A Translation-based, Automatically Derived Thesaurus of Lithuanian Adjectives

An Explorative Study

Agnė Nugaraitė

UNIVERSITETET I BERGEN

UNIVERSITETET I BERGEN

Institutt for lingvistiske, litterære og estetiske stadium

DASP350
Masteroppgave i datalingvistikk og språkteknologi
Vårsemester 2014

A Translation-based, Automatically
Derived Thesaurus of Lithuanian Adjectives

Agnė Nugaraitė

Abstract

In this master thesis, I applied the Semantic Mirrors method to the translational correspondences that I have gathered from the parallel corpus. The corpus that I was using was the English-Lithuanian parallel corpus that was borrowed from Vytautas Magnus University in Kaunas, Lithuania, particularly for this research. The data was implemented in the corpus platform Corpuscle at the University of Bergen. Using the Semantic Mirrors system I have automatically generated the explorative version of thesaurus. It contains a great amount of valid semantic relations among adjectives. Combined with manual human interaction or further automatic improvements, this explorative study might lay the foundation for a reliable thesaurus of Lithuanian adjectives.

Acknowledgements

Table of Contents

1 Introduction

The developing needs of language technology have led to an increased interest in electronically avilable multilingual dictionaries, WordNets and other lexical resources for different languages. However, the biggest number of online dictionaries are for the English language. Still, the EuroWordNet project has produced WordNets for several European languages and linked them together; these are not freely available however.

Languages of small countries either do not have online language reference books at all or only a low number of sources exist. As some languages have quite poor language resources (as my native tongue Lithuanian), my master thesis deals with this problem. There are only few Lithuanian language online dictionaries. All online sources are to a large extent manually gathered and formed – this requires much effort and it is a time and budget consuming process. Besides, in most cases the sources are not freely accessible.

Making thesauri or other web dictionaries is a laborious task, and thus big efforts are being made to automate the process. There is a clear need for methods to extract thesauri automatically or tools that assist in the manual creation and updating of these semantic resources. In this thesis I will proceed with translation-based approach that is meant to automatically construct thesauruses. The professor from Bergen University, Dyvik (who was my Master thesis supervisor) invented *the Semantic Mirrors* method as a means for automatic derivation of thesaurus entries from a parallel corpus.

Specifically, in my master thesis, I intend to try *the Semantic Mirrors* approach and to automatically generate the thesaurus of some Lithuanian adjectives. For the consistency of the experiment one part of speech was examined (the adjectives). Because during previous experiments of other researchers that have been examining *the Semantic Mirrors* method (Dyvik 2002, 2005) the adjectives get better results than nouns and verbs for instance (Dyvik 2005).

My task was done in stages. The first stage was collecting adjectives. The adjectives were being collected from the data extracted from the *English-Lithuanian Parallel Corpus* (http://tekstynas.vdu.lt/). Since this corpus is not word aligned nor contains the option of word

filtering, its data was implemented to the corpus platform *Corpuscle* of the University of Bergen, developed by Paul Meurer (http://iness.uib.no/korpuskel/main-page).

Starting with the adjective "*brilliant*" I have collected all the possible Lithuanian translations in the corpus and written them down. The 2nd stage was to collect all the possible translations of each of the translations that were derived from "*brilliant*". For example, the first adjective that translates "*brilliant*" to Lithuanian is "*nuostabus*". So I needed to collect all the translations of "nuostabus". To have a reliable data set that could possibly bring fruitful results examined on the Semantic Mirrors system it is recommended to collect translations from four levels, i.e. starting from "brilliant" (A) I collect all its translations (B) into Lithuanian and this structures the first level. The second level is structured when the translations of "brilliant" (B) gets all its own translations collected (C). The third level is being combined when I collect all the translations (D) of the previous translations (C), and the fourth level is finished when I collect all the translations (E) of the previous translations (D).

When the data sets of both language pairs were ready they were imported to SM system for its calculations of semantic relatedness of words which were derived from translations.

The semantic relatedness of automatically generated thesaurus like units was analyzed and compared with my chosen golden standard of semantic information that was generated manually (LKZ (http://www.lkz.lt/) and DLKZ (http://dz.lki.lt/)).

Using *the Semantic Mirrors* system I have automatically generated the thesaurus like semantic relations between some Lithuanian adjectives. The recall was rather high in many cases, but the adjectives listed as the noise brought in the challenging part to value the results with highest rates. However, after analyzing the results I conclude that the selected method is able to yield valuable results. I also discuss some aspects of reliability of manually generated sources of semantic information (the golden standard of the analysis of my thesis) compared with the results generated automatically (cases I "*šaunus*", II "*puikus*", III "*didelis*" VI "*geras*", VII "*tinkamas*", XIII "*dailus*"). Nevertheless, there is a high possibility that the *Semantic Mirrors* has generated one of the first experimental thesaurus of adjectives in the Lithuanian language.

2 Problem Statement: Language Resources

The Lithuanian language (which kept many archaic features) is spoken just by 4 million people in the whole world. It is one of 23 European Union official languages. The existing Lithuanian language online dictionaries are going to be discussed later in this chapter.

Language resources consist of multi-language corpuses, dictionaries, thesaurus dictionaries, reference books, various WordNets, MultiWordNets, Thesauri, etc. All these sources are to a large extent manually gathered and formed – this requires much effort and it is a time consuming process.

The first Thesaurus was created in 1911 by Roget. Thesaurus is a dictionary which lists lexical entities, in which words that have the same or similar meanings are grouped together; it also provides related words, synonyms, and may contain examples. In general usage, a **thesaurus** is a reference work that lists words grouped together according to similarity of meaning (containing synonyms and sometimes antonyms), in contrast to a dictionary, which provides definitions for words, and generally lists them in alphabetical order. The main purpose of such reference works is to help the user "to find the word, or words, by which [an] idea may be most fitly and aptly expressed" – to quote Peter Roget, the architect of the best known thesaurus in the English language (http://en.wikipedia.org/wiki/Thesaurus).

One of the classical examples of the WordNet is the one created in the Princeton university. The development of this WordNet began in 1985. The WordNet is a lexical database for the English language. It groups words into sets of synonyms called synsets, provides short, general definitions, hierarchical subordination, and records the various semantic relations between the synonym sets. Besides, it gives usage examples. The purpose of the WordNet was the development and investigational means of experimental psychology. However, it turned out to be the valuable resource for the linguists. They started using the WordNet in the twofold way: to produce a combination of dictionary and thesaurus that is more intuitively usable, and to support automatic text analysis and artificial intelligence applications.

Lithuanian language resources are really scarce. During the search for material for my master thesis I found the following online resources: the modern Lithuanian language corpus and the

English- Lithuanian-Czech Parallel Corpus (http://tekstynas.vdu.lt/tekstynas/menu?page=about), Dabartinės Lietuvių kalbos žodynas (Modern Lithuanian Dictionary) (http://dz.lki.lt), Lithuanian scientific corpus CorALit (http://coralit.lt), Lithuanian WordNet (http://metashare.dfki.de/repository/browse/lithuanian-wordnet) (but it seems to be not freely accessible).

Lithuanian WordNet is a lexical database including information about semantic relations of Lithuanian words. (It is aligned with the Princeton 3.0 WordNet). However, the problems and limitations are that unlike other dictionaries, WordNet does not include information about etymology, pronunciation and the forms of irregular verbs and contains only limited information about usage.

Due to the limited availability of WordNets and Thesauri the search of words and any corpus or digital texts based researches can be complicated. It is advisable to search for words or word combinations both in Thesaurus dictionaries and Wordnets. Words and relations which are not included in WordNet can be found in the corpus-derived thesauri. Besides, each type of Thesaurus (Corpus-based Thesaurus, Co-occurrence-based Thesaurus and Syntactically-based Thesaurus) has different characteristics and combining them provides a valuable resource to expand the query (Mandala et al., 1999).

As it was mentioned before, thesauri and WordNets for the Lithuanian language are just a few. The modern Lithuanian language corpus contains 140 921 288 words, the English- Lithuanian-Czech Parallel Corpus contains 2 643 484 words (536267 cz-lt, 2024999 en-lt, 21064 lt-cz, 61154 lt-en) (http://tekstynas.vdu.lt/page.xhtml?id=parallelCorpus), the Lithuanian scientific corpus CorALit contains 9 million words. The Lithuanian WordNet is not finished at the moment, it is under development.

2.1 Approaches

In addition to the before mentioned manually formed (in most cases) language resources, there have been attempts of semi-automatic or automatic derivation of thesaurus-relevant information. Making thesauri or other web dictionaries is a laborious task, and thus big efforts are being made to automate the process (Automatic thesaurus generation from raw text using knowledge-poor

techniques; Curran and Moens; Improvements in Automatic Thesaurus Extraction). Besides, thesauri tend to suffer from problems of bias, inconsistency and limited coverage. In addition, thesaurus compilers cannot keep up with constantly evolving language use and cannot afford to build new thesauri for the many sub-domains that NLP techniques are being applied to. There is a clear need for methods to extract thesauri automatically or tools that assist in the manual creation and updating of these semantic resources (Curran and Moens, 2002).

Lack of the electronically encoded semantic knowledge is a major obstacle in natural language applications of computers. General lexical databases such as WordNet provide limited coverage of restricted domains; domain-specific thesauri are rarely available for a given field. It is as well hard to keep manually-maintained thesauri up to date. Thus, automatically constructed thesauri offer a potential solution. They are usually built by analysing large document collections, employing statistical methods to identify concepts and semantic relations. However, the complexity of natural language and the limited possibilities of language technology means that such thesauri are inferior to manual ones in terms of accuracy and conciseness (Milne et al., 2006).

Automatic thesaurus construction approaches can be purely statistical techniques, or the alternative is to use linguistic methods. Most approaches rely on similarity (for example similarity of meaning or similarity of context). Also, there are a variety of derivations or combinations of different methods to construct thesauri automatically.

Two of the early attempts (1990) to automatize the construction of thesaurus were made by Crouch and by Hindle. Crouch presented the algorithm which clusters terms. This algorithm clusters the documents and is based on the low frequency terms in these clusters in the generated thesaurus like information. The main idea of Hindle's way to automatize thesaurus construction was to deal with the data sparseness problem of low frequency words. "The method estimates the likelihood of unseen events from that of "similar" events that have been seen before. The events are the verb/direct-object pairs of predicate-argument structure." (Park and Choi, 1996).

Most systems that are made for thesaurus construction extract co-occurrence and syntactic information from the words surrounding the target term, which is then converted into a vector-

space representation of the contexts that each target term appears in (Pereira et al., 1993). Other systems take the whole document as the context and consider term co-occurrence at the document level (Crouch, 1988). Once these contexts have been defined, these systems then use clustering or the nearest neighbour methods to find similar terms.

Much of the existing work on thesaurus extraction and word clustering is based on the observation that related terms will appear in *similar* contexts. These systems differ primarily in their definition of "context" and the way they calculate similarity from the contexts each term appears in (Curran and Moens, 2002). Further, the *co-occurrence data of terms* is analysed. This approach relies on the assumption that terms occurring in similar contexts are synonymous. The contexts of an initial term are represented by terms frequently occurring in the same document or paragraph with the initial term. (Ruge, 1997).

Alternatively, some systems are based on the observation that related terms appear *together* in particular contexts. These systems extract related terms directly by recognizing linguistic patterns which link synonyms and hyponyms (Hearst, 1992).

Hyponyms were extracted from large text corpora by Hearst (1992). She searched for relations directly mentioned in the texts and discovered text patterns that relate hyponyms. Certain expressions that usually would involve hyperonyms and hyponyms were identified. For example, the syntactic analysis of the text took place in order to find the instances of certain expressions (such as "*such as*", "*or other*") that generally involve hyperonyms-hyponyms relations. Example:

"A dog, *such as* French Bulldog..."

"Bruises, wounds, broken bones *or other* injuries…" (Hearst, 1992)

It is noticeable, that the first part of the first sentence (a dog) is the hyperonym of the last part of the sentence (French bulldog) and the middle part, the "*such as*" expression is treated as an indicator of hyperonyms-hyponyms relations. Similarly, the expression "or other" is concluded to generally indicate the hyperonyms-hyponyms relations: "*bruises*", "*wounds*" and "*broken*

bones" at some level can be understood as the kind of "*injuries*", meaning that "*bruises*", "*wounds*" and "*broken bones*" are the hyponyms of the "*injury*" (Hearst, 1992).

Now I will proceed with translation-based approach in this work which is especially of interest for linguists. Dyvik (1998, 2003, 2004) invented *the Semantic Mirrors Method* (discussed in Chapter 3 of this thesis) as a means for automatic derivation of thesaurus entries from a word-aligned parallel corpus. This word alignment is not trivial because languages can differ significantly with respect to grammar and syntactic ordering. Computational linguists have developed a variety of statistical algorithms for such word-alignment tasks. The Semantic Mirrors Method intends to extract semantic information from bilingual corpora, which are large collections of texts existing in two languages and which are aligned according to their translations (sentence to sentence or, more rarely, word to word). The assumption is that if two sentences in two different languages correspond translationally, then it should frequently be possible to align some words or phrases (or "lemmata") in the source sentence with the corresponding words or phrases in the target sentence.

An advantage of using bilingual dictionaries instead of parallel corpora is that bilingual dictionaries are freely available on the Web whereas word-aligned parallel corpora are rare and not generally available. A disadvantage of using bilingual dictionaries is that the semantic information which can be extracted from them is less complete, at least with respect to the creation of Semantic Mirrors. Therefore, Priss and Old offer to apply *conceptual exploration* (cf. Stumme, 1996) in addition to Semantic Mirrors to achieve more exact results. Conceptual exploration is a semi-automated process. It can be used to improve the incomplete information extracted from bilingual dictionaries. However, the authors suggest using these methods in ontology engineering and ontology merging (ontologies are much smaller than bilingual dictionaries).

In translations, when a word-aligned parallel corpus is presented, it is possible to extract the set of alternative translations for each lemma in the corpus. The result is an intricate network of translational correspondences uniting the vocabularies of the two languages. This network allows us to treat each language as the 'semantic mirror' of the other (Dyvik, 2005).

A translation I should say, is a rather subjective process as the translator translating from one language to another evaluates the text, uses the meanings as he/she understands. He/she may use synonyms he/she likes or he/she thinks this is the best way to express certain ideas. And as Dyvik wrote in his work, precisely because the perfect translation is impossible, actual translations can tell us a lot about semantics. In a translational approach, the semantic representations must be designed so as to capture the intricate network of translational approximations (Dyvik, 2005).

Furthermore, the criteria for correspondence between the translation and the original are not only that two words have the same meaning but also that they play the same role in the sentence. So the two words actually correspond if their surroundings in the sentence correspond in a certain way. And if the word in one language corresponds to more than one word in the other language then there is an entire set of correspondences. Since there will probably be few one to one correspondences in the translations, I can expect to extract some amount of semantic information based on translational relations.

As we will see in the following chapters the Semantic Mirrors method appears to be able to give some useful results while there is also the problem of noise.

3 Overview of the Semantic Mirrors Method and its Application

The demand of lexical and semantic knowledge in Natural Language Processing applications has spurred initiatives for resource development in recent years. Some of these attempts are aimed at the development of multilingual semantic resources. In this context, experiments were done with various translation-based methods. Plausible results were expected by using translations as sources of information about semantics (Apidianaki and Sagot 2012).

The use of manual methods in retrieving the semantics is characterized as budget and time consuming. These disadvantages led researchers to methods for the automatic development of word-nets. One of these is Helge Dyvik's 'Semantic Mirrors'. This method is meant to automatically derive thesaurus entries from a parallel corpus. In the early stages of testing, the Semantic Mirrors method provided useful results and proved that it merits further exploration. (Dyvik 2002, 2008). There are five papers written in order to present the method. The basic idea that associates papers written by Dyvik is to define a procedure for derivation of semantic representations in the form of feature matrices which are derived from the patterns of translational relations between two languages.

3.1 Semantic Mirrors: Presentation of the Method

The Semantic Mirrors method was worked out by Dyvik in the year 1998 and following years. Its development was inspired by ideas about meaning in translation. Translators evaluate meaning relations between expressions, not as a part of some meta-linguistic, philosophical or theoretical reflection, but as a part of the normal and common linguistic activity of translation (Dyvik 1998b). This circumstance makes the translational relation emerge as epistemologically prior to more abstract and theory-bound notions such as 'meaning' and 'synonymy', which supports the idea of using bilingual corpora as a plausible source of data for deriving semantic relations (Dyvik 2005). This method appears to be innovative, brings new potential for the derivation of semantic features and requires deeper investigation.

The exploration of the method and its functions begins with the paper '*Translations as Semantic Mirrors*' (1998b) which goes into details of basic concepts of translational relations, the distinction of full and relative synonymy, the distinction of ambiguous, and vague, and underspecified signs in the patterns of t-relation - these are the important terms to be defined

exactly because they will be mentioned not once in the paper and also are important while explaining the Mirrors method. The meaning of *full synonymy* in our case is the identical translational properties with respect to all possible target languages. The *relative synonymy*, on the other hand, hold with respect to subsets of target languages.

The *ambiguous signs* are characterized as that in some sense it is 'accidental' that it is a cluster property of one word in a language that happens to be associated with two or more different meanings. Apart from ambiguous signs there are the *vague signs* whose alternative interpretations seem to have more to do with different aspect of what is being denoted (like in the adjective 'good'), and hence less 'accidental'. Vague words are clustered within the family of things that have something in common irrespective of language.

The *underspecified signs* that we will be taking into consideration are explained in this way: "The highest-ranking signs are the ones that have the widest range of translational possibilities within the sense concerned, which, it intuitively seems, must be associated with a wide "meaning potential" as compared to the lower-ranking signs. This may mean that they have a kind of 'prototype' status vis à vis the lower-ranking signs, or that they are somehow 'underspecified' in relation to them, as hyperonyms to hyponyms — we will return to the exact nature of the semantic relationship." (Dyvik, 1998b).

After defining terms important to our exploration of the method we may proceed with defining the translational relation between two languages. The translational relation between two languages is being regarded as an epistemological primitive, accessible in actual translations. The idea is to attempt to define semantic properties in terms of this relation. These translational relations are accessible in a parallel corpus which provides a relation between situated texts. The paper '*Translations as a Semantic Knowledge Source*' (Dyvik 2005) reviews the classical structuralist approach to the description of word meanings within the field in the use of componential analysis which is expressed by assigning semantic features to the words, capturing their interrelations. Also the author notes that 'the corresponding sets of terms in two languages are connected by a relation of translation' which motivates the development of Semantic Mirrors method.

Semantic Mirrors' source consists of the translations of corresponding words appearing in a word-aligned bilingual corpus. The procedure derives semantic representations in the form of feature matrices from the patterns of translational relations between semantic fields in two languages, as further defined below.

In general Semantic Mirrors method can be conveyed through explaining the actual operation of t-images and their further grouping into senses, the introduction to the semantic fields, ranking signs in a semantic field and deriving semantic representations. All the steps are mentioned in Dyvik (1998b) and will be presented further in my thesis (p 13-17).

The procedure of assigning semantic features and creating the semantic field begins by extracting the set of alternative translations for each lemma from the word-aligned bilingual corpus. The data that have been extracted are being written down in a way that will be suitable for the Semantic Mirrors system to read it. First, word in L1 is found as the translations in L2 and all the equivalents are written down. Then, the collection of words gathered from L2 one by one are being collected from L1 as the translations of equivalent in L2. For a comprehensive data set it is recommended to collect the words in four levels: L1 to L2, L2 back to L1, a new collection from the latest set from L1 back to L2 and a new collection from L2 back to L1. In this way one creates a multiplex network of translational correspondences between the vocabularies of two languages.

A possible data sample is presented:

```
("aiškus" "clear" "obvious" "definite" "evident" "intelligible" "sharp" "distinctive" "clean"
"plain" "overt" "unmistakable")
("akinantis" "dazzling" "blinding" "perfect" "blazing" "intense")
("akivaizdus" "apparent" "self-evident" "clear" "obvious")
("aktyvus" "active" "high" "widespread" "proactive" "interactive" "dramatic"
"intervening" "reactive" "intense" "robust")
("amžinas" "eternal" "timeless" "infinite" "permanent" "unalterable")
("antgamtiškas" "supernatural")
("aršus" "fast" "bad" "fierce" "relentless" "urgent" "ferocious" "insolent" "savage"
"angry" "rabid" "bitter" "extreme")
```

Listing. 3.1 A possible data sample containing adjectives of L2 and their translations in L1

What we see in the figure above is that the first word in each row is taken from, let us say, L1 and all the rest of the row is the translations of the first word, in L2. The words have been organized as follows: for some Lithuanian words in the source text their translations in the target

text have been found. The same procedure was done for new words that were noted down as the translations for Lithuanian words - they were searched for in the English source texts and their translations collected from the target texts. This possible data sample represents the words from L2 which were collected as translations of the word in L1. This form of data can be now read by the Mirrors system which calculates from it some semantic relations between words in L1 or between words in L2.

