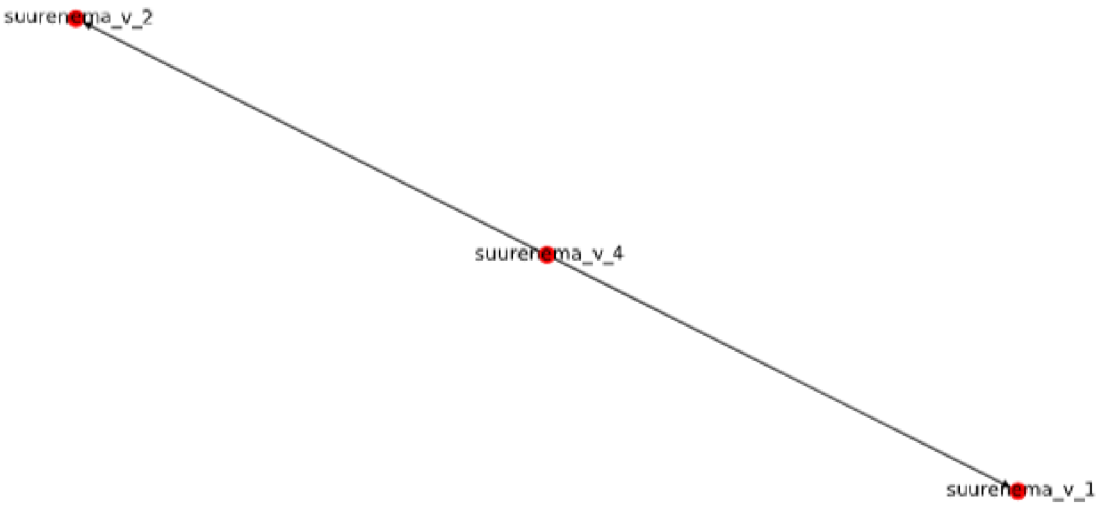

Figure 7. Example of "suurenema" graph after new approach

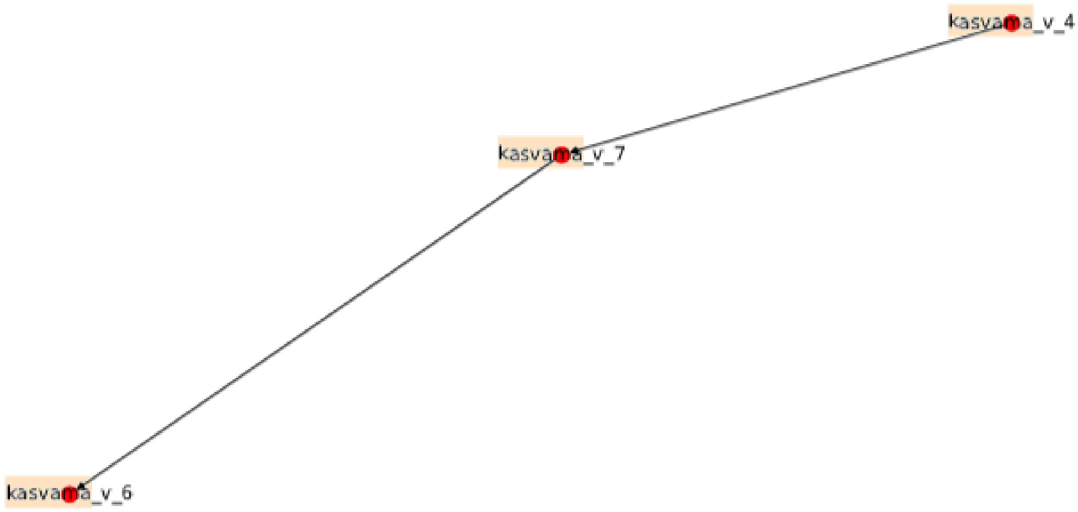

Figure 8. Example of a "kasvama" graph after new approach

## 3.4 Visualizing

We apply the NetworkX spring layout for all of the created graphs to achieve better readability of the plots. Finally, we iterate over the list of subgraphs, plotting each of the them individually using Matplotlib Pyplot module and save them to a multipage .pdf type file. Such a .pdf file can be searched for specific words or strings to find a graph with that word or string.