The corpus that was used in this case to collect a possible data sample is reachable on line http://iness.uib.no/korpuskel/main-page. The data originally was taken from the Lithuanian corpus which can be reached via this link http://tekstynas.vdu.lt/page.xhtml?id=parallelCorpus. The '*Corpuscle*' corpus contains 9,169,274 English tokens translated by translators into Lithuanian language and 7,727,039 tokens in Lithuanian language and their translations into English.

The network of one language's correspondences (as we saw in the figure above) is being treated as the 'semantic mirror' of the other as we state five assumptions (Dyvik 2005, 2008):

"1. Semantically closely related words tend to have strongly overlapping sets of translations.

2. Words with wide meanings tend to have a higher number of translations than words with narrow meanings.

3. If a word sense *a* is a hyponym of a word sense *b* (such as *tasty* of *good*, for example), then the possible translations of *a* will probably be a subset of the possible translations of *b*.

4. Contrastive ambiguity, i.e., ambiguity between two unrelated senses of a lemma, such as the two senses of *band* ('orchestra' and 'piece of tape'), tends to be a historically accidental and idiosyncratic property of individual lemmas. Hence we do not expect to find instances of the same contrastive ambiguity replicated by other lemmas in the language or by lemmas in other languages.

5. Words with unrelated meanings will not share translations into another language, except in cases where the shared target lemma is contrastively ambiguous between the two unrelated meanings. By assumption 4 there should then be at most one such shared lemma."

The first stage in applying the method is the generation of 't-images'. In Dyvik's works, the set of translations of a word x from one language (L1) is the first 't-image' in the other language (L2). Then, taking the first t-images back in L2 of all the members of x's first t-image brings a set of intersecting sets of words in L2. This set of sets is being called 'the inverse t-image' of x. Later, finding the first t-images in L1 of all the members of the set of x's inverse t-image allows us to have a set of intersecting sets of words in L1. These are called x's 'second t-image'. Figure 3.1 illustrates this example where x is "*wood*" and L1 is English, and the equivalent translations in L2 (Lithuanian) are the first t-image. The inverse t-image is generated while taking all the first t-image's units back to L1 and finding their equivalents:
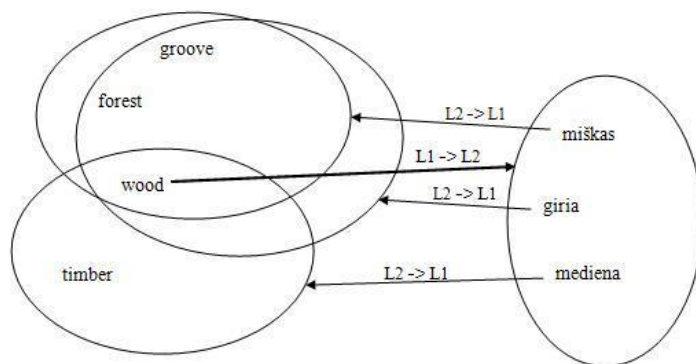

Figure. 3.1 The first (on the right) and inverse (on the left) t-images of the noun "wood"

The second stage is to partition the t-image of a word x into senses. Each partition contains semantically related words. As an example, a t-image of English 'wood' in Lithuanian could be {miškas, giria, mediena}. Intuitively, these three words belong to two senses: the sense of 'wood' as a collection of trees ('miškas' and 'giria') and the sense of 'wood' as a building material ('mediena'). As shown in Figure 2, these senses can be derived automatically by analysing the inverse t-image. In our example, it is assumed that the t-image of "miškas" is {wood, grove, forest}, the t-image of "giria" is {wood, forest, grove}, and the t-image of "mediena" is {wood, timber}. Because the t-image of "miškas" and "giria" overlap in more than one word, they are considered one sense of "wood" denoted by 'wood1'. The t-image of "mediena" overlaps with other two t-images only in the beginning word "wood", "mediena" is considered a second sense of "wood" denoted by 'wood2'.

At the third stage the word senses across different lemmas in each language are grouped into 'semantic fields' based on shared translational properties. Two senses belong to the same semantic field if they have intersecting first t-images (Dyvik 2005). A feature hierarchy is formed based on the set-structure. The semantic fields in L1 and L2 are being grouped symmetrically one-to-one according to the relation that determines field membership. Each field in such a pair projects a subset structure onto the other field (Dyvik 2008). The example shows paired semantic fields from Norwegian and English.
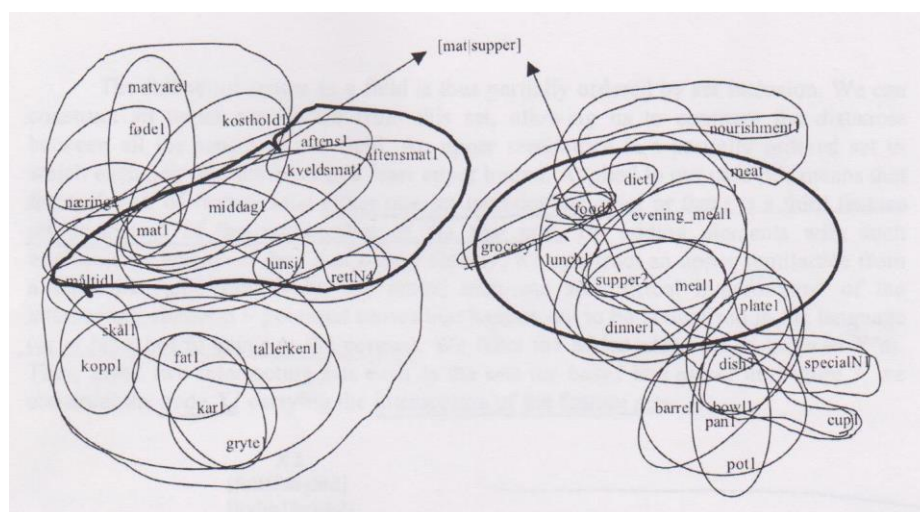

Figure 3.2 Paired semantic fields from Norwegian and English (Dyvik 2005)

By assumption 2, the senses that are a members of many subsets, are expected to have wide meanings. In the figure 3.2 we see that *mat1* ('food') constitute peak in the subset structure. Surprising is that the sense *supper2* happens to outrank the sense *food5* in the English field. Furthermore, the fact that two senses are co-members of many subsets means that they share many translations and hence ought to be closely related semantically (Dyvik, 2005).

At the fourth stage the semantic features are being assigned to each sense in each field, encoding the sense's relationships to the other senses in its field. The aim is to assign few features to wide senses and supersets of those features to their hyponyms. This is achieved through the notion of 'rank' within a semantic field. The number of t-image subsets of which a sense is a member is called its 'rank'. The sense with the highest rank is the 'peak' of a field and assumed to have the widest sense. A feature [a|b] constructed from source sense *a* and target sense *b* is assigned to *a* and *b* and inherited by all members of the first t-images of *a* and *b* which are ranked lower than *b*

and *a* (Dyvik 2008). After this is done, the feature set inclusion now, by hypothesis, reflects a hyponymy/hyperonymy relation. Feature structures are graphically displayed as lattices (Figure 3.3).

food1
[mat1|supper2]
[middag1|food1]

lunch1
[mat1|supper2]
[middag1|food1]
[lunsj1|meal1]
[lunch1]

Listing 3.2 Feature assignment to two senses (Dyvik 2008)

What we see in listing 3.2 is two randomly picked senses by Dyvik (2005). The senses are both derived from the semantic field (Fig. 3.2) with food-related nouns. These examples come from English-Norwegian Parallel Corpus (ENPC). One of them is *food1* and another - *lunch1*. We see that *lunch1* inherited all the *food1* features what makes sense *food1* to carry a subset of the features of its hyponym *lunch1*. From the semantic field one can now construct the lattice.
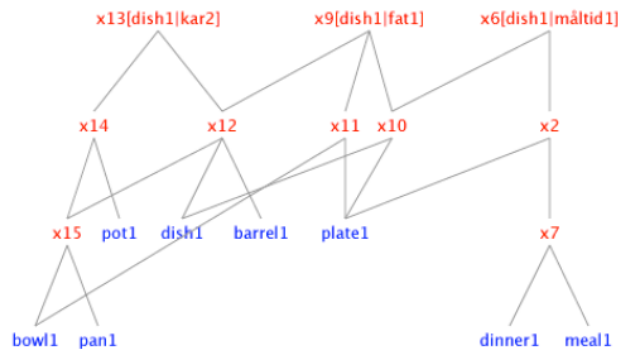
Figure 3.3 A sublatice for nouns (Dyvik 2008)

Feature structures are shown in figure 3.2. We can easily notice how two fields of related words formed in here: one contains nouns like *dinner, meal*, another - *bowl, pan, pot, dish, barrel, plate*. It already refer to two different *food* senses - the vessel and food senses.

At the final stage in the Semantic Mirrors method, thesaurus entries are generated. Hyperonyms, hyponyms and synonyms are the thesaurus entries that the Semantic Mirrors can determine. To

maintain a plausible Mirrors generated hyperonym the user can set a lower bound on its number of hyponyms. The variable called *SynsetLimit* specifies the number of senses that must have inherited a feature *f* constructed from a sense *s* for *s* to be counted as their hyperonym (Dyvik 2008).

The work in determining thesaurus entries is comprehensively described in the paper 'Translations as Semantic Mirrors: From Parallel Corpus to Wordnet' (Dyvik 2002). The possibility of inducing WordNets and thesauri as an output of the Mirrors method is studied. The concept of semantic lattices is presented as the tools to generate the thesaurus entries. The procedure of deriving senses from t-images of translations from one language to another, grouping into semantic fields and feature assignment is presented without major changes (compared to Dyvik 1988b). The ordering of the senses in the lattice is based on feature set inclusion. In order to derive semilattices from the partly overlapping feature sets a new feature called **x-node** is introduced in Dyvik 2002. X-nodes occur when two senses *a* and *b* have intersecting feature sets without either of the sets including the other, and no existing third sense is assigned the intersection of the features. In such cases an x-node is introduced as a 'virtual hyperonym' of *a* and *b*.

Also Dyvik (2002) provides details on the derivation of thesaurus entries which involves determining subsenses, hyperonyms, near-synonyms and hyponyms of each sense on the basis of the information in the semilattices. Semilattices in some cases are extremely complex because they contain 'noise'. It is resulting from accidental biases and gaps in the corpus. For this reason two new parameters are being introduced to the reader: OverlapThreshold and SynsetLimit. OverlapThreshold is the parameter which decides the granularity of the division into subsenses in the thesaurus entry; the parameter SynsetLimit defines the maximal size which the set denoted by a feature can have in order to be included among the near-synonyms.

Testing the method on different language pairs (English-Norwegian, English-Greek, English-German) and different input data confirmed its plausibility. However, there are some conditions for the method to generate reliable semantic relations.

First of all, the method is vulnerable to the increase of noise, such as errors in the automatic alignment of words. Precision and recall in the thesaurus entries from automatically aligned data

seems to be lower than the ones from manually aligned data. A similar finding was found in Thunes' research according to the Mirrors method: "the usefulness of the method is more dependent on 'low noise' than on full coverage.' (Thunes 2003). Therefore, the manually aligned data is the proper input compared with automatically aligned data for Semantic Mirrors method.

The reliability of the method itself is not easily measured because of troubles in finding a suitable gold standard. In Dyvik (2005, 2008) it was written about a comparison of Mirrors derived entries with corresponding entries in Merriam-Webster's Thesaurus as a gold standard. This exemplary evaluation is taken from Thunes 2003. Dyvik (2005) found that "When using Merriam-Webster's Thesaurus or Princeton WordNet as gold standards for the sets of semantically related words associated with the thesaurus entries, precision and recall is low, but not very much worse than the results obtained when we compare the established resources Merriam-Webster and Princeton WordNet with each other." Moreover, Thunes (2003) claims that the Mirrors method gives a more complete picture of a language compared with the gold standard in her study, i.e. the Merriam-Webster's Thesaurus.

Another finding is that the method gives different results for different parts of speech. Adjectives get better results than nouns and verbs, and abstract nouns get better results than concrete nouns (Dyvik 2005).

Another indicator for the plausibility of the results is the distance between the two languages that are being tested. In order to get fruitful results the distance between languages should be suitable (not *too close* and not *too distant*). For example, close languages (like Norwegian and Danish) share a big part of entries in the vocabulary. This hampers the process of mirroring languages within this combination because there will be less alternative translations of a given word. And vice versa, too distinct languages bring the difficulties to obtain high-quality input data because of very complicated, ambiguous, and debatable translations and word alignments.

To conclude we can say that the Mirrors method generates fruitful results with the main condition of high-quality translations being used as an input resource, and a suitable distance between languages.

3.2 A Review of Applications of the Method

There is a body of work that in some way applies or tests the method. For instance, Apidianaki (2008) obtains senses from a parallel corpus by combining contextual (distributional) information and information gained from translations. She uses Semantic Mirrors as a model method for lexical-semantic information extraction from translations taken from a parallel corpus. Lyse (2010) investigates word sense disambiguation in order to evaluate the productivity and reliability of Mirrors method as a semantic knowledge source. She applies the information about word senses derived from the Mirrors method as a knowledge source in a corpus-based machine learning approach to word sense disambiguation. Thunes' (2003) work is based on evaluation of thesaurus entries derived from translational features using the Semantic Mirrors method.

Priss and Old (2005) apply the Mirrors method to an English-German dictionary instead of a parallel corpus. These authors agree that semantic information extracted from a dictionary is less complete (in the terms of Semantic Mirrors method) than extracted from a parallel corpus and they add conceptual exploration (Formal Concept Analysis) to improve the incomplete information from Mirrors method. A review and summaries of the above-mentioned papers follows.

Thunes (2003) systematically evaluates results from applying the Semantic Mirrors method to English-Norwegian data. In her paper she tests the adjective *pleasant*. The translations for the base word are derived from data which was manually extracted from English-Norwegian (ENPC) parallel corpus which is sentence aligned. The author sets Merriam-Webster Online Thesaurus and Princeton WordNet as the standard to evaluate the quality of Mirrors thesaurus entries. Then the precision and recall are being considered with respect to total sets of lemmas listed as semantically related words of an entry across sense descriptions and divisions into subsenses in Webster and WordNet (Thunes 2003). The final evaluation supports the hypothesis of Mirrors method's abilities to provide user with thesaurus like entries. The example of thesaurus generated by Mirrors Method and used in Thunes (2003) paper follows:

pleasant
    (Translation: *god, hyggelig, pen, snill, vakker.*)
    Hyperonyms: *gentle, good.*
    Synonyms: *all right, amiable, benign, friendly, good-humoured, good-natured, jolly, kind, kindly,*
        lovely, mild, pleasing, polite, smiling, soft, sweet.
    Related words: *attractive, beautiful, charming, comfortable, cozy, cute, delightful, dishy,*
        enchanting, fair, fancy, graceful, handsome, happy, magnificent, ornate, picturesque,
        pleasurable, pretty, well.

Listing 3.3 An example of Mirrors' generated example of thesaurus (Thunes 2003)

In the listing 3.3 we can see the thesauri entry for English word *pleasant*. It contains translations into Norwegian, plausible hyperonyms, synonyms and related words. In this case only one sense was identified but partition into senses is also an outcome of the Mirrors method. Antonymy in the Mirrors method is not being retrieved because translations do not reflect the reverse sense of a word.

The paper 'Conceptual Exploration of Semantic Mirrors' (Priss and Old 2005) models the Semantic Mirrors method with Formal Concept Analysis. It investigates to what extent the method is applicable to a bilingual dictionary instead of a parallel corpus. The use of a bilingual dictionary as data for the Semantic Mirrors method did not prove to be very suitable but this method can be improved by applying conceptual exploration. Dyvik (1998b, 2002, 2005, 2008) uses Venn diagrams to draw his examples, but in many cases it may be difficult (or even impossible) to draw complex examples (figure 3). To introduce easier way of representation Priss and Old (2005) use 'neighbourhood lattices' to visualize the first three steps of the Mirrors method. Furthermore, they apply the techniques developed for the neighbourhood lattices by combining the Mirrors method with FCA. For the first step they derive a t-image with its lattice and the neighbourhood lattice (inverse t-image: the translation of the translation). As a second step in modelling the Mirrors method with FCA, the authors identify senses and include a 'contingent' of a concept which is defined as a set of attributes and objects, which are in the extension of the concept (Priss and Old 2005). The difference from the classic Mirrors method in this part is that these attributes and objects belong directly to the concept and are not inherited from sub- or super-concepts.
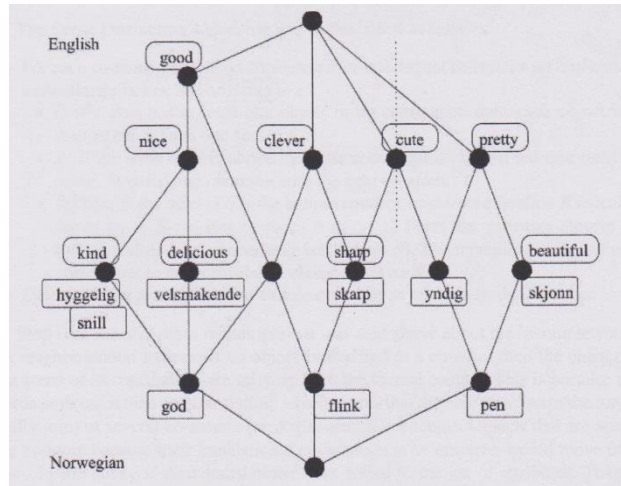
Figure 3.4 An example of a neighbourhood lattice for "good/god" (Priss and Old 2005)

Figure 3.4 illustrates the neighbourhood lattice instead of Venn diagram for words *good* and *god*. The authors got the results which we can see in the figure 3.5: the neighbourhood lattice is almost symmetric with respect to a horizontal line in the centre of lattice ('Semantic Mirror' between two languages).
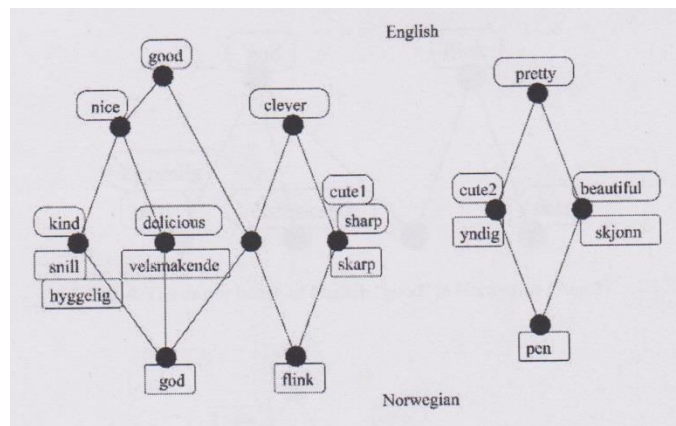

Figure 3.5 Two semantic fields (Priss and Old 2005)

Furthermore, in the figure 3.5 we can notice the division into different semantic fields. One (left) contains adjectives meaning features of character or person itself like *nice, good, clever, cute, sharp* (most likely meaning a sharpness of the mind in this case) or feature of food - *delicious*. On the right we can see features describing the appearance: *pretty, cute* (it can both be the feature of the appearance or character), *beautiful*.
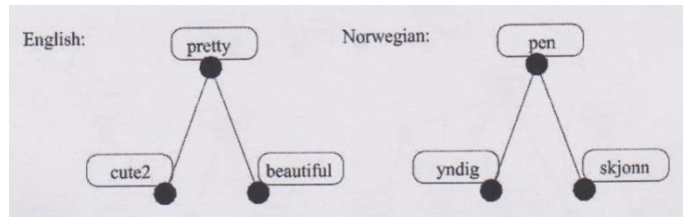
Figure 3.6 The mirror image for "pretty" in English (on the left) and Norwegian (on the right)
(Priss and Old 2005)

Finally, from the neighbourhood lattices mirroring two languages we can take into consideration the plausible example of 'pretty-pen' adjectives (Figure 3.6). Both 'pretty' and 'pen' are connected to two other adjectives.

This paper concludes that the combination of the Semantic Mirrors method and conceptual exploration could give promising results in some areas.

Apidianaki's Doctoral Dissertation (2008) is concerned with the combination of translational data with context based data in order to identify semantic similarity between words. The author is interested in the possibility of disambiguating word senses using a monolingual corpus - the researcher looks for various senses in various contexts. The data is sorted and translational correspondences are added. In a comparison of her method with the Mirrors method (using a Greek-English parallel corpus) she finds an interesting degree of correspondence between the results (familiar grouping of words, division of senses). She concludes that this similarity implies trustworthiness of the results of her own method. The Mirrors results showed up to be compatible with her method's results. The finding that Mirrors method gives similar outcome inspired the author to give better grades to her own method. Therefore she finds it fruitful to combine these two methods.