Due to constraints of the NetworkX framework, information held inside the nodes, such as the polysemous lexical unit, could not be displayed on top of the node. To get around this a additional dictionary type variable is made. Using a NetworkX function to get node attributes, this dictionary is populated with the data from nodes, using the node id as the key value. This dictionary is then added as an argument for the plotting function, which, using the keys, matches the appropriate lexical unit to their respective node as a label.

# 4 Results and their validation

For this thesis an algorithm for detecting and visualising polysemous structures was developed. Specifically structures that are formed when lexical units in synsets, that are connected with *hyperonymy/hyponymy* relationship, share the same orthography.

In this chapter the results of running this algorithm will be discussed and a overview of how the results were validate will be given.

## 4.1 Results

As input the algorithm was given the EstWn, in the form of a .xlsx file. From the input database with 68 000 different concepts and 210 000 relations, 881 different polysemous structures, with 836 different polysemous words, were detected and plotted.

As seen in table 3. the two largest of these structures were ones that contained six nodes/synsets. There were 779 structures with the smallest of 2 nodes detected.

Table 3. The amount of structures of different sizes that were detected

| **Size of structure** | **Frequency** |
|---|---|
| 2 | 779 |
| 3 | 77 |
| 4 | 16 |
| 5 | 7 |
| 6 | 2 |

The fact that such a large portion of the resulting structures were the smallest possible of only 2 nodes is surprising. This is possibly due to larger polysemous structures being fragmented into smaller structures, because some of the synsets did not have lexical units with matching orthography. An option to skip such synsets that do not have any lexical units with matching orthography during the formation of a polysemous structures is needed to verify this. With such a option the resulting structures could potentially provide a better overview of the occurrence of these structures.

## 4.2 Verifying the results

In order to acquire an understanding of the effectiveness of the produced algorithm it is necessary to analyse the validity of the resulting polysemous graph structures. For this purpose, during the runtime of the algorithm all the times the polysemous words are identified to be in a polysemous relationship are counted and stored in list of tuples. Correspondingly the occurrences of lexical units in the wordnet are counted. These two values are then compared for each word.

Table 4. Number of occurrences of words

| **Word** | **Nr in EstWn** | **Nr in graphs** | **Nr not in graphs** |
|---|---|---|---|
| tee | 12 | 9 | 3 |
| tühi | 12 | 0 | 12 |
| andja | 12 | 0 | 12 |
| vaba | 12 | 0 | 12 |
| saamine | 13 | 4 | 9 |
| andma | 13 | 0 | 13 |
| pesa | 13 | 3 | 10 |
| kindel | 13 | 0 | 13 |
| ajaja | 13 | 0 | 13 |
| pidama | 14 | 6 | 8 |
| andmine | 15 | 0 | 15 |
| pidamine | 15 | 0 | 15 |
| võtma | 16 | 0 | 16 |
| ajama | 16 | 0 | 16 |
| ajamine | 16 | 0 | 16 |
| võtmine | 17 | 0 | 17 |
| minema | 17 | 5 | 12 |
| minemine | 17 | 2 | 15 |
| käimine | 23 | 0 | 23 |
| käima | 24 | 2 | 22 |

Table 4. number of occurrences of words

There are multiple possible reasons for the discrepancies in these numbers, such as:

1. The synsets with these lexical units not having any *hyperonymy/hyponymy* relationships and therefore no vertical polysemy can be present.
2. The synsets with which there are IS-A relationships do not hold any lexical units with matching orthography.
3. The wordnet is incomplete and missing one or both of the rows that should be holding the *hyperonymy/hyponymy* relations and therefore the vertical polysemy could not be detected.

# 5 Unresolved Issues and Future works

Due to this paper being written as a bachelor's graduation thesis, the limited scope of such work and time constraints some features were not implemented.

**Plotting** is one area that could be improved upon. Pyplot which was used in this work is a simple yet powerful tool for drawing simple graphs, but in the future a specialized application might be used for this purpose. The current graph does not have the possibility of easily showing different information based on the current needs of the user, also it does not support interactivity, such that the nodes in the graph might be individually moved and arranged for a more easily readable plot.

Furthermore, the method for signifying a polysemous relationship in the plot, could be improved upon by adding different colours for every different polysemous word in a connected subgraph.

In addition, the **visualisation of hierarchy** between different nodes could significantly be improved upon. Currently the only indication of hierarchy between different nodes in the graphs are arrows at the end of edges. A solution where nodes were shown at a different level relative to others depending on the hierarchy would be helpful to reading the graphs.