Paper written by Lyse (2010) combines classical statistical (data-driven) methods and translation-based lexical information from the Mirrors method in order to disambiguate word senses. This approach assumes a correlation between word senses and context words: some words' senses occur more in some context and other words' senses - in other contexts. After this finding, the system then may classify unseen instances of the target word, based on what it has learnt from the training corpus. Since sense information is not overtly present in a text corpus, each training instance must be labelled with its relevant sense prior to learning (Lyse 2010). This

is where the author has combined the Semantic Mirrors method with statistically based machine learning approach. She uses the Mirrors as a knowledge source for WSD. The automatic sense tagging feasible in the Mirrors system serves as a source of senses for Lyse's machine learning approach.

However, apart from using the Semantic Mirrors generated senses in her approach, Lyse also tests the plausibility of these senses. This may involve some circularity for the research hypothesis because the author uses senses that have been derived from Mirrors system and then she tests how plausible the senses derived by the same method actually are.

3.3 General Review

The papers reviewed in this chapter include **different goals** of using the method of our interest. For example, Apidianaki's main goal was *to compare* two methods for the same material (Mirrors method with context-based word sense disambiguation method). Thune's main goal was *to evaluate* the results produced by applying the Mirrors method to the English-Norwegian language pair. Priss and Old while applying the method had a goal *to improve* (adding the change in the development and presentation of stages in the method) the understanding of method with the help of FCA and *to investigate* (using different source of data) the Mirrors method's efficiency in a case of using bilingual dictionary instead of a parallel corpus. The goal of Lyse (2010) was *to combine* classical statistical (data-driven) methods and translation-based lexical information from the Mirrors method.

The authors apply the method in these ways: Thune and Apidianaki use the same patterns of method application as Dyvik in the first group; Priss and Old apply the techniques developed for neighbourhood lattices to semantic Mirrors by modelling the method with FCA; Lyse applies the Mirrors method to her investigation in order to label each training instance with its relevant sense prior to learning.

Some papers (Dyvik 1988b, 2002, 2005, 2008) reviewed in this article have one goal – *to present* the method to the audience with interest. However, the goal of evaluating the method on any significant scale is left for others.

There also is a **different application** of the Mirrors method in the papers overviewed above. In the first group of papers that is pointed to presenting the method, the application of the method keeps stable during all five works. The method is being applied to the English-Norwegian parallel corpus using four steps of t-images derivation, individualization of senses, assigning the features and deriving hierarchy of senses (hyponymy-hyperonymy), and the derivation of thesaurus entries.

One more distinction within the papers is the **different data** that was tested. In the first group Dyvik is testing English-Norwegian and Norwegian-English results. Moreover, the author is referring to different parts of speech - nouns (2002, 2005, 2008) adjectives (1998b, 2002, 2005, 2008).

One other paper, namely that of Thunes is testing the data from the English-Norwegian language pair. The part of speech that is tested is adjectives. Priss and Old are testing the data from English-German bilingual dictionary and investigate adjectives in their paper as well Lyse uses the data from English-Norwegian parallel corpus as an input.

The authors of the papers had different goals, different ways of applying the method, different data, this led to the conclusion of **different results**. Dyvik in his articles (1998a, 1998b, 2002, 2005, 2008) got promising results. Dyvik (2005) claims, that 'the method gives better results for adjectives'. Thunes came out with a conclusion that the method is more reliable when it gets input with 'low noise' rather than it gets full coverage of source. Also, she claims, that it is able to capture the distinction between a wide sense of the word and more narrow senses (Thunes, 2003). Priss and Old came out with the result that the combined method of conceptual exploration and Semantic Mirrors provides a useful toolkit specifically for smaller size bilingual resources (Priss and Old, 2005). Apidianaki also got the plausible results after comparing Semantic Mirrors method with her own method's results.

3.4 Conclusion

Various ways of applying the method, the possibilities to apply the method to different data, positive results and different goals of application within several researchers test its reliability, its potential and wide range of possible usage.

However, using the Mirrors method one has to consider the distance between the language pair, be precise in data collection, and careful when choosing the corpus as a source. If these conditions are satisfied, the researcher is very likely to automatically derive valid, or potentially useful, semantic representations from the patterns of translational relations between semantic fields in two languages.

4 The Corpus of Contemporary Lithuanian Language

This chapter is going to introduce the Corpus of Contemporary Lithuanian Language and the English-Lithuanian parallel corpus with which I have been working. The corpus was compiled at the Vytautas Magnus University by joint effort of the scientists of the Faculty of Philology and the Faculty of Mathematics and Informatics. The aim of the building of the parallel corpus is to accumulate a large database of authentic Lithuanian and English texts which could be used for objective studies of languages, would reveal peculiarities and characteristics of genre variety and would give information about the existing or possible trends of the development of languages and their interpretations. Considering its practical application, this corpus is invaluable in teaching and learning of different languages (http://coralit.lt/en/node/18).

The corpus of contemporary Lithuanian language is a collection of electronic texts provided with special software and is meant for philological, statistical, sociological or other analysis of language usage. It is generally recognized as a data source used by researchers and practitioners in various spheres. This corpus is the largest Lithuanian language corpus and has been developed since 1992.

The current size of the corpus (140 921 288 words) as well as the variety of genres and topics (fiction 11.6%, nonfiction 14.2%, administrative texts 10.0%, publicistic texts 63.8%, spoken language 0.3%, 0.1% - other) determine a wide spectrum of its application possibilities. It can be used as an encyclopedic or contextual dictionary or as a practical means for language teaching and learning for those who are searching for numerous authentic examples of the Lithuanian language or some other languages, too (http://tekstynas.vdu.lt/tekstynas/menu?page=statistics).

The English-Lithuanian parallel corpus is the part of the Corpus of Contemporary Lithuanian Language which is being used to collect the necessary data in the research described in this thesis. It is explicitly described in the following (4.1) part.

4.1 The English-Lithuanian Parallel Corpus

The parallel corpus may first of all be used searching for information that is not given or due to the limited size is not included in traditional bilingual dictionaries. Most equivalents of translation are contextual, therefore it is impossible to put them all into dictionaries. In this case

it is advisable to use parallel corpora where the wider variety of translation equivalents can be found. The corpus data show that in actual translations there exist important usage differences between apparently synonymous equivalents presented as interchangeable items. A parallel corpus thus can provide a more extensive inventory of cross-linguistic correspondences than a bilingual dictionary (Ruzaite, 2010). The parallel corpus can be especially useful for making the quality of translations better. It helps the translator to choose the closest translation equivalent in the concrete context.

The Parallel corpus prepared in the Centre of Computational Linguistics is the first accessible on the internet in Lithuania, being as well the biggest in its amount. Both theorists and practitioners can access it easily, namely dictionary writers and the specialists of language teaching can use it as an objective source of language usage and translation. As well as other parallel corpora of such type, this corpus may be used as a valuable means for drawing language parallels and as a means for comparison, teaching, creating automatic translation, and to compile bilingual, multilingual and term dictionaries.

In the Centre of Computational Linguistics at Vytautas Magnus University the English-Lithuanian parallel corpus was compiled in 2000 and the compilation is still in progress. It is compiled in accordance with the modern theory and practice of corpus compilation and following the TEI P5 (Text Encoding Initiative) text encoding guidelines. Since 2005 the Parallel Corpus has been freely accessible on the internet at http://donelaitis.vdu.lt or at http://tekstynas.vdu.lt.

The Parallel Corpus comprises two parts: English-Lithuanian and Czech-Lithuanian. The whole English-Lithuanian parallel corpus contains 2024999 words (Fiction – 40.9%, Non-fiction – 22.5%, Internet texts – 6.1%, Documents – 26.7%, Press – 3.4%, User Manuals – 0.1%, and Other texts – 0.3%). All texts are in XML format (Vytautas Magnus University, The Centre of Corpus Linguistics http://tekstynas.vdu.lt/page.xhtml?id=parallelCorpus). The information about this parallel corpus was collected directly from the people working at the Vytautas Magnus University, The Centre of Corpus Linguistics. To collect the information about parallel corpus was rather challenging task because of information's absence in the online sources. There is no ready information on the web page of the corpus, no published papers explicitly describing the organization and content of this corpus, too.
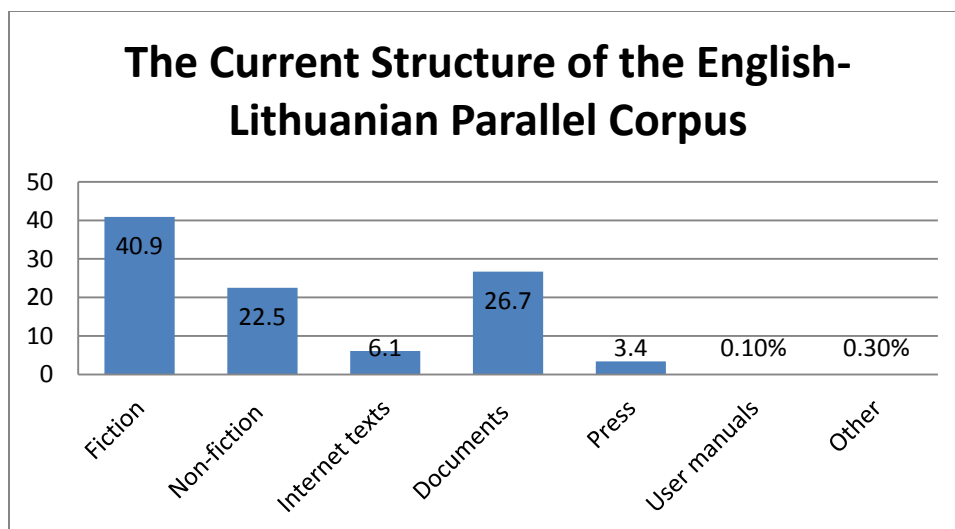
Figure 4.1 The Current Structure of the English-Lithuanian Parallel Corpus

The internet access to parallel corpora allows reaching only part of the English-Lithuanian and Lithuanian-English corpora, while the rest of the data is being prepared at the moment. However, in this thesis the full set of English-Lithuanian parallel corpus was borrowed from Vytautas Magnus University (VDU), The Centre of Corpus Linguistics and used for the research.

The search was not done in the Parallel corpus from VDU because the corpus is not word-alligned. This complicated the manual search which I was performing in order to collect translational correspondences. For this reason, the VDU English-Lithuanian Parallel Corpus was borrowed and implemented into the Corpuscle at UIB. There, my manual search was facilitated with the possibility of *filtering* (filtering procedure is explained in the Chapter 5.1 (p. 29).

As for the research which is the main interest of this paper, parallel corpora are indispensable tools for collecting a wide variety of translation equivalents which are being used in the language and translated into another one under consideration of the context. The translations of particular tokens in the context not always represent only tokens noted in the dictionaries. A much wider spectrum of translation possibilities is being found using the parallel corpus compared with the dictionary. Also, using a parallel corpus gives a more natural view of language usage. That is why the research which is going to be presented in this Master's thesis will be based on the most fruitful data which reflect the translation relations of adjectives in Lithuanian and English.

5 Data Collection Process

In the Master thesis Corpuscle (Korpuskel) (http://iness.uib.no/korpuskel/main-page) has been used in the data collection process.

Corpuscle is a corpus management platform for annotated corpora. It is a new corpus query engine and corpus management system. The tool has been developed by Paul Meurer at Uni Computing (http://www.computing.uni.no/?session-id=234905593953605#_blank), in collaboration with colleagues at the University of Bergen. The work has been supported by grants from NFR and the Meltzer foundation.

The site hosts several types of corpus collections:

- Open-access corpora
- Restricted-access corpora
- CLARINO corpora: a collection of language corpora that are part of the Norwegian CLARINO infrastructure.

The Corpus list consists of the Norwegian NewsPaper-corpus ann. (Part of the Norwegian NewsPaper-corpus, grammatically annotated and classified); Wikipedia ENG (English Wikipedia set from April 2010, containing all articles with more than 2000 words); and Аҧсуа (texts from www.abkhaziagov.org). Some corpora are only available when you have signed in. The information obtained from the English- Lithuanian Parallel Corpus http://tekstynas.vdu.lt/tekstynas/menu?page=about was structured into the Corpuscle http://iness.uib.no/korpuskel/main-page of the University of Bergen by Meurer and I have had the possibility to try and use the Corpuscle system with the Lithuanian corpora while collecting the data.
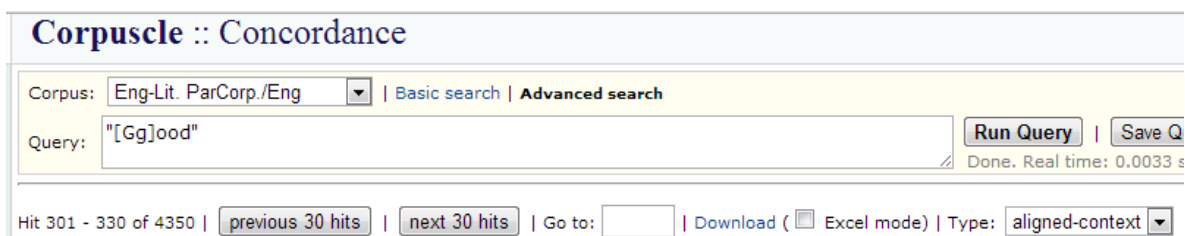
5.1 Searching for Translational Correspondences

For the consistency of the experiment one part of speech, adjectives, was selected to be examined - because during the previous experiments of other researchers that have been examining the Semantic Mirrors method (Dyvik 2002, 2005; Apidianaki 2008) the adjectives get better results than nouns and verbs for instance (Dyvik 2005).

The Semantic Mirrors program takes as an input translational correspondences on a lexical level. To be able to search for words manually in the parallel corpus, ideally we need to have a corpus that is aligned on a word level. Since the corpus that I have been using is not word-aligned, but sentence-aligned, while searching I needed to perform a rather complex procedure called *filtering*. This procedure is explicitly explained in the following part.

Searching for the Lithuanian adjectives. Filtering

The process of data collection was carried out as follows. First of all, in the Concordance you have to choose the required or needed language search option, for example Eng-Lit ParCorp/Eng or Eng-Lit ParCorp/Lit, or the Norwegian NewsPaper Corpus, etc. While collecting the data I chose the relevant language search which is Eng-Lit ParCorp/Eng or Eng-Lit ParCorp/Lit. It depends on which side of the Corpus Lithuanian or English you are going to search for the words. Some texts are in Lithuanian, other texts are translated into English, and some other texts are translated from English to Lithuanian. Then, in the option *Type,* the aligned context is chosen. In the case of searching in Eng-Lit ParCorp/Eng corpus the English word that you are searching for is written in inverted commas in the search query. Since the English adjectives are not inflected except for comparison, only the main form of the word is searched for, not including degrees of comparisons of adjectives because the main form in English using the chosen corpora is informative enough. In addition, the manual search of the translations is very time consuming. This is the search expression for the adjective "*good*": "[Gg]ood". This expression finds all occurrences of the adjective form "*good*" irrespective of the case of the initial letter. We disregard degrees of comparison and take into consideration only basic forms of adjectives because as previous studies indicate (Dyvik 1998, 2002) they are probably comprehensive enough to base our experiment on.

Figure 5.1 Search for the adjective "good" in the Corpuscle.

314	8436667	Good is what conduces to this end and evil what has the opposite effect.	1127.1820	A History Of Political Theor
314		Gėris yra tai, kas prisideda prie šio tikslo, o blogis — tai, kas daro priešingą poveikį.		
315	8483430	Good laws are the only means of making men virtuous.	1127.3269	A History Of Political Theor
315		„Geri įstatymai yra vienintelė priemonė padaryti žmones dorybingus.		
316	8656725	"Ouch!" Cory reacted. "Good point.	1128.763	Who Moved My Cheese?
316		- O, čia tai tiesa, - sureagavo Cory,		
317	8672305	Good people equals good management and good management equals good people.	1132.63	How To Get Rich
317		Geri darbuotojai tolygu geram vadovavimui, o geras vadovavimas - geriems darbuotojams.		
318	8731376	Good luck with the show. "	1132.2881	How To Get Rich
318		Linkiu sėkmės.		
319	8751610	Good work will not make him a rich man, but good work may make him a good workman.	1134.157	NIL
319		Geras darbas nepadarys jo turtingu žmogum, bet geras darbas padarys jį geru darbininku.		
320	8781891	Affirm Your Good	1136.433	NIL
320		Pagaukite savo sėkmę		
321	8948307	Some theatrical and musical groups were invited and performed, for example the Brussels Madrigals Singers and the Brussels Shakespeare Society, who showed their art with the piece, 'Pastime with Good Company', followed by excerpts from the musical, 'Oliver!', performed by the English Comedy Club, Brussels, and the American Theatre Company, Brussels.	1155.90	EESC information
321		Briuselio madrigalų dainininkai ir Briuselio Šekspyro draugija atliko trumpą kūrinį „Pastime with Good Company". Anglų teatras iš Briuselio (English Theatre Brussels) ir Briuselio amerikietiškojo teatro trupė (American Theatre Company Brussels) atliko miuziklo „Oliver!"ištraukas.		
322	6829	2. It shall examine the accounts of all Union revenue and expenditure, and shall ensure good financial management.	1.274	EU constitution
322		2. Jie tikrina visų Sąjungos pajamų ir išlaidų sąskaitas bei užtikrina patikimą finansų valdymą.		
323	11346	1. In order to promote good governance and ensure the participation of civil society, the Union institutions, bodies, offices and agencies shall conduct their work as openly as possible.	1.429	EU constitution
323		1. Siekdamos skatinti tinkamą valdymą ir užtikrinti pilietinės visuomenės dalyvavimą, Sąjungos institucijos, įstaigos ir organai veikia kuo atviriau.		
324	12534	1. The Union shall develop a special relationship with neighbouring countries, aiming to establish an area of prosperity and good neighbourliness, founded on the values of the Union and characterised by close and peaceful relations based on cooperation.	1.472	EU constitution
324		1. Sąjunga plėtoja ypatingus santykius su kaimyninėmis šalimis, siekdama sukurti Sąjungos vertybėmis grindžiamą gerovės ir geros kaimynystės erdvę, kuriai būdingi artimi ir taikūs bendradarbiavimo santykiai.		

Figure 5.2 The example of search results giving both upper and lower case beginning of the word.

The adjective "*brilliant*" was selected as a starting point. This choice was determined because it is an interesting adjective to begin with. For instance, it is obvious that the adjective "*brilliant*" can both refer to literal brilliance (like something shining) and to mental brilliance (intelligence). So the expectation is to get a division into two semantic fields.

After searching for the adjective "*brilliant*", aligned sentences containing this word in English contexts and translations of these contexts were found. Each of the sentences was read and studied, and each of the matching translations for the adjective "brilliant" was written down in a separate file. Only single words were taken into consideration. Since frequency is not relevant, whenever a translation has been registered the search was repeated with the registeredtranslation filtered out in order to simplify the search for further translations. The final set of translations of "*brilliant*" according to the Eng-Lit ParCorp/Eng parallel corpus is:

("brilliant" "nuostabus" "žvilgantis" "spindintis" "ryškus" "šaunus" "talentingas" "išvaizdus" "nušviestas" "šviesus" "akinantis" "puikus" "žavingas" "žėrintis" "genialus" "spindulingasis" "įstabus" "protingas" "įžymus" "iškilus" "galingas" "tobulas" "nepakartojamas" "išradingas" "išmintingas" "nuovokus" "žavus" "klestintis" "žydintis" "reikšmingas" "gabus")

As will be explained further in the following paragraphs (p. 35), during the search of translations, I was going back and forth between them, in order to ensure that the sets of adjectives extracted would be connected translationally. "Starting with the adjective "*brilliant*" I have collected all the possible translations and original adjectives (all the possible translational correspondences) in Lithuanian and written them down. The next stage was to collect all the possible translations of the translations that were derived from "*brilliant*". For example, the first adjective that translates "*brilliant*" to Lithuanian is "*nuostabus*"." (p. 30). The adjective "*nuostabus*" among its translations into English (based on the corpus data) contains the adjective "*beautiful*". But because "*beautiful*" was met in the corpus data not only once, but was included in the vast amount of sentences that repeatedly translated "*nuostabus*" into "*beautiful*", after registering once that "*beautiful*" is one of the translational equivalents to "*nuostabus*", the search was repeated with the adjective "*beautiful*" filtered out.

The filtering process was pursued as follows: when one searches for the adjective "*beautiful*" the most common translation found is "*gražus*". To filter out the repetition of this word as a translation of "*beautiful*" in the target sentences after the search combination for word "[Bb]eautiful" we use symbol combination ">m" which allows to filter out already extracted units. This symbol combination is followed by the root of the adjective in both capital and lower cases followed by the full stop punctuation mark and kleene star (*) which both stand for any other symbols that might follow: [Gg]raž.* . This filter gives us all the sentences of "*beautiful*" with the sentences containing its translation except the matches in the source language (in this case Lithuanian) that contains a word root "*graž*". These are: *gražus*, *graži*, *gražūs*, *gražiai*, *gražiam*, *gražiems* (gender, declensional differences and degrees of comparison of the word), etc. This filtering option simplifies the search a lot. However, in some cases we lose some data, too. For instance, the noun "*gražbylys*" (translates into "*orator*", "*rhetorician*", "*mouther*", "*talker*") falls into the same search combination but it does not have anything to do with the adjective "*gražus*" ("*beautiful*"). It means, that in the case of this word appearing in the source sentence while searching for the adjective "*beautiful*" it will be filtered out too even though it is not what we are meant to filter out. Still, because all the sentences are being read and translational matches collected manually, this filtering system was decided to be used, accepting the loss of the rare coincidental sentences that are being filtered out together with the repetition words in

order to reduce manual work. It is expected that this small possibility of data loss will not harm the investigation. An example of an unfiltered seach window and a filtered one follows:

Figure 5.3 Without filtering out of the adjective "gražus"

Figure 5.4 The example of filtering in the corpus (no sentences in the source that would contain already written down equivalent "gražus" for the target adjective "beautiful")

Because of the complex grammar of the Lithuanian language the search of an adjective in English (eg. "*beautiful*") was giving a vast range of words differing in gender, degrees of comparison and declensional endings. For example, as "*beautiful*" was commonly being translated into "gražus" it might have come in forms like gražus (m. Sg. - Beautiful (boy).), graži (f. Sg. - Beautiful (girl).), gražūs (m. Pl. - Beautiful (mountains).), gražiai (f. Sg. - For a beautiful (girl).), gražiam (m. Sg. - For a beautiful (boy).), gražiems (m. Pl. - For beautiful (mountains).), gražiausias (m. Sg. Superlative degree - The most beautiful (boy), etc. Whenever any of these forms of adjectives were found as a translation of "beautiful" the main form (i.e. m. Sg. Positive degree) of this adjective in Lithuanian was written out and then the root of this adjective was being filtered out to simplify the further search.
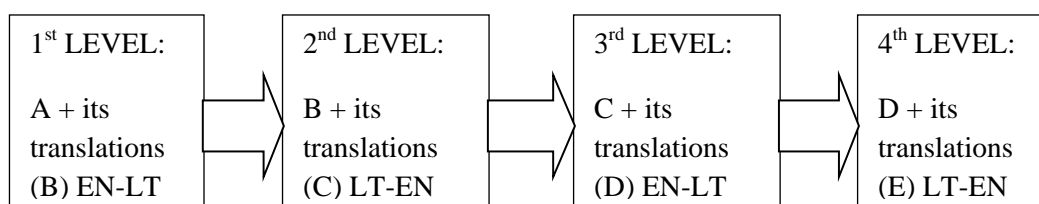
Searching for the English Adjectives

Starting with the adjective "*brilliant*" I have collected all the possible translations in Lithuanian and written them down. The next stage was to collect all the possible translations of the translations that were derived from "*brilliant*". For example, the first adjective that translates

"*brilliant*" to Lithuanian is "*nuostabus*". To collect all the translations of "*nuostabus*" we need to switch the corpus side from Eng-Lit ParCorp/Eng to Eng-Lit ParCorp/Lit. Then we need to search for both kinds of equivalents starting with a capital or a lower letter. This is being done in the same way as it was described above in the search of the English adjectives. Also, Lithuanian words are more difficult to search for in the corpus because of their different changeable endings of gender and case inflexions. That is why during the search we need to set the search syntax to be capable of finding all possible words differing in their gender, degrees of comparison and declensions. This is being done by ignoring the ending of the word where gender, degrees of comparison and declensional suffixes change, and using only word's root for the search, i.e. "[Nn]uostab.*" . When we run the search system, we go to the vertical menu on the left side and select the option '*Word List*'. It gives us the full list of possible words that were found searching for "[Nn]uostab.*", i.e. nuostabus (m. Sg. - Brilliant (boy).), nuostabi (f. Sg. - Brilliant (woman).), nuostabūs (m. Pl. - Brilliant (waiters).), nuostabioms - (f. Pl. - For brilliant (women).), nuostabiais (m. Pl. - (Proud) of brilliant (boys).), nuostabesnis (m. Sg. Comparative degree - (A) more brilliant (man).), nuostabiausia (f. Sg. Superlative degree - (The) most brilliant (woman)) , nuostaba (tr. wonder, surprise, astonishment), nuostabumas (tr. brightness), etc. Not all the words that were found were adjectives, so not all the words in the listed sentences of translations were being checked through (e.g. the last two words exhibited in the list above are nouns). All the rest of adjectival equivalents of "*nuostabus*" were one by one examined and all the unique translations were written down. The final set of translations of "*nuostabus*" is:

("nuostabus" "amazing" "exciting" "great" "terrific" "remarkable" "superb" "wonderful" "fine" "miraculous" "fascinating" "magnificent" "expectant" "breathtaking" "perfect" "awesome" "powerful" "magical" "brave" "marvellous" "extraordinary" "exquisite" "splendid" "delightful" "striking" "beautiful" "swell" "admirable" "brilliant" "wondrous" "good" "tremendous" "lovely" "glorious" "interesting" "ingenious" "loving" "intimate" "keen" "bright" "precious" "rare" "incomparable" "impressive" "stunning" "spectacular" "astonishing" "successful" "exciting" "exceptional" "strange" "quaint" "uninitiated" "divine" "sensational" "glowing" "young" "lively" "charming" "dazzling" "immense" "graceful" "loveliest" "princely" "incredible" "elaborate" "pretty" "gratifying" "advanced" "improbable" "famous" "wild")

To have a reliable data set that could possibly bring fruitful results by applying the Semantic Mirrors system it is recommended to collect translations from four levels, i.e. starting from "brilliant" (A) we collect all its translations (B) into Lithuanian and this structures the first level. The second level is structured when all the translations of "*brilliant*" (B) get all their own translations collected (C). The third level is being combined when we collect all the translations (D) of the previous translations (C), and the fourth level is finished when we collect all the translations (E) of the previous translations (D). Below there is the graphical example of four levels.

1st LEVEL:	2nd LEVEL:	3rd LEVEL:	4th LEVEL:
A + its translations (B) EN-LT	B + its translations (C) LT-EN	C + its translations (D) EN-LT	D + its translations (E) LT-EN

Figure 5.5 Collection process of the translational correspondences

English to Lithuanian translations are being collected from both the 1st and the 3rd levels into one file. Lithuanian to English translations are being stored into a separate file as well. There is a specific syntax how the words and their translations have to be written down in order to be located for analysis by the Semantic Mirrors system. The syntax is: the adjective in one language combine one group together with all its translations:

("amžinas" "eternal" "timeless" "infinite" "permanent" "unalterable")

The first word in the brackets is a target adjective and all the following ones are its translations into English that have been found in the corpus during the search procedure. Brackets indicate the group of adjectives in one language with all the found translations from another language. Double quotation marks separate words and denote a frame of one word-unit for the Semantic Mirrors system.

During the data extraction process the direction of translation is disregarded. This means that we extract as the correspondences of a given L1 adjective *a* all the words into which *a* has been

35

translated in the L1-to-L2 part of corpus as well as all the words that have been translated into *a* in the L2-to-L1 part of the corpus.

5.2 Overview of the Data

After all the unique translations of the adjectives in all four levels were searched for and collected we ended up with two data sets Lithuanian-English and English-Lithuanian. In the **English-Lithuanian data set** there have been documented 350 unique English adjectives with their translations. The first level contains 1 English adjective and the third level contains 349 adjectives. Starting from the adjective "*brilliant*" and further going from a translation to its translations semantically unique forms of adjectives were noted down. The translations led to very uncommon and rare adjectives like *"bully", "clear-cut", "first-class", "heavy-duty"*, etc. The Lithuanian translations have also been at some times hard to distinguish according to their suitability. For example, the possible translation for the adjective *"harsh"* is *"užkimęs"*. In Lithuanian *"užkimęs"* is a participle in the past tense, though in English it is being translated as an adjective. The other example: one of the possible translations for the adjective *"perfect"* is *"lygut lygutėlis"* which stands for reduplicated word adjective which is not very popular in use.

During the search of translations for the English adjectives there were some cases where translations were not extracted. For example:

- Metaphoric translation of the English adjective in Lithuanian:

"Everything he had ever believed about the demise of the Illuminati was suddenly looking like <u>a brilliant sham</u>.

Jo ankstesnysis įsitikinimas, kad Iliuminatų brolija nebeegzistuoja, ūmai pavirto <u>sprogusiu muilo burbulu</u>. (1010.3013 Angels and Demons 4456996)"

"*A brilliant sham*" is being translated into "*exploded soap-bubble*" literally.

- Escaping the use of the adjective in the translation:

"<u>He was a brilliant talker</u>, and when he was arguing some difficult point he had a way of skipping from side to side and whisking his tail which was somehow very persuasive.

Ką jau ką, o <u>pakalbėti jis mokėjo</u>: įrodinėdamas kokį painų dalyką, turėjo įprotį straksėti į šalis ir vizginti stimburį, - tai kažkodėl padėdavo įtikinti pašnekovą. (1024.180 Animal Farm 5846987)"

"*He was a brilliant talker*" is translated into "*He knew how to talk*" literally.

- The adjective in English becomes another part of speech (a verb in this exact case) in Lithuanian and does not meet the criteria for the correspondence of translations:

"The capital is established near Zhong Mountain; The palaces and thresholds <u>brilliant and shining</u>; The forests and gardens are fragrant and flourishing; Epidendrums and cassia complement each other in beauty. The forbidden palace is magnificent; Buildings and pavilions a hundred stories high.

Prie Zhongo kalno yra įsikūrusi sostinė; rūmai ir įvažiavimai <u>spindi ir tviska</u>; kvepia, svaigina vešintys miškai ir sodai; epidendrumai ir kasijos savo grožiu papildo vienas kitą; uždraustieji rūmai yra puikūs; šimtaaukščiai pastatai ir paviljonai. (1107.50 The Power of Identity 7003736)"

"Brilliant and shining" is translated into *"glisten and shimmer"*.

In these cases nothing is extracted as a translation of "*brilliant*".

These choices of indirect translation are the result of human translation. It is still considered as the best translation option and agreed to have more value than machine translation. Though, because of the translation specifics we have to experience some data loss.

In the **Lithuanian-English data set** there were found 716 adjectives having unique meanings. In the second level there were 30 Lithuanian unique adjectives and in the fourth level 686 adjectives. Again, because of wide range of declensional forms there were some very familiar but still slightly different adjectives noted down. For example, *"didelis"* and *"didžiulis"*, meaning *"big", "huge"*, or *"didingas"* and *"didis"* meaning *"great", "grand"*. These are close synonyms, but in the dictionary (http://dz.lki.lt/search/, http://www.lkz.lt/startas.htm) each of this word holds its own, separate unit, and meaning. That is why each of these very close synonyms

were noted down as separate adjectives. This abundance in Lithuanian language was one of the causes of a larger data set in the Lithuanian-English data set.

The other reason for a larger data set in this side of the language pair is that it got the final, the 4th level where translations have been found. The 4th level is the deepest and derives from the 3rd – the last biggest set of unique adjectives. The 1st level starts from the data search and includes only the beginning word (in our case "*brilliant*") with all its translations (in our case 30 unique translations). However, it is the smallest one and has disadvantage in extent to the final English-Lithuanian data set. I assemble the English words on steps (levels) 1 and 3 and the Lithuanian words on steps (levels) 2 and 4. So the total number of the Lithuanian words is therefore going to be substantially larger than the total number of the English words. There have been documented 350 unique English adjectives with their translations.

The same as in the English-Lithuanian translations, the mining process in Lithuanian-English language side (when I was searching in the corpus for the Lithuanian adjectives and their translational correspondences) contained some sentences which because of human translator specifics were not resulting in translating adjective into adjective. Examples are the following:

- Escaping the use of the adjective in the translation:

"– Koks nuostabus sutapimas, mes kaip tik turime laisvą gydymo kambarį, kuriuo niekada nesinaudojame.

'Funnily enough, we've got an empty healing room that we never use.' (1020.2733 Kaip buti geru žmogumi 4770588)"

The expression "*Koks nuostabus sutapimas*" literally meaning "*what an amazing coincidence*" was chosen to be translated into "funnily enough".

- The adjective in Lithuanian becomes the other part of speech (an adverb in this exact case) in English:

"Vakaras buvo nuostabus, ir visa naktį sapnavau paslaptingąją ir puikiąją Mere Kavendiš.

The evening passed pleasantly enough; and I dreamed that night of that enigmatical woman, Mary Cavendish. (994.225   Paslaptingas atsitikimas Stailze 2754324)"

The expression "*Vakaras buvo nuostabus*" which literally means "*The evening was amazing*" was chosen to be translated into '*The evening passed pleasantly enough*".

The translations of fiction tend to be less literal than the other kinds of texts. This leads to higher proportion of sentences that cannot be used. However, this is the nature of fiction translations made by people. Nevertheless, any the so called loss of data is natural and will not unbalance the investigation.

The amount of unique units of translations at the end of the search allows the presumably efficient investigation of the Semantic Mirrors method generated on Lithuanian adjectives.

6 Data Processing and Analysis. The Lithuanian Thesaurus

To select some adjectives for my analysis I have chosen the Lithuanian novel titled "Sukilėliai" (En. - The Rebels; author Vincas Mykolaitis-Putinas, 1986) and while reading it through I collected 20 adjectives that were first met in the context:

*Šaunus (*tr. *dear, dashing, valiant, decent)*

*puikus (*tr. *exellent, superb, splendid, beautiful)*

*didelis (*tr. *large, big, great, considerable)*

*griozdiškas (*tr. *clumsy, unwieldy, cumbersome, bulky)*

*neturtingas (*tr. *poor, poverty-stricken, indigent, penniless)*

*nuskurdęs (*tr. *poverty-stricken)*

*darbštus (*tr. *industrious, laborious, diligent, hard-working)*

*sumanus (*tr. *clever, great, intelligent, bright)*

*senas (*tr. *old, aged, used, eldery)*

*geras (*tr. *good, nice, kindly, gentle)*

*apgriuvęs (*tr. *dilapidated, crazy, tumbledown, decrepit)*

*tinkamas (*tr. *suitable, appropriate, happy, relevant)*

*jaunas (*tr. *young, green, juvenile, youthful)*

*gražus (*tr. *beautiful, lovely, pretty, good)*

*meilus (*tr. *nice, loving, lovely, kind)*

*kuklus (*tr. *modest, humble, conservative, quiet)*

*sklandus (*tr. *fluent, smooth, round, fluent)*

*dailus (*tr. *pretty, handsome, nice, elegant)*

*tikras (*tr. *real, sure, certain, positive)*

*klusnus (*tr. *obedient, humble)*

Since these 20 adjectives were found in one book and I collected them as they were occurring in the text while reading, they might appear to be semantically close.

Next, I searched for them in the Thesaurus and analysed them. This is the setting of the search:

Figure 6.1 The settings of my search

The seetings were set to be: Word Base: **agne** extended (from 0, to 1000; totally 715 entries, 282 starred entries); Synset Limit: **automatic** (**20**); Overlap Threshold: **0.05**. This yielded a thesaurus with 14 of the 20 randomly selected adjectives.

There will be 14 randomly selected (listed in the page above) thesaurus entries followed by the examples of the same adjectives and their senses found in the Modern Lithuanian Dictionary and in the Dictionary of the Lithuanian Language. The results from the Semantic Mirrors will be listed first and the entries of the two dictionaries will follow. In each thesaurus all the adjectives which according to my interpretation can be matched with senses described in the golden standards (DLKZ and LKZ dictionaries) will be colour marked. Adjectives matched with senses in the Modern Lithuanian Dictionary are coloured green and the ones matched with senses in the Dictionary of the Lithuanian Language are coloured blue. The words which will be matched with senses in both dictionaries will be underlined and coloured green.

Antonyms will be marked red.

In most cases where a high number of words appear in all three data sets that we exhibit for each adjective (SM thesaurus entry, DLKZ and LKZ), this is because of the wide meaning of the adjective (for example adjective "*puikus*"; see page 51).

6.1 Evaluation of the Results: Good, Average and Bad Automatically Generated Results

In this thesis, I limited the study to evaluating the sets of related words, disregarding the distinctions between hyperonymy, hyponymy and synonymy, and disregarding the sense

individuation. The results will be sorted out according to their performance in ability to match the sense groups in the dictionaries based on my interpretations.

The results will be compared with the information from golden standards (GS) and discussed. The indicators of the senses of words in the natural language were chosen to be the Modern Lithuanian Dictionary and the Dictionary of the Lithuanian Language.

The Modern Lithuanian Dictionary *is a universal one-volume explanatory normative work of standard language intended for the wide circle of readers. It contains a huge amount of modern Lithuanian words, some regional dialectal and more widely used spoken language words. Moreover, it also contains words from past and contemporary fiction, especially classical papers, which are necessary for the studying youth to cultivate their language, to reflect various language styles, often suitable for specific new concept expression* (http://dz.lki.lt/static/english.html).

The Dictionary of the Lithuanian Language *is the largest work of twentieth-century Lithuanian linguistics. The idea of the Dictionary was conceived (and its compilation begun) by the eminent Lithuanian philologist Kazimieras Buga at the turn of the twentieth century. Since then several generations of lexicographers of the Institute of the Lithuanian Language have been engaged in the preparation of its twenty volumes for six decades (published between 1941 and 2002) (http://www.lkz.lt/en/dze.htm).*

The Dictionary aims to give the words and illustrate their usage by quotations culled from all kinds of writings and dialect records from the period between the year 1547, i.e. the publication of the first Lithuanian book, and 2001 (http://www.lkz.lt/en/dze.htm).

The twenty volumes of the Dictionary make up about 22,000 pages, comprising half a million headwords and over 11,000,000 words of text. This academic edition of the Dictionary of the Lithuanian Language is significant not only as a major landmark of Lithuanian philology, it is also an authoritative source for comparative Indo-European studies. It presents the origin, history and spread of a word, its grammatical and accentual forms and categories, and its peculiarities with respect to word-formation, semantic structure, stylistic usage, etc. The Dictionary abounds in extra-linguistic information: the illustrative material carries much

background information about the everyday life of the speakers of the language, their social relations, ethical values, ethnographical details, etc. (http://www.lkz.lt/en/dze.htm).

The Dictionary of the Lithuanian Language is accessible online at www.lkz.lt.

The reason why I have used two dictionaries as a gold standard in my thesis is because since there are no any Lithuanian thesauri generated manually or automatically at all, comparing the results with a dictionary in this case has been the only way to evaluate my results. Since sometimes one or another dictionary (DLKZ or LKZ) contains more explicit information than the other, I was taking them both into consideration. However, while comparing the results generated by SM with the results from the dictionaries, the different senses in the dictionaries were merged into one list and compared with the list generated by SM. Besides, the same two senses were not taken in consideration twice during the calculations.

The other reason why it is useful to have two golden standards in this case is that it is interesting to see the separate dictionaries' results and the adjectives from SM that fits them. As we will see further on the pages 94-95 (in the case of the adjective "*dailus*") SM is capable of sorting out the different senses listed in different dictionaries. Also it is capable of suggesting senses which do not appear in any of the dictionary but might as well be concluded to be an unclassified sense rather than noise. However, as it was mentioned before, only the merged results from two dictionaries will be taken into consideration while doing the countings.

There happened to be many cases suggested by SM of hyperonymy-hyponymy relations among the adjectives. For this reason, a few further paragraphs will discuss the possibility of hyperonymy-hyponymy relations among the adjectives.

The interpretation of hyperonymy. Hyponym is the term derived from Greek word formation unit "*onyme*" which means "name", "*hypo*" - "located inside, lower", "*hyper*" - "higher". This dimensional model describes hyperonym and hyponym relations and represents the widening or narrowing the meaning of a word. Any hyponymous sense of the word can be included (as a special case) into the concept of hyperonym. A specialized, narrowed meaning of a word is usually more concrete precise. Hyperonyms convey broader, more abstract meaning of the word. Often it is harder to understand it and it is interpreted in more different ways (Tajarobi, 1998).