As previously mentioned a **method for skipping synsets/nodes** during the detection and formulation of polysemous structures should be studied. Such an option has the potential to reduce the total number of polysemous structures detected from a wordnet, by combining some of the structures that appear to be fragmented due to connecting synsets not holding and lexical units with matching orthography. The resulting structures would give a better overview of the occurrence of polysemous structures in wordnet.

# 6 Conclusion

Machine-readable WordNet-type dictionaries (or wordnets) are widely used for different NLP tasks, in particularly where semantic analysis is needed. However, wordnets often suffer from too fine-grain distinction of polysemous words. Algorithms for reducing polysemy have been proposed, but they tend to be limited in scope.

This work took into consideration a narrower but previously unrealized approach to study polysemous structures. With the program created for this thesis, polysemous structures of the Estonian wordnet version 70 were detected. These are the substructures of semantic hierarchies for nouns as well as verbs, and in which lexical units with the same orthography are connected by an IS-A (i.e., *hyperonymy*) relationship as well as a relationship of vertical polysemy.

The resultant substructures show to a lexicographer the spots in semantic hierarchies where the use of wordnet in NLP processes might turn out to be obviously problematic due to a lexical unit (words or multi word expressions) large ambiguity, especially in cases where it is necessary to annotate different words based on their definition. For this reason, it is planned to forward the results to the workgroup of the Estonian Wordnet at the University of Tartu.

The created programs are also applicable to wordnets of other languages, of which over 70 currently exist in the world.

# References

[1] "Estonian Wordnet," [Online]. Available: https://www.cl.ut.ee/ressursid/teksaurus/. [Accessed 22 May 2019].

[2] "Wordnet | A lexical Database for english," [Online]. Available: https://wordnet.princeton.edu/. [Accessed 22 May 2019].

[3] J.-Y. Y. C.-H. T. S.-C. H. C.-R. Lin, "The Structure of Polysemy: a Study of Multi-sense Words Based on WordNet," in *Proceedings of the 16th Pacific Asia Conference on Language, Information, and Computation*, 2002.

[4] "Homonym | Definition of Homonym by Merriam-Webster"," [Online]. Available: https://www.merriam-webster.com/dictionary/homonym. [Accessed 22 May 2019].

[5] "[EKSS] "Eesti keele seletav sõnaraamat"," [Online]. Available: https://www.eki.ee/dict/ekss. [Accessed 22 May 2019].

[6] A. G. F. D. B. Freihat, "Regular Polysemy in WordNet and Pattern based Approach," *International Journal on Advances in Intelligent Systems,* vol. VI, pp. 199-212, 2013.

[7] A. Lohk, A System of Test Patterns to Check and Validate the Semantic Hierarchies of Wordnet-type Dictionaries, 2015.

[8] "Metaphor | Definition of Metaphor by Merriam-Webster," [Online]. Available: https://www.merriam-webster.com/dictionary/metaphor. [Accessed 22 May 2019].

[9] A. Koskela, "Metonymy, category broadening and narrowing, and vertical polysemy.," in *Defining Metonymy in Cognitive Linguistics: Towards a consensus view,* Amsterdam, John Benjamins Publishing Co., 2011, pp. 125-146.

[10] "WordNet Search - 3.1," [Online]. Available: http://wordnetweb.princeton.edu/perl/webwn. [Accessed 22 May 2019].

[11] "Turning Wordnet into an Information Retrieval Resource: Systematic Polysemy and Conversion to Hierarchical Codes," *International Journal of Pattern Recognition and Artificial Intelligence,* no. 17, pp. 689-704, 2003.

[12] "NetworkX - NetworkX," [Online]. Available: https://networkx.github.io/. [Accessed 22 May 2019].

[13] "Usage Matplotlib 2.0.2 documentation," [Online]. Available: https://matplotlib.org/faq/usage_faq . [Accessed 22 May 2019].

[14] L. S. Sterling, The Art of Agent-Oriented Modeling, London: The MIT Press, 2009.

[15] A. G. F. D. B. Freihat, "Approaching Regular Polysemy in WordNet," in *eKNOW 2013, The Fifth International Conference on Information, Process, and Knowledge Management*, 2013.

# Appendix 1 – AlgorithmForDetectingPolysemy.py

The program with the algorithm developed for this thesis can be found in the included .zip folder.

# Appendix 2 – output.pdf

The output file with the created structures can be found in the included .zip folder.