Since in the Lithuanian language there is no dictionary of hyperonyms and hyponyms and there are not many published works that analyse those relations, it is (by most cases of adjectives) an impossible task to determine these relations when we talk about adjectives. Also, analysing adjectives we experience a problem of clearly determining the content of semantic range in the lexical unit. "Consider the problem of defining the adjective "*good*". A good pencil is one that writes easily, a good knife is one that cuts well, a good paint job is one that covers completely, a good light is one that illuminates brightly, and so on. As the head noun changes, "*good*" takes on a sequence of meanings: writes easily, cuts well, covers completely, illuminates brightly, etc. It is unthinkable that all these different meanings should be listed in a dictionary entry for good." (Fellbaum, 1998).

There are few works of hyperonyms and hyponyms relations written by Lithuanian linguists (J. Navakauskienė, 2005; E. Jasaitienė, 1988, 2009; O. Armalytė; L. Pažūsis). A. Gudavičius (2007) published the paper about problems in translating these relations from one language to another. But none of the papers review hyperonymy-hyponymy relations among adjectives in the Lithuanian language. However, hierarchy of adjectives is a questionable topic - "it is not clear what it would mean to say that one adjective ''is a kind of'' some other adjective" (Fellbaum, 1998) (except in cases of colours, when one colour can be a hyperonym of the other (ex. "blue" is a hyperonym of "turquoise" and "royal blue")). For this reason categorization into hyperonymy-hyponymy will be not taken under deep consideration, except some special cases where the possibility of hierarchial relations will be discussed (adjectives *šaunus, puikus, didelis*).

To evaluate if a thesaurus entry created by SM reflects manually generated results (golden standard) well, averagely or badly, I will calculate the **recall**, the **precision** and the **F-score** for each entry from the thesaurus.

6.2 The Recall the Precision and the F-score

The precision is the number of correct results divided by the number of all returned results:

P = nr of correct results / nr of returned results

The recall is the number of correct results divided by the number of results that should have been returned:

R = nr of correct results / nr of results that should have been returned

According to this we can define what each of **the variables according to our investigation** is (to give examples I will use the adjective "*šaunus*" (p. 46):

In order to get <u>the number of correct results</u> we count the number of words returned by SM that are actually mentioned as related words in the golden standard (GS) DLKZ and LKZ. In the case of "šaunus" they were 2 ("*gražus*" and "*puikus*").

In order to <u>get the number of returned results</u> we count all the words returned as related by SM, including those that did not match any senses in the GSs (27). 2 + 27 = 29.

In order to get <u>the number of results that should have been returned</u> we count all the semantically related words given in the GSs merged together. It is 5 (in DLKZ: "*puikus*", "*gražus*", "*smarkus*", "*greitas*", "*garbingas*") + 14 (in LKZ: "*judrus*", "*žvitrus*", "*apsukrus*", "*atžarus*", "*skaudus*", "*rūstus*", "*piktas*", "*stiprus*", "*intensyvus*", "*didelis*", "*ilgas*", "*gerokas*", "*nemažas*", "*gausus*"). From this set we substract the words which did not happen to be in the corpus and hence could not have been found by the SM method. They are: "*atžarus*" and "*gerokas*". 5 + 14 - 2 = 17

Qualitatively modified precision, recall and the F-score:

Since the gold standards are regular dictionaries and not thesauruses or synonym dictionaries, the number of synonyms listed for each entry is very limited. The result of this is that the precision calculated in the way described above is unreasonably low: many good synonyms are counted as noise because of the limitation of regular dictionaries in providing synonyms. We therefore also provide a qualitatively modified calculation of precision and recall, where the <u>number of correct results</u> is taken to be the number of words returned by SM which **intuitively fit** one of the senses described for the entry in the gold standards. In the case of "šaunus" there are 15 senses recognized by SM which in this way is included in the set of correct results.

The number of returned results is the number of senses in GSs that we managed to connect with words from SM (15 in this case) + all the other single words that did not match any senses in the GSs. So, 15 + 14 = 29 - is the number of returned results in the case of adjective "*šaunus*".

The number of results that should have been returned will then be the union of the synonyms actually mentioned in the gold standards and the qualitatively augmented set of correct results described above (17 + 15 = 32).

To calculate the F score we will use this formula: $F = 2 \cdot \frac{P \cdot R}{P + R}$

To label the entries according to their idealized precision, the idealized recall and the idealized F-score I will use this system: when the recall/ precision/ F score is 1 - 0.67 the entry will be labelled as of a **"good"** recall/ precision/ F-score. Entries that have got score between 0.66 and 0.33 will be labelled as **"average"** and the entries with the score of 0.32 to 0 will be labelled as "**bad**" performance of idealized automatic generation of thesaurus entries.

Now we can calculate the precision and the recall of our cases:

I. Šaunus

The Semantic Mirrors

> (Translation: smart, good, bright, brilliant, fine, nice, great.)
>
> *Hyperonyms:* tikslus‹2›, gražus‹1›, geras‹1›, ryškus‹1›, nuostabus‹1›, puikus‹1›.
>
> *Hyponyms:* apdairus, apgalvotas, aukštas‹1›, aštrus‹1›, dailus‹1›, elegantiškas‹1›, grakštus, gudrus‹1›, iškalbingas‹2›, išmintingas‹1›, išmoningas‹1›, išsilavinęs, išvaizdus, madingas‹1›, nuovokus, pasipūtęs‹2›, patrauklus, prašmatnus‹1›, protingas‹1›, puošnus‹1›, racionalus‹1›, sumanus‹1›, sąmojingas‹1›.

DLKZ (Modern Lithuanian Dictionary)

> šaunus, ~ì
>
> 1. Puikus, gražus; smarkus, greitas (tr. Excellent, beautiful, intense, fast)
>
> **Correspondences from SM**: gražus‹1›, nuostabus‹1›, puikus‹1›, dailus‹1›, išvaizdus, patrauklus, prašmatnus‹1›, puošnus‹1›
>
> 2. Garbingas (tr. Honourable)
>
> **No correspondences from SM**

LKZ (the Dictionary of the Lithuanian Language)

šaunus, -ì adj.

1. keliantis pasigėrėjimą savo puikia išvaizda, geromis savybėmis, darbu, veikla (tr. causing admiration for his/hers good looks, good qualities, work, activities)

Correspondences from SM: gražus‹1›, geras‹1›, nuostabus‹1›, dailus‹1›, elegantiškas‹1›, grakštus, išvaizdus, prašmatnus‹1›, puošnus‹1›

2. puikiai, pasigėrėtinai atliekamas, padarytas, sutvarkytas (tr. perfectly performed, admirably made, arranged)

Correspondences from SM: puikus‹1›, geras‹1›, nuostabus‹1›, apgalvotas

3. greitas, judrus, žvitrus, apsukrus (tr. fast, agile, sprightly, shifty)

No correspondences from SM

4. atžarus, skaudus, rūstus (tr. offensive, painful, severe)

Correspondences from SM: aštrus‹1›, pasipūtęs‹2›

5. kuris smarkaus būdo, piktas (tr. The one with a severe manner, angry)

No correspondences from SM

6. smarkus, stiprus, intensyvus (tr. vigorous, strong, intensive)

Correspondences from SM: aštrus‹1›

7. didelis (erdvės, apimties atžvilgiu) (tr. large (in a space volume))

Correspondences from SM: aukštas‹1›

8. ilgas, gerokas (laiko atžvilgiu) (tr. a long, long while (in a time volume))

No correspondences from SM

9. nemažas, gausus (tr. significant, abundant)

No correspondences from SM

Because the amount of the exact matches in the GSs and SM for adjective "šaunus" (and all the rest of the 14 cases of adjectives that I am investigating in this work) is so low: the recall for "šaunus" is 0.12 and the precision is 0.07, I will go on with the qualitative evaluation.

The Qualitative Evaluation

In this case, the dictionaries which I have used as a golden standard have significantly different amount of senses: the Modern Lithuanian Dictionary (further named as DLKZ) has two senses, while the Dictionary of the Lithuanian Language (further named LKZ) contains nine senses. The Semantic Mirrors has not determined any senses of this adjective, but it identified hyperonyms and hyponyms.

Even though hyperonymy and hyponymy relations cannot be determined in this case (because of absence of hyperonymy-hyponymy relations among adjectives), four of six words listed as adjective's "*šaunus*" hyperonyms can be classified as some of the senses listed in the dictionaries. The same tendency can be noticed among the listed hyponyms. This proves that the Semantic Mirrors has the ability to sort out words from the translations that are concluded to be the sense of the adjective in the golden standard but not always sorts them out accurately according to their semantic relations. The words listed as adjective's "*šaunus*" hyponyms but rather falling to sense group of this word are added just after the sense (meanings) given by dictionaries (the target in the borders).

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 15.

All the rest of the adjectives that were returned by SM and named noise are 14. $14 + 15 = 29$

P $= 15 / 29 =$ **0.517**

The number of the results that should have been returned is 19. And the new correct results are added: $19 + 15 = 34$.

R $= 15 / 34 =$ **0.441**

F $=$ **0.476**

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:

Figure 6.2.1 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

In the figure 6.2.1 it is easy to see the senses of "*šaunus*" that are given by SM , LKZ and DLKZ. LKZ shares some senses that are the same in both SM and the DLKZ and also it contains some senses that DLKZ does not contain, but SM has been able to find them out ("*geras*" (tr. good)*, "nuostabus"* (tr. amazing)*, "apgalvotas"* (tr. considered)*, "aukštas"* (tr. tall)*, "aštrus"* (tr. sharp)*, "elegantiškas"* (tr. elegant)*, "grakštus"* (tr. graceful)*, "pasipūtęs"* (tr. arrogant)). The other intersection is the one that contains senses of "*šaunus*" that are listed in the LKZ, DLKZ and have been listed by SM too ("*puošnus*" (tr. ornate)*, "patrauklus"* (tr. attractive), *"prašmatnus"* (tr. luxurious)*, "gražus"* (tr. beautiful)*, "puikus"* (tr. great)*, "dailus"* (tr. pretty), *"išvaizdus"* (tr. presentable)).

The adjectives that share some senses in LKZ and DLKZ but have not been found by SM are: ("*greitas*" (tr. fast), "*smarkus*" (tr. intense)).

There are also adjectives that are listed as senses of the adjective "*šaunus*" only in LKZ (*judrus* (tr. mobile)*, žvitrus* (tr. sprightly), *apsukrus* (tr. shifty)*, piktas* (tr. angry)*, ilgas* (tr. long)*, nemažas* (tr. not small), *stiprus* (tr. strong), *intensyvus* (tr. intensive), *didelis* (tr. big), *gausus* (tr. plentiful)), only in DLKZ (*garbingas (tr. respectable)*), or only in SM (*tikslus (tr. accurate),*

ryškus (tr. bright), apdairus (tr. cautious), gudrus (tr. sly), iškalbingas (tr. eloquent), išmintingas (tr. wise), išmoningas (tr. ingenious), išsilavinęs (tr. educated), madingas (tr. trendy), nuovokus (tr. sensible), protingas (tr. clever), racionalus (tr. rational), sumanus (tr. ingenious), sąmoningas (tr. conscious). All these words found and listed only by SM have to be sorted out as noise. However, the reason why the words appeared in SM is based on the information that was gathered from the corpus - information of natural language usage and the translational relations (decisions that were made by human translators choosing one or the other word to represent the target word). We can intuitively notice, that SM next to all the senses that are defined in the dictionaries, suggests one more sense - the sense of <u>*great mental abilities*</u>: *apdairus (tr. cautious), gudrus (tr. sly), iškalbingas (tr. eloquent), išmintingas (tr. wise), išmoningas (tr. ingenious), išsilavinęs (tr. educated), nuovokus (tr. sensible), protingas (tr. clever), racionalus (tr. rational), sumanus (tr. ingenious), sąmoningas (tr. conscious).* It is eleven out of fourteen hits from the full set of what we have to call noise because it did not match any sense in the dictionaries. Nevertheless, when the number is so high (78,5%), we might consider that the Semantic Mirrors suggests some other sense that should have been in the golden standard - the sense of *mental brightness*, which for a person with native Lithuanian competence seems to be legitimate to represent the adjective "*šaunus*". Example:

Tomas yra šaunus studentas. (tr. Tom is a <u>sly/ wise/ ingenious/ sensible/ clever</u> student.).

Tai yra tikrai šaunus fizikos vadovėlis. (tr. This is really <u>wise/ sensible/ clever/ rational/ conscious</u> textbook of physics.).

Kiek turėjai tikrai šaunių mokytojų? (tr. How many really <u>wise/ sly/ ingenious/ sensible/ clever/ rational/ conscious/</u> teachers have you had?).

Likewise, adjectives "*ryškus*" and "*madingas*" in particular sentences mean "*šaunus*" too:

Kokios šaunios spalvos šioje dažų paletėje! (tr. What <u>bright</u> (tr. ryškus) *colors are in this paint palette!);*

Kokia šauni šiandien saulė! (tr. What the bright (tr. "ryškus") *sun is today!).*

Šios merginos aprangos stilius yra išties šaunus. (tr. The dressing style of this girl is really fashionable.).

If the dictionaries would include senses with the synonyms of "*šaunus*" listed above, the 'noise' would decrease to only 2 adjectives ("*tikslus*" and "*ryškus*") in all the samples of the adjective "*šaunus*". This would result in extremely high precision, recall and the F-score.

II. Puikus

The Semantic Mirrors

(*Translation:* great.)

Hyponyms: aitrus‹2›, aktyvus‹1›, apgalvotas, apsukrus, apčiuopiamas‹1›, aršus‹1›, atidus, augalotas, auksinis‹1›, aukštas‹1›, baisus‹1›, begalinis‹1›, bekraštis, beribis‹1›, besaikis‹1›, brandus‹1›, brangus‹1›, būtinas‹1›,dažnas‹1›, didelis‹1›, didingas‹1›, didis, didus‹1›, didysis, didžiulis‹1›, dieviškas‹1›, dosnus‹1›, dramatiškas, drūtas‹1›, egzotiškas‹1›, ekstravagantiškas, erdvus‹1›, fantastiškas‹1›, fenomenalus‹1›, galingas‹1›, galintis‹1›,galvotas‹1›, garsus‹1›, gausus‹1›, genialus‹1›, geras‹1›, giluminis‹1›, gilus‹1›, glaudus, globalinis, globalus, gražus‹1›, gražutis, gremėzdiškas‹1›, griežtas‹1›, grėsmingas, gudrus‹1›, ilgas‹1›, intensyvus‹1›, intriguojantis‹1›,įspūdingas, išauklėtas‹1›, išdidus‹1›, išimtinis‹1›, iškalbingas‹2›, iškilmingas, iškilus‹1›, išmintingas‹1›, išmoningas‹1›, išplėstas‹1›, išpūstas‹1›, išraiškingas‹1›, išsipūtęs‹1›, išskaidytas‹1›, išskirtinis‹1›, ištvermingas, išvaizdus,jaudinantis‹1›, kapitalinis‹1›, kategoriškas‹2›, kilnus, klestintis‹1›, kokybiškas‹1›, kruopštus‹1›, lemiamas‹1›, lemtingas‹1›, liūdnas‹2›, magiškas‹1›, maksimalus‹1›, malonus‹1›, masyvus‹1›, matomas, mažas‹2›, milžiniškas‹1›,mįslingas, narsus, naudingas‹1›, naujas, neaprėpiamas‹1›, neapsakomas‹1›, neatidėliotinas, neatskiriamas‹2›, neblogas, neeilinis, nekasdieniškas‹1›, nemažas‹1›, nemenkas, nenormalus‹1›, nenumaldomas‹1›,nenusakomas‹1›, nepakartojamas, nepaprastas‹1›, neperskiriamas‹1›, neprilygstamas‹1›, neregėtas‹2›, nesavas‹2›, netikėtas‹2›, neįkainojamas‹1›, neįsivaizduojamas‹1›, neįtikėtinas, neįveikiamas‹2›, nuodugnus,nuostabus‹1›, nuoširdus‹1›, nusipelnęs‹1›, opus‹1›, orus‹1›, padidėjęs‹1›, pagarsėjęs‹1›, pagrindinis‹1›, pakankamas‹1›, pakilus, pamatinis‹1›, paplitęs‹1›, pasibaisėtinas, pasigėrėtinas, pastebimas‹1›, patenkintas‹1›,patikimas‹1›, patobulintas, pavojingas, pašėlęs‹2›, pelningas‹1›, platus, pompastiškas‹2›, populiarus‹1›, pozityvus, prasmingas‹1›, pribloškiantis, putnus‹1›, radikalus, rafinuotas‹1›, reikiamas, reikšmingas‹1›, reikšminis, retas,riebus‹1›, rinktinis‹1›, ryžtingas‹1›, savotiškas‹1›, sensacingas‹1›, siaubingas‹1›, siautulingas, simpatingas, skambus‹1›, skardus‹1›, skaudus‹1›, skiriamasis‹1›, slaptas‹2›, smagus‹1›, smailus‹2›, smarkus‹1›, smulkmeniškas‹2›,sodrus, solidus‹1›, sotus‹1›, spartus‹1›, spindulingasis, stambus, stangrus‹1›, status‹1›, stebinantis, stebuklingas‹1›, stebėtinas, stiprus‹1›, storas‹1›, stulbinamas‹1›, stulbinantis‹1›, stuomeningas‹1›, sumanus‹1›,sumanytas‹3›, sunkus‹1›, supratingas‹1›, susikaupęs‹2›, suveltas‹1›, svaiginantis‹1›,svarbus, svarus, sąmojingas‹1›, sėkmingas, tankus‹1›, tariamas‹1›, tarptautinis, taurus, teigiamas‹1›, tirštas‹1›, triuškinantis,turtingas‹1›, tuščias‹2›, unikalus, valstybinis‹2›, veiksmingas‹1›, vertingas, vešlus‹1›, vibruojantis, vidutinis, viliojantis‹1›, visiškas, vyraujantis‹1›, vyriausias, ypatingas‹1›, įdomus‹1›, įdėmus‹1›, įkritęs‹1›, įsidėmėtinas‹1›,įspūdingas‹1›, įstabus, įtaigus‹1›, įtakingas‹1›, įtikimas, įvairus‹1›, įžvalgus‹2›, įžymus‹1›, šaunus‹1›, šiuolaikiškas, šiurpus‹1›, šlovingas‹1›, šviesus‹1›, žavingas‹1›, ženklus‹1›, žinomas‹1›, žymus.

DLKZ (Modern Lithuanian Dictionary)

---

puikus, ~ì

1. labai gražus, dailus (tr. very nice, pretty)

Correspondences from SM: gražus‹1›, gražutis, išvaizdus, neapsakomas‹1›, simpatingas, žavingas‹1›

2. labai geras, šaunus (tr. very good, nice)

<div style="border:1px solid black; padding:10px;">

Correspondences from SM: <u>pasigėrėtinas</u>, fantastiškas‹1›, fenomenalus‹1›, <u>geras‹1›</u>, <u>išskirtinis‹1›</u>, kokybiškas‹1›, nekasdieniškas‹1›, nepakartojamas, nepaprastas‹1›, nuostabus‹1›, <u>pasigėrėtinas</u>, pribloškiantis, stulbinamas‹1›, stulbinantis‹1›, <u>įspūdingas‹1›</u>, <u>šaunus‹1›</u>

3. Išdidus (tr. Proud)

Correspondences from SM: <u>didingas‹1›</u>, <u>didis</u>, <u>didus‹1›</u>, <u>išdidus‹1›</u>, <u>orus‹1›</u>

</div>

LKZ (the Dictionary of the Lithuanian Language)

<div style="border:1px solid black; padding:10px;">

puikus, -ì adj.

1. labai gražus, dailus (tr. very nice, pretty)

Correspondences from SM: <u>gražus‹1›</u>, <u>gražutis</u>, <u>išvaizdus</u>, <u>neapsakomas‹1›</u>, <u>simpatingas</u>, <u>žavingas‹1›</u>

2. tinkamas, patogus, parankus, geras (tr. appropriate, convenient, handy, good)

Correspondences from SM: <u>geras‹1›</u>, <u>išskirtinis‹1›</u>, naudingas‹1›

3. malonus, patrauklus (tr. pleasant, attractive)

Correspondences from SM: <u>pasigėrėtinas</u>, <u>pasigėrėtinas</u>, malonus‹1›

4. doras, teisingas, žmoniškas (tr. honest, fair, humane)

No correspondences from SM

5. puošnus, prabangus, turtingas (tr. gorgeous, luxurious, rich)

Correspondences from SM: iškilmingas, iškilus‹1›, auksinis‹1›, brangus‹1›, ekstravagantiškas, iškilmingas, iškilus‹1›, pompastiškas‹2›, turtingas‹1›, vertingas

6. šaunus, smarkus; garbingas (tr. cool, intense, honourable)

Correspondences from SM: <u>didingas‹1›</u>, <u>didis</u>, <u>didus‹1›</u>, <u>įspūdingas‹1›</u>, <u>šaunus‹1›</u>, nenumaldomas‹1›, siautulingas, smarkus‹1›

7. nemažas, didelis, žymus; tikras (tr. significant, big, considerable, real)

Correspondences from SM: <u>didingas‹1›</u>, <u>didis</u>, <u>didus‹1›</u>, augalotas, aukštas‹1›, begalinis‹1›, bekraštis, beribis‹1›, didelis‹1›, didysis, didžiulis‹1›, erdvus‹1›, gausus‹1›, gremėzdiškas‹1›, ilgas‹1›, masyvus‹1›, milžiniškas‹1›, žymus nemažas‹1›, nemenkas, platus, ženklus‹1›

8. išdidus, išpuikęs (tr. proud, haughty)

Correspondences from SM: <u>išdidus‹1›</u>

9. kuris aukštos kilmės, kilnus (tr. a high-ranking, noble)

Correspondences from SM: didysis, kilnus

</div>

Because the amount of the exact matches in the GSs and SM for the adjective "*puikus*" is low (the recall is 0.48 and the precision is 0.05, I will go on with the qualitative evaluation in this case too.

The Qualitative Evaluation

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 62.

All the rest of the adjectives that were returned by SM and named noise are 175. 175 + 62 = 237.

P = 62 / 237 = **0.261**

The number of the results that should have been returned is 24. And the new correct results are added: 62 + 24 = 86.

R = 62 / 86 = **0.72**

F = **0.384**

The adjective "*puikus*", similarly as the previously examined adjective "*šaunus*" has a rather wide meaning. It can be compared with the problem that was previously mentioned in this paper - problem of the adjective "*good*" in the English language (Fellbaum, 1998). It contains such a wide semantic field that it can be replaced with nearly any other adjective (in most cases more concrete) that is of the positive meaning, according to the noun and the specific feature that is meant to be specified. A few illustrative examples:

The possible replacement of the adjective "*puikus*" with a more accurate adjective according to the noun and its feature which is meant to be specified:

Puikus namas - šiltas gyventi/erdvus/švarus/prabangus, etc (tr. Great house - warm to live in, commodious, clean, luxurious);

Puikus vaikas - klusnus/gudrus/gerai išauklėtas, etc (tr. Great child - humble, clever, polite);

Puikus stalas - gražus/didelis/ant kurio yra daug maisto, etc (tr. Great table - nice, big, containing a lot of food on it).

The "*puikus*" and "*šaunus*" examples show that adjectives with very wide meanings were classified by SM rather fair (the recall of the adjective "šanus" is 0.44 and of the adjective "puikus" is 0.72). However, the manual classification of the adjectives with the wide meanings is rather complicated, wide and noticeably vague: the linguists formatting dictionaries try to capture the meaning of vague words but through the time and language change it continues to be elusive, it varies and changes according to the contexts.

However, this case obviously contains most of the noise units (73.84 % noise).

One more noticeable aspect in this case of words listed as the hyponyms of adjective "*puikus*" is that two different senses in two different dictionaries have got many word matches from the SM generated thesaurus. In DLKZ the second sense (2. *labai geras, šaunu*s (tr. very good, nice)) inherited 16 equivalents from the SM thesaurus, while the first sense was connected with 6 and the third one with 5 equivalents. This would suggest that the second sense in the golden standard might rather have be listed as the first one. However, getting to the other golden standard LKZ, we notice that since the first sense is exactly the same as in the previous dictionary, it does not have the exact equivalent of the second sense in DLKZ but there appears to be an other sense which gains most of equivalents from SM - the seventh one (7. *nemažas, didelis, žymus; tikras* (tr. significant, big, strong, real)). It inherits 22 equivalents from SM, while the 1st one inherits 6, 2nd -3, 3rd - 3, 4th - 0, 5th - 10, 6th - 8, 8th - 0 and 9th - 2. A word senses in DLKZ are arranged so that first would be listed the most widely known ones, followed by the less common, older and more specific senses. The senses that were matched with more equivalents from the SM can be concluded to be of a wder meaning. The prioritization according to the SM gives us the suggestion of different kind of hierarchy of senses than in our gold standards. According to SM, senses listed in DLKZ would describe an adjective "*puikus*" more accurately if they were listed in this sequence:

puikus, ~ì

1. labai geras, šaunus (tr. very good, nice)

Correspondences from SM: pasigėrėtinas, fantastiškas‹1›, fenomenalus‹1›, geras‹1›, išskirtinis‹1›, ›, kokybiškas‹1›, nekasdieniškas‹1›, nepakartojamas, nepaprastas‹1›, nuostabus‹1›, pasigėrėtinas, pribloškiantis, stulbinamas‹1›, stulbinantis‹1›, įspūdingas‹1›, šaunus‹1›

2. labai gražus, dailus (tr. very nice, pretty)

Correspondences from SM: grežus‹1›, gražutis, išvaizdus, neapsakomas‹1›, simpatingas, žavingas‹1›

3. Išdidus (tr. Proud) -

Correspondences from SM: didingas‹1›, didis, didus‹1›, orus‹1›

In LKZ it would be more accurate if it were in this sequence:

puikus, -ì adj.

 1. nemažas, didelis, žymus; tikras (tr. significant, big, strong, real)

Correspondences from SM: didingas‹1›, didis, didus‹1›, augalotas, aukštas‹1›, begalinis‹1›, bekraštis, beribis‹1›, didelis‹1›, didysis, didžiulis‹1›, erdvus‹1›, gausus‹1›, gremėzdiškas‹1›, ilgas‹1›, masyvus‹1›, milžiniškas‹1›, žymus nemažas‹1›, nemenkas, platus, ženklus‹1›

 2. puošnus, prabangus, turtingas (tr. gorgeous, luxurious, rich)

Correspondences from SM: iškilmingas, iškilus‹1›, , auksinis‹1›, brangus‹1›, ekstravagantiškas, iškilmingas, iškilus‹1›, pompastiškas‹2›, turtingas‹1›, vertingas

 3. šaunus, smarkus; garbingas (tr. cool, intense, honourable)

Correspondences from SM: didingas‹1›, didis, didus‹1›, įspūdingas‹1›, šaunus‹1›, nenumaldomas‹1›, siautulingas, smarkus‹1›

 4. labai gražus, dailus (tr. very nice, pretty)

Correspondences from SM: gražus‹1›, gražutis, išvaizdus, neapsakomas‹1›, simpatingas, žavingas‹1›

 5. tinkamas, patogus, parankus, geras (tr. appropriate, convenient, handy, good) -

Correspondences from SM: geras‹1›, išskirtinis‹1›, naudingas‹1›

and

 malonus, patrauklus (tr. pleasant, attractive)

Correspondences from SM: pasigėrėtinas, pasigėrėtinas, ›, malonus‹1›

 6. kuris aukštos kilmės, kilnus (tr. a high-ranking, noble)

Correspondences from SM: didysis, kilnus

 7. doras, teisingas, žmoniškas (tr. honest, fair, humane)

No correspondences from SM

and

 išdidus, išpuikęs (tr. proud, haughty)

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:

Dictionary of the Lithuanian Language
išdidus
išpuikęs

Modern Lithuanian Dictionary

Semantic Mirrors

aitrus	jaudinantis
aktyvus	kapitalinis
apgalvotas	kategoriškas
apsukrus	klestintis
apčiuopiamas	kruopštus
aršus	lemiamas
atidus	lemtingas
besaikis	liūdnas
brandus	magiškas
būtinas	maksimalus
dažnas	matomas
dosnus	mažas
dramatiškas	mįslingas
drūtas	narsus
egzotiškas	naujas
galingas	neaprėpiamas
galintis	neatidėliotinas
galvotas	neatskiriamas
garsus	neblogas
genialus	neeilinis
giluminis	nenormalus
gilus	nenusakomas
glaudus	neperskiriamas
globalinis	nesavas
globalus	netikėtas
griežtas	neįkainojamas
grėsmingas	neįsivaizduojamas
gudrus	neįtikėtinas
intensyvus	neįveikiamas
intriguojantis	nuodugnus
įspūdingas	nuoširdus
išauklėtas	nusipelnęs
išimtinis	opus
iškalbingas	padidėjęs
išmintingas	pagarsėjęs
išmoningas	pagrindinis
išplėstas	pakankamas
išpūstas	pakilus
išraiškingas	pamatinis
išsipūtęs	paplitęs
išskaidytas	pasibaisėtinas
ištvermingas	pastebimas

augalotas iškilmingas
auksinis iškilus
aukštas kilnus
begalinis malonus
bekraštis masyvus
beribis milžiniškas
brangus naudingas
didelis nemažas
didysis nemenkas
didžiulis nenumaldomas
ekstravagantiškas platus
erdvus pompastiškas
gausus siautulingas
grėmėzdiškas smarkus
ilgas turtingas

didingas
didis
dieviškas
geras
gražus
gražutis
išskirtinis
išvaizdus
neapsakomas
pasigėrėtinas
simpatingas
įspūdingas
šaunus
žavingas

patenkintas
patikimas
patobulintas
pavojingas
pašėlęs
pelningas
populiarus
pozityvus
prasmingas
putnus
radikalus
rafinuotas
reikiamas
reikšmingas
reikšminis
retas
riebus
rinktinis
ryžtingas
savotiškas
sensacingas
siaubingas
skambus
skardus
skaudus
skiriamasis
slaptas
smagus
smailus

smulkmeniškas
sodrus
solidus
sotus
spartus
spindulingasis
stebuklingas
stebėtinas
stiprus
stambus
stangrus
status
stebinantis
stebuklingas
stebėtinas
stiprus
storas
stuomeningas
sumanus
sumanytas
sunkus
supratingas
susikaupęs
suveltas
svaiginantis
svarbus
svarus
samojingas
sėkmingas
tankus

tariamas
tarptautinis
taurus
teigiamas
tirštas
triuškinantis
tuščias
unikalus
valstybinis
veiksmingas
vešlus
vibruojantis
vidutinis
viliojantis
visiškas
vyraujantis
vyriausias
ypatingas
įdomus
įdėmus
įkritęs
įsidėmėtinas
įstabus
įtaigus
įtakingas
įtikimas
įvairus
įžvalgus
įžymus

fantastiškas
fenomenalus
nekasdieniškas
nepakartojamas
nepaprastas
neprilygstamas
kokybiškas
neregėtas
nuostabus
orus
priblokštiantis
stulbinamas
stulbinantis

šiuolaikiškas
šiurpus
šlovingas
šviesus
žinomas

Figure 6.2.2 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

There are adjectives among the ones that are 'noise', which do mean "*puikus*" in some way, and represent its semantics, even though they were not listed as a separate sense in the dictionaries. Here is a list of adjectives that can be connected with the meaning of the adjective "*puikus*": 1 mental brightness - *"atidus"* (tr. attentive), *"galvotas"* (tr. brainy), *"genialus"* (tr. genial), *"gudrus"* (tr. clever), *"išmintingas"* (tr. wise), *"išmoningas"* (tr. ingenious), *"sumanus"* (clever), *"sąmojingas"* (tr. conscious), *"įžvalgus"* (tr. shrewd); 2 of a fast, energetic movement - *"aktyvus"* (tr. active), *"intensyvus"* (tr. intensive), *"spartus"* (tr. quick)*; 3* well considered or well done - *"apgalvotas"* (tr. considered), *"atidus"* (tr. attentive), *"kruopštus"* (tr. thorough), *"nuodugnus"* (tr. thorough), *"patikimas"* (tr. reliable); 4 *"brandus"* (tr. mature); 5 stong, powerful, well built - *"drūtas"* (tr. thick), *"galingas"* (tr. powerful), *"ištvermingas"* (tr. persevering), *"neįveikiamas"* (tr. compulsive), *"stambus"* (tr. large), *"stangrus"* (tr. resilient), *"stiprus"* (tr. strong), *"stuomeningas"* (tr. handsome); 6 *"dosnus"* (generous); 7 *"egzotiškas"* (tr. exotic); 8 *"glaudus"* (tr. close); 9 *"intriguojantis"* (tr. gripping), *"jaudinantis"* (tr. moving, exciting); 10 of a good maners, taste - *"išauklėtas"* (tr. polite), *"rafinuotas"* (tr. sophisticated), *"solidus"* (tr. grave); 11 not casual - *"išimtinis"* (tr. exceptional), *"neeilinis"* (tr. uncommon); 12 *"išraiškingas"* (tr. expressive); 13 profitable, (firm/business) that works well - *"klestintis"* (tr. prosperous), *"pelningas"* (tr. profitable), *"sėkmingas"* (tr. successful); 14 special - *"magiškas"* (tr. magic), *"nenusakomas"* (tr. nondescript), *"neįtikėtinas"* (tr. unbelievable), *"svaiginantis"* (tr. heady), *"unikalus"* (tr. unique), *"ypatingas"* (tr. special); 15 *"maksimalus"* (tr. maximal); 16 *"narsus"* (tr. brave); 17 *"neįkainojamas"* (tr. invaluable); 18 *"nuoširdus"* (tr. frank), *"taurus"* (tr. sublime); 19 *"nusipelnęs"* (tr. deserved); 20 *"pagarsėjęs"* (tr. renowned); 21 *"pakilus"* (tr. elated), *"smagus"* (tr. funny); 22 *"populiarus"* (tr. popular), 23 *"pozityvus"* (tr. positive), *"teigiamas"* (tr. positive), 24 valuable work, results - *"prasmingas"* (tr. meaningful), *"reikšmingas"* (tr. significant), *"svarus"* (tr. weighty), *"veiksmingas"* (tr. efficient); 25 *"sodrus"* (tr. lush), *"šviesus"* (lucid); 26 *"supratingas"* (tr. understanding); 27 *"viliojantis"* (tr. tempting); 28 *"įdomus"* (tr. interesting); 29 *"įvairus"* (tr. various); *30 "šlovingas"* (tr. glorious).

Again, we experience the limitation of our golden standard. Therefore, it is a challenging task to evaluate the SM method in a best way.

III.   Didelis

The Semantic Mirrors

(*Translation:* deep, good, great.)

Hyperonyms: nuostabus‹1›, puikus‹1›.

Hyponyms: absoliutus‹1›, augalotas, aukštas‹1›, baisus‹1›, begalinis‹1›, beribis‹1›, didus‹1›, didžiulis‹1›, drūtas‹1›, duslus‹2›, erdvus‹1›, esminis, galingas‹1›, garsus‹1›, geras‹1›, giluminis‹1›, gilus‹1›, globalus, griežtas‹1›,grėsmingas, išpūstas‹1›, išsamus‹2›, kapitalinis‹1›, kimus‹1›, maksimalus‹1›, masyvus‹1›, milžiniškas‹1›, nemalonus‹2›, nenumaldomas‹1›, neįveikiamas‹2›, nuodugnus, nuoširdus‹1›, pagrindinis‹1›, pakankamas‹1›, pilnas‹2›,putnus‹1›, ryškus‹1›, skambus‹1›, skardus‹1›, skaudus‹1›, slaptas‹2›, sodrus, stiprus‹1›,storas‹1›, subtilus, sudėtingas‹1›, suveltas‹1›, tamsus‹2›, tankus‹1›, tarptautinis, tobulas‹1›, triuškinantis, tvirtas‹1›, tylus‹2›, užkimęs‹1›,vibruojantis, visuotinis‹1›, vyraujantis‹1›, vyriausias, čaižus‹1›, įkritęs‹1›, įtaigus‹1›, šaižus‹2›, žemas‹2›.

DLKZ (Modern Lithuanian Dictionary)

didelis, ~ė

prš. m a ž a s. (tr. Opposite - small)

1. žymus savo apimtimi, dydžiu (aukščiu, ilgiu, pločiu) (tr. significant in scope, size (height, length, width))

Correspondences from SM: augalotas, aukštas‹1›, begalinis‹1›, beribis‹1›, didžiulis‹1›, erdvus‹1›, gilus‹1›, globalus, masyvus‹1›, milžiniškas‹1›, putnus‹1›, storas‹1›

2. suaugęs, subrendęs (tr. grown-up, mature)

No correspondences from SM

3. gausus, apstus (tr. Rich, numerous)

Correspondences from SM: garsus‹1›, maksimalus‹1›, pakankamas‹1›, pilnas‹2›

4. smarkus, stiprus, intensyvus (tr. Vigorous, strong, intensive)

Correspondences from SM: drūtas‹1›, galingas‹1›, stiprus‹1›, triuškinantis, tvirtas‹1›, nenumaldomas‹1›

5. svarbus, reikšmingas (tr. important, significant)

Correspondences from SM: esminis, visuotinis‹1›, vyraujantis‹1›

6. žymus, garsus, garbingas; didis (tr. Considerable, famous, respectable, great)

Correspondences from SM: didus‹1›

LKZ (the Dictionary of the Lithuanian Language)

didelis, -ė adj.

1. žymus savo apimtimi, dydžiu (aukščiu, ilgiu, pločiu, storiu...) (tr. significant in scope, size (height, length, width, thickness ...))

Correspondences from SM: augalotas, aukštas‹1›, beribis‹1›, didžiulis‹1›, erdvus‹1›, gilus‹1›, globalus, masyvus‹1›, milžiniškas‹1›, putnus‹1›, storas‹1›

2. suaugęs, subrendęs (tr. Grown-up, mature)

No correspondences from SM

3. žymus skaičiumi, kiekiu, gausus, apstus (tr. significant in number or content, rich, numerous)

Correspondences from SM: maksimalus‹1›, pakankamas‹1›, pilnas‹2›

4. smarkus, stiprus, intensyvus (tr. Vigorous, strong, intensive)

Correspondences from SM: drūtas‹1›, galingas‹1›, stiprus‹1›, triuškinantis, tvirtas‹1›, nenumaldomas‹1›

5. svarbus, reikšmingas, žymus (tr. Important, significant, considerable)

Correspondences from SM: esminis, garsus‹1›, visuotinis‹1›, vyraujantis‹1›, didus‹1›

6. ypatingas, nepaprastas, tikras (tr. special, extraordinary, real)

Correspondences from SM: geras‹1›, ryškus‹1›

7. išdidus (tr. Proud)

No correspondences from SM

8. ilgas (laiko atžvilgiu) (tr. Long (in time vector))

Correspondences from SM: begalinis‹1›

9. pats, tikras (tr. Same, real)

Correspondences from SM: absoliutus‹1›, ryškus‹1›

10. nemažybinis, nemaloninis (tr. not diminutive, not hypocoristic)

No correspondences from SM

The Qualitative Evaluation

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 29.

All the rest of the adjectives that were returned by SM and named noise are 37. 37 + 29 = 66.

P = 29 / 66 = **0.439**

The number of the results that should have been returned is 17. And the new correct results are added: $29 + 17 = 46$.

R = 29 / 46 = **0.63**

F = 0.518

In this case, the first sense in both dictionaries is the same and appeared to be matched with the most adjectives generated by SM as hyponyms (which were rather classified as senses). However, by ability to automatically determine adjectives which can be sorted as senses of the adjective "*didelis*" in our golden standard (even though the list was named as 'hyponyms'), the result in this case is quite plausible: the main sense from both golden standards was recognised the best.

There is a slight difference among the classification of automatically generated adjectives and senses of the golden standard. It appeared, that the adjective "*begalinis*" can be classified as two senses in DLKZ. It happens because the adjective "*begalinis*" according the DLKZ means:

begalinis
neribotas erdvės, laiko, dydžio, kiekio atžvilgiu (tr. unlimited in terms of quantity (space, time, and size)

The equivalents to which this adjective can be matched are the 1st sense in LKZ and the 1st and 8th sense in DLKZ:

LKZ

1. žymus savo apimtimi, dydžiu (aukščiu, ilgiu, pločiu) (tr. significant in scope, <u>size</u> (height, length, width))

DLKZ

1. žymus savo apimtimi, dydžiu (aukščiu, ilgiu, pločiu, storiu...) (tr. significant in scope, <u>size</u> (height, length, width, thickness ...))
8. ilgas (laiko atžvilgiu) (tr. Long (in <u>time</u> vector))

This case shows the intersection among the senses in one dictionary (DLKZ): the same adjective can be classified to two different senses. This suggests, that these two senses (1and 8) in DLKZ should rather be merged and become one sense.

As in most cases it is a challenging task to determine the hierarchical relations between adjectives, in this case we might try to prove the hyperonym-hyponym relations among adjectives listed as hyperonyms of "*didelis*". We may take a look to how dictionaries explain the hyperonyms "*nuosabus*" and "*puikus*" in this case:

LKZ

nuostabus, -ì adj. - įstabus, nepaprastas (tr. wonderful, extraordinary)

puikus, -ì adj.

1. labai gražus, dailus (tr. very nice, pretty)

 2. tinkamas, patogus, parankus, geras (tr. appropriate, convenient, handy, good)

3. malonus, patrauklus (tr. pleasant, attractive)

 4. doras, teisingas, žmoniškas (tr. honest, fair, humane)

 5. puošnus, prabangus, turtingas (tr. gorgeous, luxurious, rich)

 6. šaunus, smarkus; garbingas (tr. cool, intense, honourable)

 7. nemažas, didelis, žymus; tikras (tr. significant, big, strong, real)

 8. išdidus, išpuikęs (tr. proud, haughty)

 9. kuris aukštos kilmės, kilnus (tr. a high-ranking, noble)

DLKZ

nuostabus - keliantis susižavėjimą, nepaprastas (tr. admirable, extraordinary)

puikus

1. labai gražus, dailus (tr. very nice, pretty)

2. labai geras, šaunus (tr. very good, dashing)

3. išdidus (tr. proud)

Starting with the adjective "*nuostabus*" to be the hyperonym of the adjective "*didelis*" there can be given few examples confirming these relations:

One stating that "*Šitas namas yra nuostabus*" (tr. This house is wonderful/admirable/extraordinary) can specify the adjective "*nuostabus*" with other adjectives which are more concrete, and which mostly are the features of a house that gains a label of being wonderful/admirable/extraordinary ("*nuostabus*"). One of those features is being big (commodious, roomy) - "*didelis*".

Since any hyponymous word can be included into the concept of hyperonym and it should have more specialized, narrowed meaning, to be more concrete and understandable we might conclude that "*nuostabus*" can be a hyperonym of "*didelis*". Also, as in theory, a hyperonym expresses a more abstract meaning, intuitively we may decide that the adjective "*nuostabus*" has a wider meaning than the adjective "*didelis*".

As for the adjective "*puikus*" to be explained as the hyperonym of the adjective "*didelis*", the very same demonstrational statement can be used: "*Šitas namas yra puikus*" (tr. *This house is amazing*.). Again, for the house to be labeled as "*puikus*" (most probably for any language user to decide) it will require the feature of big size. By the same logical interpretation we can conclude that SM represented not popular, but possible hierarchical relations among the adjectives rather than only vertically spacial (synonymy, antonymy) ones. However, even if the hyperonymy-hyponymy relations existed in this way, they would be possible only in this particular meanings of sentences that I discuss above.

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:

Figure 6.2.3 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

There are adjectives among the ones that are 'noise', which do mean "*didelis*" in some way, and represent its semantics, even though they were not listed as a separate sense in the dictionaries. Here is a list of adjectives that can be connected with the meaning of "*didelis*": "*nuostabus*", "*puikus*", "*vyriausias*". For example:

Šis namas yra nuostabus/puikus. (tr. This house is amazing, spacious, <u>big</u>.).

Darius yra Tomo vyriausias brolis. (tr. Darius is Tom's <u>big</u> brother.).

The repetitive note can be made in this case too: we experience the limitation of our golden standard, which do not allow to evaluate the SM method with the fullest coverage.

IV. Sumanus

The Semantic Mirrors

Hyperonyms: šaunus‹1›, geras‹1›, puikus‹1›.

Subsense (i)
(*Translation:* bright, smart, clever.)
Synonyms: apdairus, apgalvotas, apsukrus, galvotas‹1›, gudrus‹1›, išradingas‹1›, išsilavinęs, miklus‹1›, ne prilygstamas‹1›, nuovokus, sąmojingas‹1›.

Subsense (ii)
(*Translation:* quick, acute.)
Synonyms: aktualus‹1›, staigus‹2›, ūmus‹1›.
Related
words: akivaizdus‹1›, dažnas‹1›, energingas‹1›, greitas‹1›, iškalbingas‹2›, judrus‹1›, miklus‹1›, opus‹1›, s kubus, spartus‹1›, svarbiausias, trumpas‹2›, įžvalgus‹2›, žvitrus‹1›.

Subsense (iii)
(*Translation:* capable.)
Synonyms: galimas‹2›.
Related
words: gabus‹1›, galintis‹1›, išmintingas‹1›, kompetentingas‹1›, pajėgus, pasirengęs‹1›, prieinamas‹1›, tal entingas‹1›.

DLKZ (Modern Lithuanian Dictionary)

sumanus

greitai suvokiantis, gudrus (tr. quickly perceiving, clever)

Correspondences from SM:
apdairus, apgalvotas, apsukrus, galvotas‹1›, gudrus‹1›, išradingas‹1›, miklus‹1›, nuovokus, sąmojingas‹1›, įžvalgus‹2›, gabus‹1›, galintis‹1›, išmintingas‹1›, kompetentingas‹1›

LKZ (the Dictionary of the Lithuanian Language)

sumanus adj.

1. kuris greitai sumeta, susiorientuoja, nuovokus (tr. Someone who is quickly perceiving, orienting himself, perceptive)

Correspondences from SM: apdairus, apsukrus, miklus‹1›, nuovokus, sąmojingas‹1›, įžvalgus‹2›, kompetentingas‹1›

2. kuris sugeba ką su išmone, išradingai padaryti; galvotas (tr. who are capable of doing something with notion, ingeniously done, intelligent)

Correspondences from SM: apgalvotas, galvotas‹1›, gudrus‹1›, išradingas‹1›, gabus‹1›, galintis‹1›, išmintingas‹1›, talentingas‹1›

The Qualitative Evaluation

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 15.

All the rest of the adjectives that were returned by SM and named noise are 25. 25 + 15 = 40.

P = 15 / 40 = **0.375**

The number of the results that should have been returned is 3. And the new correct results are added: 25 + 3 = 28.

R = 15 / 28 = **0.536**

F = 0.441

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:
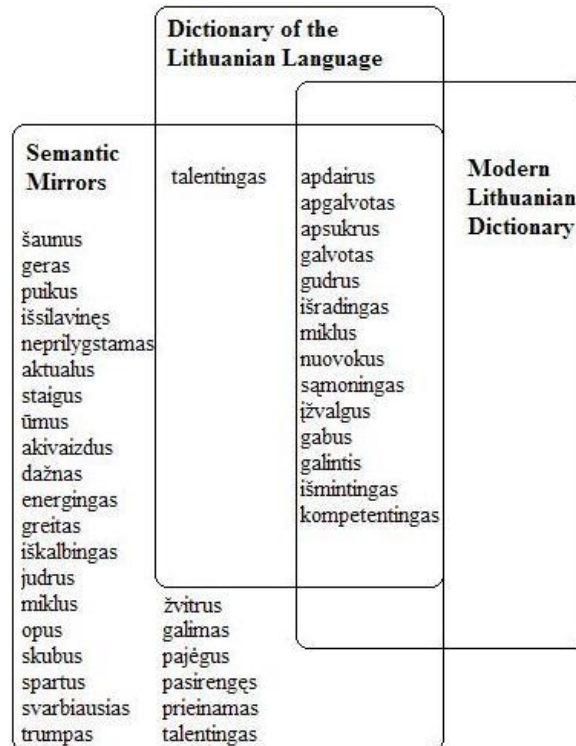

Figure 6.2.4 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

There are no adjectives among the ones that are listed as 'noise', which mean "*sumanus*" in any way. So all the 'noise' in this case is grouped correctly.

V. Senas

The Semantic Mirrors

Sense 2
Subsense (i)
(*Translation:* old.)
Synonyms: žilas‹3›.

Subsense (ii)
(*Translation:* early.)
Synonyms: pradinis‹1›.
Related words: ankstesnis‹2›, jaunas‹2›, naujas, vėlus‹2›.

Sense 3
Hyperonyms: svarbus.

Subsense (i)
(*Translation:* stark.)
Synonyms: skaudus‹1›.
Related words: absoliutus‹1›, atkaklus‹2›, atšiaurus‹1›, išraiškingas‹1›, stulbinamas‹1›, sustingęs‹2›.

Subsense (ii)
(*Translation:* long.)
Synonyms: ilgas‹1›.

DLKZ (Modern Lithuanian Dictionary)

---

senas

1. turintis daug amžiaus; (tr. in its old age)

Correspondences from SM: žilas‹3›

2. jau kuris laikas esantis; prš. n a u j a s 1 (tr. existing for some time already; opposite - new)

Correspondences from SM: ilgas‹1›

3. ilgai vartotas, palaikis (tr. long used, aged)

No correspondences from SM

4. anksčiau buvęs, nedabartinis, pasenęs (tr. former, not modern, outdated)

No correspondences from SM

---

5. pirma, prieš tai buvęs (tr. something what was before, previous)

Correspondences from SM: pradinis‹1›, ankstesnis‹2›

6. kuris pabuvęs, palaikytas, pastovėjęs; prš. š v i e ž i a s 1 (tr. something that stayed longer, was held longer; opposite - fresh)

No correspondences from SM

LKZ (the Dictionary of the Lithuanian Language)

1. turintis daug amžiaus, sulaukęs senatvės, nejaunas (apie žmones, gyvulius); ilgai augantis (apie augalus); pršn. Jaunas (tr. being old, reached eldery age, not young (human or animal); opposite - young)

seniai praėjęs, ilgai užsitęsęs (apie amžių, metus) (tr. long gone protracted (age, years))

Correspondences from SM: žilas‹3›

2. jau kuris laikas esantis, seniai atsiradęs, padarytas, įsigytas (tr. some time existing, appeared or made or bought long time ago)

patyręs, įjunkęs (tr. experienced, practiced)

išlaikytas (apie gėrimą) (tr. maintained (drink, e.g. vine))

Correspondences from SM: ilgas‹1›

3. ilgai naudotas, vartotas, palaikis (tr. long used, maintained)

nebegaliojantis, pavartotas (tr. expired, used)

No correspondences from SM

4. seniai praėjęs (apie laiką) (tr. long ago passed (about time))

susijęs su ankstesniais laikais, ne šiuolaikinis (tr. relates to the past, not modern)

No correspondences from SM

5. prieš tai buvęs, ankstesnis už esamą (tr. previous, earlier than the current)

Correspondences from SM: pradinis‹1›, ankstesnis‹2›

praeinantis, besibaigiantis (tr. finishing, about to expire)

Correspondences from SM: vėlus‹2›.

susijęs su atgyvenusia santvarka (tr. associated with an outdated system)

6. nešviežias (tr. not fresh)

<div style="border:1px solid black;">

No correspondences from SM

7. pilnas (apie mėnulio fazę) (tr. full (phase of the moon))

No correspondences from SM

 8. (germ.) sulaukęs tam tikro amžiaus (tr. reached a certain age)

seniai žinomas dalykas (tr. long known thing)

senoviškas (tr. antediluvian)

daug matęs, patyręs žmogus (tr. veteran, experienced man)

No correspondences from SM

</div>

The Qualitative Evaluation

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 5.

All the rest of the adjectives that were returned by SM and named noise are 11. $11 + 5 = 16$.

P $= 5 / 16 =$ **0.313**

The number of the results that should have been returned is 10. And the new correct results are added: $5 + 10 = 15$.

R $= 5 / 15 =$ **0.33**

F = 0.323

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:
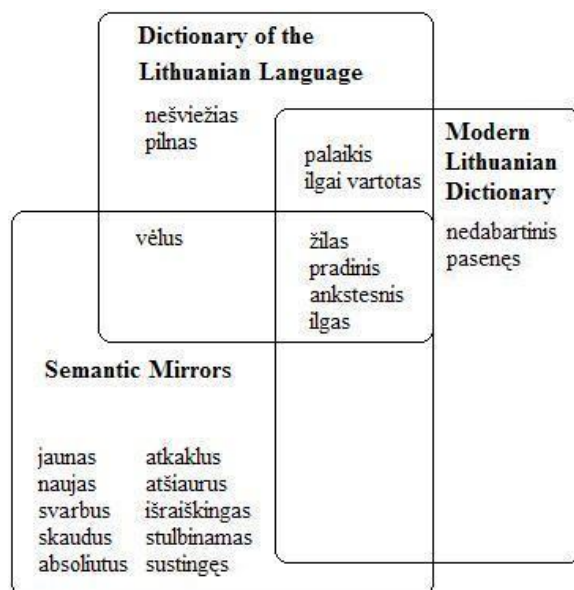
Figure 6.2.5 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

In this case two senses with two subsenses were generated by SM. The first sense contains subsenses and synonyms. Both of the synonyms and the subsenses very well represent senses in the golden standard. However, in the second subsense next to the synonym that matches a certain sense in the dictionary and two related words, two antonyms ("*jaunas*" (tr. young) and "*naujas*" (tr. new)) are given as well. This was an odd result so the lattice was inspected:
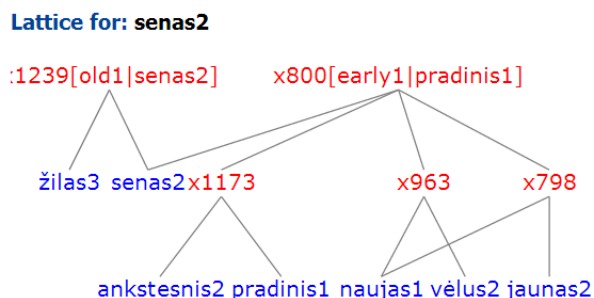

Figure 6.2.5.1 The lattice of the adjective "*senas*"

From the lattice above we can see that the antonym of "*senas*" is obtained from the adjective "*early*". We can easily notice how two fields of related words are formed here: one contains adjectives "*žilas", "senas", "ankstesnis", "pradinis*", another - "*naujas", "vėlus", "jaunas*".

This refers to a different partition of adjective's "*senas*" meaning - the sense "*old*" and the antonymy "*young*/*new*" (jaunas, naujas).

The appearance of antonyms in this case might come from various possibilities of translating the adjective "*early*" to the Lithuanian language. For example:

"Early" as "*ansktyvas"* (tr. early)

That was the box I kept by my typewriter where my ideas lay and spoke to me <u>early</u> mornings to tell me where they wanted to go and what they wanted to do.

Šią dėžę laikiau prie rašomosios mašinėlės, ten gulėjo mano idėjos, <u>ankstyvą</u> rytą prabildamos ir sakydamos man, kur jos nori eiti ir ką nori daryti.

"Early" as "*jaunas"* (tr. young*)*:

When the Barcelona European Council called for the establishment of the indicator, it also observed that the teaching of at least two foreign languages from a very <u>early</u> age was an important part of the basic skills – part of the birthright of all European citizens.

Barselonos Europos Vadovų Taryba, pakviesdama sukurti indikatorių, taip pat pastebėjo, kad mažiausiai dviejų užsienio kalbų mokymas nuo labai <u>jauno</u> amžiaus buvo svarbi pagrindinių įgūdžių dalis – visų Europos piliečių prigimtinės teisės dalis.

Early as *senas* (tr, old):

Others, again, spoke of some <u>early</u> love affair, and of a fair-haired girl who had pined away on the shores of the Atlantic.

Kai kurie tvirtino, kad tokio užsispyrimo priežastis esanti <u>sena</u> meilė ir kad kur nors Atlanto vandenyno pakrantėje jo laukianti šviesbruvė gražuolė.

The previous three examples are taken from the corpus which was previously used for data collection.

So the wide range of the translation possibilities of the adjective "*early*" in the Lithuanian language distorted the thesaurus results and made them contain not only senses representing the adjective "*senas*" but its antonyms too.

VI. Geras

The Semantic Mirrors

(*Translation:* bright, good, strong, deep, great.)

Hyperonyms: svarbus, didelis‹1›, nuostabus‹1›, puikus‹1›.

Hyponyms: akinamas, apgalvotas, apšviestas‹1›, aštrus‹1›, baltas‹2›, blizgantis, blykčiojantis, dailus‹1›, džiugus‹1›, gabus‹1›, giedras, gudrus‹1›, guvus‹1›, gyvas‹1›, intensyvus‹1›, išradingas‹1›, linksmas‹2›, mirguliuojantis,nutviekstas‹1›, paslaptingas‹1›, protingas‹1›, saulėtas‹1›, skaistus‹1›, spindintis‹1›, spindulingas, spindulingasis, spinduliuojantis, stačiokiškas, sumanus‹1›, sąmojingas‹1›, tirštas‹1›, tviskantis‹1›, vaiskus‹1›, šaunus‹1›,švarus‹1›, šviesus‹1›, šviečiantis‹1›, švytintis, žibantis, žvilgantis‹1›, žėrintis, žėruojantis‹1›.

DLKZ (Modern Lithuanian Dictionary)

---

geras
prš. b l o g a s. (tr. Opposite - bad)
1. turintis teigiamų ypatybių, tinkamas, naudingas (tr. with positive characteristics, appropriate, beneficial)

Correspondences from SM: apgalvotas

2. malonus, gailestingas, nepiktas (tr. pleasant, gracious, not angry)

No correspondences from SM

3. mokantis savo darbą, gabus, sumanus (tr. knowing their work, gifted, smart)

Correspondences from SM: gabus‹1›, gudrus‹1›, išradingas‹1›, protingas‹1›, sumanus‹1›, sąmojingas‹1›

4. teikiantis pasitenkinimą, patogus, jaukus (tr. satisfying, comfortable, cozy)

Correspondences from SM: džiugus‹1›, šaunus‹1›, švarus‹1›, šviesus‹1›

5. nemažas, didelis (tr. significant, large) -

Correspondences from SM: didelis‹1›

6. pelningas (tr. Profitable)

No correspondences from SM

7. sveikas, stiprus (tr. healthy, strong)

| **Correspondences from SM:** guvus‹1›, gyvas‹1›, intensyvus‹1› |

LKZ (the Dictionary of the Lithuanian Language)

---

geras, -à (nom. pl. gerì, gẽrūs) adj.

1. meilus, malonus; gailestingas; nepiktas (tr. affectionate, kind, compassionate, not angry)

No correspondences from SM

2. doras, teisingas, žmoniškas (tr. honest, fair, humane)

No correspondences from SM

3. gabus, gudrus, apsukrus (tr. gifted, clever, shifty)

Correspondences from SM: gabus‹1›, gudrus‹1›, išradingas‹1›, protingas‹1›, sumanus‹1›, sąmojingas‹1›

4. tinkamas; patogus; naudingas; vertingas; tikras; parankus (tr. appropriate, comfortable, useful, valuable, genuine, handy)

Correspondences from SM: apgalvotas

 (sausas, giedras) (tr. dry, cloudless)

Correspondences from SM:
giedras, saulėtas‹1›, skaistus‹1›, spindintis‹1›, spindulingas, spindulingasis, spinduliuojantis

5. nemažas, didelis, žymus; tikras, atsakomas (tr. significant, big, strong, confident, answerable)

Correspondences from SM: didelis‹1›

6. sveikas, stiprus(tr. healthy, strong)

Correspondences from SM: guvus‹1›, gyvas‹1›, intensyvus‹1›

7. laimingas, vykęs (tr. happy, felicitous)

Correspondences from SM: džiugus‹1›, šaunus‹1›, šviesus‹1›

8. pelningas (tr.profitable)

No correspondences from SM

9. garbingas (tr. honorable)

No correspondences from SM

The Qualitative Evaluation

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 22.

All the rest of the adjectives that were returned by SM and named noise are 24. 24 + 22 = 46.

$P = 22 / 46 = $ **0.48**

The number of the results that should have been returned is 31. And the new correct results are added: 31 + 22 = 53.

$R = 22 / 53 = $ **0.42**

$F = $ **0.444**

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:
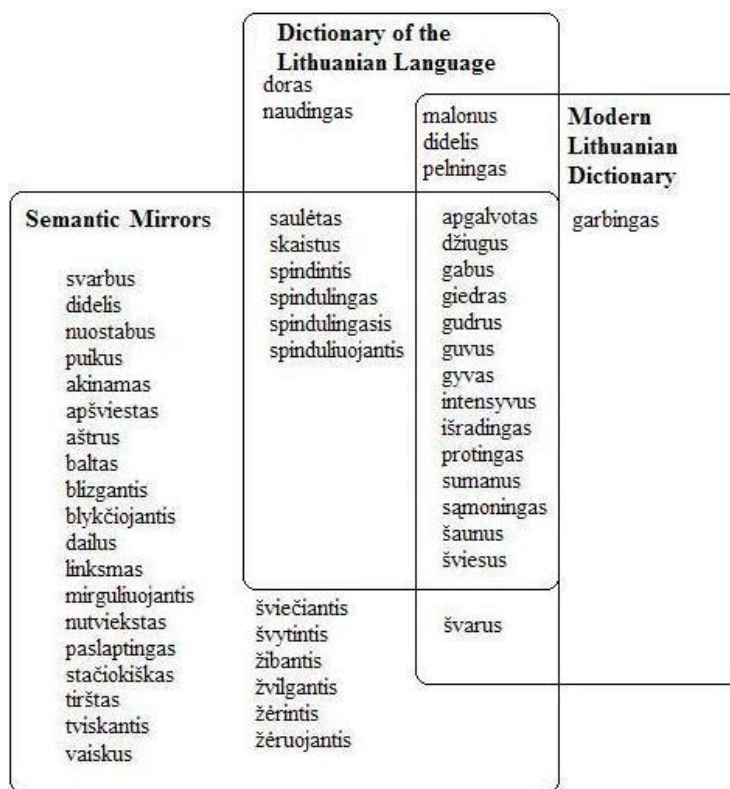

Figure 6.2.6 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

There are adjectives among the ones that are 'noise', which do mean "*geras*" in some way, and represent its semantics, even though it was not listed as a separate sense in the dictionaries. Here is a list of adjectives that can be connected with the meaning of "*geras*": "*puikus*" (tr. excellent), "*aštrus*" (tr. sharp), "*dailus*" (tr. nice), "*linksmas*" (tr. cheerful). For example:

Puikus darbas! (tr. <u>Good</u> job!).

Aštrus peilis. (tr. Sharp/<u>good</u> knife.).

Dailus suknelės pasiuvimas. (tr. <u>Good</u>/nicely made dress.).

Linksmas buvo senelio būdas. (tr. The grandfather was always in a <u>good</u> mood.)

Again, the limitations of the golden standard is noticeable in the examples listed above.

In this case we got four adjectives which were automatically listed as hyperonyms of the adjective "*geras*". However, as it has been already mentioned before, it is not easy to tie these adjectives with hierarchical relations. The adjective "*svarbus*" (tr. important) has a rather distant meaning from the meaning of the adjective "*geras*" (tr. good). The adjective "*didelis*" (tr. big) is considered to be one of the sense of the adjective "*geras*", but not the hyperonym (DLKZ - 5. nemažas, didelis (tr. significant, large), LKZ - 5. nemažas, didelis, žymus; tikras, atsakomas (tr. significant, big, strong, confident, answerable)). The adjective "*puikus*" (tr. amazing) is the synonym of the adjective "*geras*":

The synonyms of "*geras*":

"*Labas*" (tr. good), "*lemtas*" (tr. fateful), "*doras*" (tr. honest), "*padorus*" (tr. decent), "*šaunus*" (tr. dear), "*puikus*" (tr. <u>amazing</u>), "*taurus*" (tr. sublime), "*vykęs*" (tr. successful), "*panašus*" (tr. similar), "*valyvas*" (tr. tidy), "*žmoniškas*" (tr. humane), "*neblogas*" (tr. not bad), "*nepeiktinas*" (tr. irreaproachable), "*nebartinas*" (tr. not blamable), "*nepriekaištingas*" (tr. perfect). (http://www.lietuviuzodynas.lt/sinonimai/Geras).

Because the adjective "*nuostabus*" (tr. amazing) is semantically related to "*geras*" but "*svarbus*" has a rather distinct meaning, the lattice of this thesaurus entry was inspected:
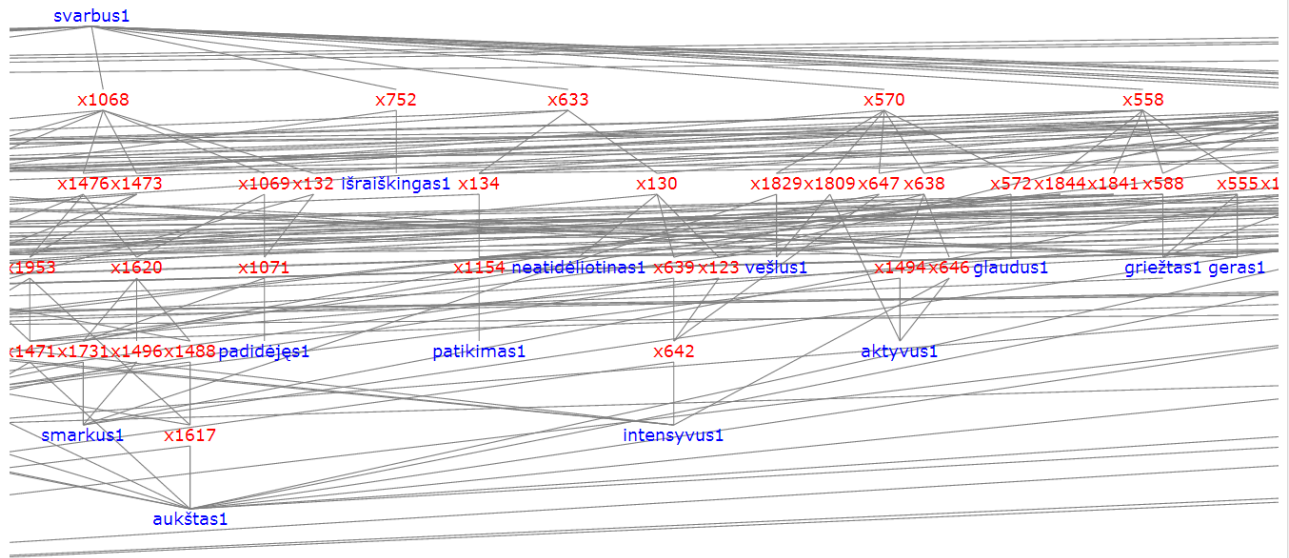
Lattice for: geras1

svarbus1

x1068 x752 x633 x570 x558

x1476 x1473 x1069 x132 išraiškingas1 x134 x130 x1829 x1809 x647 x638 x572 x1844 x1841 x588 x555 x1

x1953 x1620 x1071 x1154 neatidėliotinas1 x639 x123 vešlus1 x1494 x646 glaudus1 griežtas1 geras1

x1471 x1731 x1496 x1488 padidėjęs1 patikimas1 x642 aktyvus1

smarkus1 x1617 intensyvus1

aukštas1

Figure 6.2.6.1 The lattice of the adjective "*geras*" (part 1)

Lattice for: geras1

nuostabus1

101 x1099 x1087 x1067 x1062 x1000 x997 x989 x844 x839 x837 x836

x133 vidutinis1 x1097 pakilus1 x1618 x1475 x1092 supratingas1 x991 kruopštus1 x1066 x830 viliojantis1

x1100 x1096 x988 neblogas1 x1089 prasmingas1 orus1 nuoširdus1 kokybiškas1
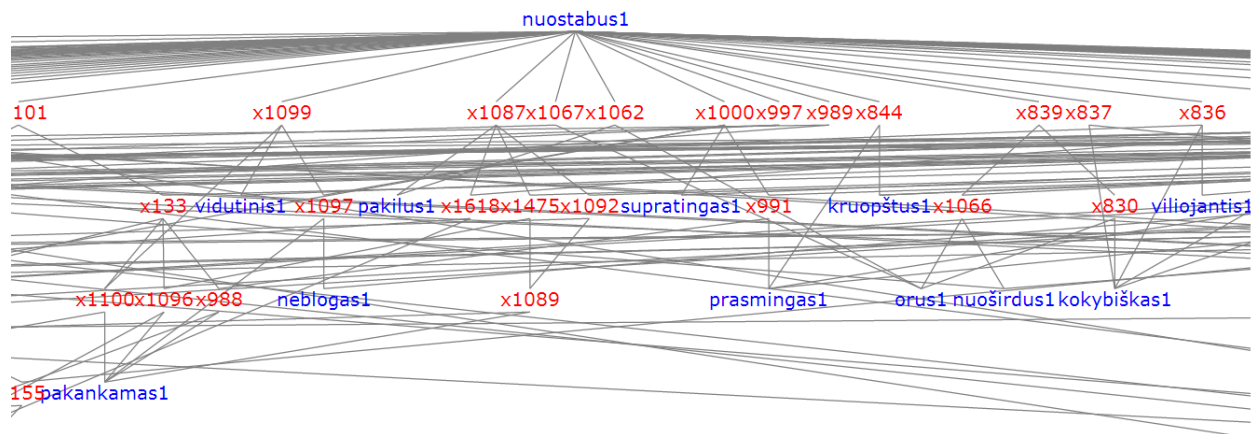
155 pakankamas1

Figure 6.2.6.2 The lattice of the adjective "*geras*" (part 2)

This is one lattice divided into two pictures since it was too detailed to contain all of it into one picture. We see that the adjective "*geras*" can be found in between two senses "*svarbus*" and "*nuostabus*". The adjectives "*svarbus*" and "*nuostabus*" are the upper nodes in the lattice in respect to the adjective "*geras*", that is why they were automatically classified to become the hyperonyms of the adjective "*geras*".

In the dictionary the adjective "*svarbus*" have such senses (DLKZ):

---

svarbus

1. turintis didelę reikšmę (tr. of critical importance)

2. tarm. svarus, sunkus (tr. (dialectal) weighty, heavy)

"*Nuostabus*" (DLKZ):

---

nuostabus

keliantis susižavėjimą, nepaprastas (tr. admired and extraordinary)

There is a link of positive feature between the adjective "*geras*" and "*nuostabus*". But it does not make the adjective "*nuostabus*" the hyperonym of "*geras*", rather it is the synonym. Though, it did not happenend to have any equivalent sense in the GSs. "*Svarbus*" does not seem to have a strong link to these two adjectives in the natural language or in the manually created resources of it (dictionaries).

VII.   Tinkamas

The Semantic Mirrors

(*Translation:* sufficient, reasonable, proper, decent, right, successful, perfect, good, useful.)

Hyperonyms: dailus‹1›, nuostabus‹1›.

Synonyms: atitinkamas, atskiras‹1›, deramas, doras‹1›, garsus‹1›, kokybiškas‹1›, naudingas‹1›, nuoširdus‹1›, orus‹1›, padorus‹1›, pagrįstas‹1›, pakankamas‹1›, protingas‹1›, reikiamas, vienintelis‹1›.

Related words: brangus‹1›, gardus‹1›, gausus‹1›, idealus‹1›, išmintingas‹1›, klestintis‹1›, laimingas‹1›, logiškas‹1›, neblogas, nemažas‹1›, nepriekaištingas‹1›, normalus‹1›, nusipelnęs‹1›, patenkintas‹1›, patogus‹1›,pelningas‹1›, perspektyvus‹1›, pilnas‹2›, populiarus‹1›, pozityvus, prasmingas‹1›, prieinamas‹1›, priimtinas‹1›, racionalus‹1›, realus‹1›, santūrus‹2›, sklandus‹2›, specialus‹1›, supratingas‹1›, sveikas‹1›, sėkmingas, taiklus‹1›,taisyklingas‹1›, talentingas‹1›, teigiamas‹1›, teisingas, tikras, tvarkingas‹1›, vertingas, įtikinamas‹2›.

DLKZ (Modern Lithuanian Dictionary)

---

tinkamas

> 1. atitinkantis reikalavimus, geras (tr. conforming to the requirements, good)
>
> **Correspondences from SM:** atitinkamas, deramas, doras‹1›, kokybiškas‹1›, padorus‹1›, pagrįstas‹1›, pakankamas‹1›, neblogas, sėkmingas, taiklus‹1›, taisyklingas‹1›, teigiamas‹1›, teisingas, tikras
>
> 2. patinkamas (tr. pleasing)
>
> **Correspondences from SM:** priimtinas‹1›

LKZ (the Dictionary of the Lithuanian Language)

> tinkamas
>
> 1. kuris kam tinka, pritaikomas kokiam reikalui, deramas (tr. something what is suitable, applicable to some particular matter, appropriate)
>
> **Correspondences from SM:** atitinkamas, deramas, naudingas‹1›, pagrįstas‹1›, pakankamas‹1›, taiklus‹1›, taisyklingas‹1›, teigiamas‹1›, teisingas, tikras
>
> 2. norimas, prideramas, reikiamas, atitinkamas (tr. desired, proper, necessary, appropriate)
>
> **Correspondences from SM:** kokybiškas‹1›, reikiamas
>
> 3. atitinkantis skonį, patinkamas (tr. matching the taste, pleasing)
>
> **Correspondences from SM:** priimtinas‹1›
>
> 4. teigiamas, padorus, geras (tr. positive, decent, good)
>
> **Correspondences from SM:** doras‹1›, nuoširdus‹1›, padorus‹1›, neblogas, pozityvus, sėkmingas

The Qualitative Evaluation

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 19.

All the rest of the adjectives that were returned by SM and named noise are 38. 38 + 19 = 57.

P = 19 / 57 = **0.33**

The number of the results that should have been returned is 10. And the new correct results are added: 10 + 19 = 29.

R = 19 / 29 = **0.66**

F = 0.442

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:
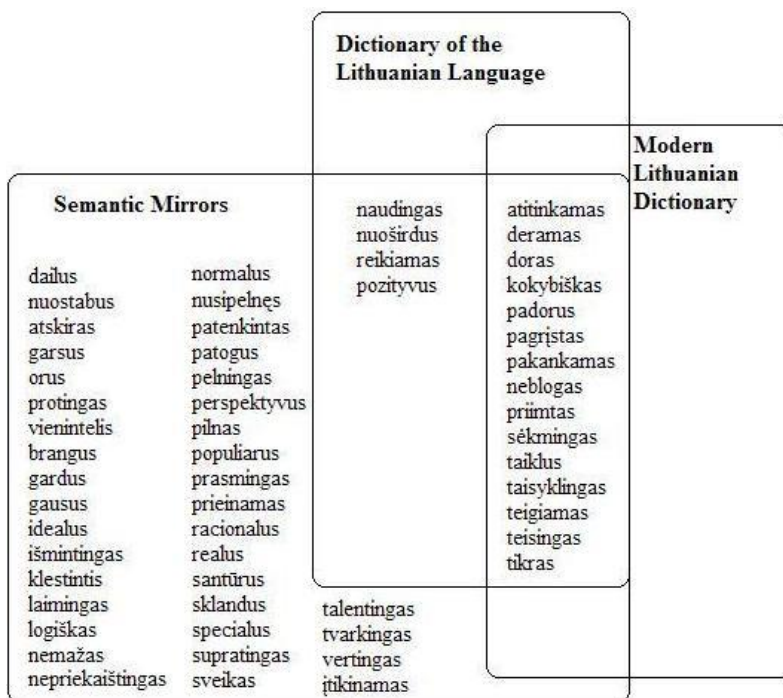

Figure 6.2.7 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

There are adjectives among the ones that are 'noise', which do mean "*tinkamas*" (tr. suitable) in some way, and represent its semantics, even though they were not listed as a separate senses in the dictionaries. Here is a list of adjectives that can be connected with the meaning of "*tinkamas*": "*idealus*", "*patogus*". For example:

Drožinėti medžio skulptūras yra idealus darbas kruopštiems žmonėms. (tr. Whittling sculptures of wood is the perfect/ suitable/ ideal work for thorough people.).

Šie batai yra patogūs. (tr. These shoes are comfortable/ suitable.).

And again the question of golden standard's coverage may be raised.

VIII.   Jaunas

The Semantic Mirrors

<div style="border:1px solid black; padding:8px;">

Subsense (i)

(*Translation:* low, little.)

Synonyms: nepakankamas‹1›, tylus‹2›, žemas‹2›.
Related
words: duslus‹2›, kuklus‹1›, liūdnas‹2›, nereikšmingas‹1›, nesudėtingas‹1›, siauras‹1›, smulkus‹1›.

Subsense (ii)
(*Translation:* early, new.)
Synonyms: eilinis‹1›, naujas, pradinis‹1›.
Related
words: ankstesnis‹2›, gatavas‹1›, kitoks‹1›, modernus‹1›, naujasis‹1›, neregėtas‹2›, netikėtas‹2›, senas‹2›, vėlus‹2›, šiuolaikinis, šviežias‹2›.

</div>

DLKZ (Modern Lithuanian Dictionary)

<div style="border:1px solid black; padding:8px;">

jaunas

nedaug amžiaus turintis; mažiau amžiaus turintis už kitus (tr. not old, younger than others)

No correspondences from SM

</div>

LKZ (the Dictionary of the Lithuanian Language)

<div style="border:1px solid black; padding:8px;">

Jaunas adj.

1. nedaug metų turintis, nesenas (tr. a few years old, the recent)

No correspondences from SM

2. neseniai įkurtas, įsteigtas (tr. recently set up)

Correspondences from SM: pradinis‹1›

3. naujas (tr. new)

Correspondences from SM: naujas, naujasis‹1›, neregėtas‹2›

4. šviežias (tr. fresh)

Correspondences from SM: šviežias‹2›

5. tik ką pasirodęs (apie mėnulio fazę) (tr. just appeared (about a phase of the moon))

</div>

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 5.

All the rest of the adjectives that were returned by SM and named noise are 19. 19 + 5 = 24.

P = 5 / 24 = **0.21**

The number of the results that should have been returned is 3. And the new correct results are added: 3 + 5 = 8.

R = 5 / 8 = **0.63**

F = 0.313

The second subsense was the one which contained senses from the golden standard. There also appeared antonyms. The reason of antonyms appearing among the output of the SM is already mentioned in the case of the adjective "*senas*" (p. 67-72).

For the same reason as we got the antonym in the entry of the adjective "*senas*". Here we can see how the adjective "*jaunas*" is linked with the adjective "*senas*":
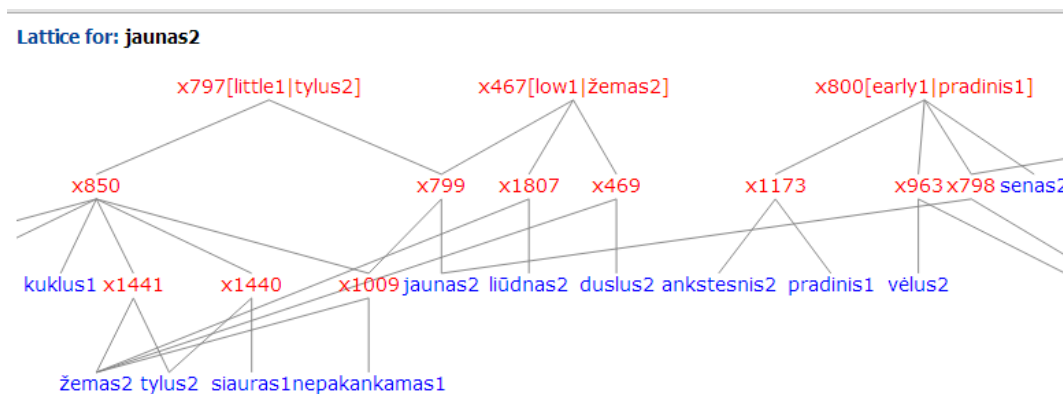

Figure 6.2.8.1 The lattice of the adjective "*jaunas*"

From the lattice above (figure 6.2.8.1) we see that the roots of the antonym in this case are located in the means of the translation of the adjective "*early*" (see "The evaluation of the adjective "*senas*" pages 67-72). The adjective "*jaunas*" can be translated to the adjective "*early*" in English. And "*early*" can mean "*old*". That is why the SM assigned this antonym to the adjective "*jaunas*".

IX.   Gražus

The Semantic Mirrors

(*Translation:* great, nice, good, fine.)

Hyperonyms: tikslus‹2›, stiprus‹1›, nuostabus‹1›, puikus‹1›.

Hyponyms: dailus‹1›, doras‹1›, gležnas‹2›, grakštus,
gražutis, grynas, išauklėtas‹1›, išdidus‹1›, lengvas‹1›, malonus‹1›, mielas‹1›,
nepakartojamas, normalus‹1›, pagirtinas‹2›, painus‹2›, pasigėrėtinas, paslaptingas‹1›,pilnas‹2›, plonas‹1›,
prašmatnus‹1›, priimtinas‹1›, simpatingas, skanus‹1›, smulkus‹1›, stambus, subtilus, sveikas‹1›,
taktiškas‹1›, taurus, tikras, trapus‹1›, vidutinis, viliojantis‹1›, šaunus‹1›, švelnus‹1›, šviesus‹1›, žavus‹1›.

DLKZ (Modern Lithuanian Dictionary)

gražus

1. teikiantis pasigėrėjimo, džiuginantis savo išvaizda ar skambėjimu, dailus; prš. b j a u r u s 1 (tr. providing admiration, pleasing in appearance or sound, beautiful; opposite - ugly)

Correspondences from SM: nuostabus‹1›, puikus‹1›, dailus‹1›, grakštus, gražutis, malonus‹1›,
mielas‹1›, nepakartojamas, simpatingas, pasigėrėtinas, prašmatnus‹1›, viliojantis‹1›, šaunus‹1›, žavus‹1›

2. Giedras (tr. cloudless)

No correspondences from SM

3. riebus, vešlus, derlingas (apie gyvulius ar javus) (tr. thick, lush, fertile (about animals or crops))

Correspondences from SM: stambus

4. didelis, gausus, apstus (tr. big, rich, numerous)

No correspondences from SM

5. geras, vertingas (tr. good, valuable)

No correspondences from SM

6. doras, mandagus (tr. honest, courteous)

Correspondences from SM: doras‹1›, taktiškas‹1›, taurus

LKZ (the Dictionary of the Lithuanian Language)

gražus adj.

1. dailus, darnaus sudėjimo; mielas (tr. beautiful, with harmonious constitution, cute)

Correspondences from SM: dailus‹1›, grakštus, gražutis, pasigėrėtinas, prašmatnus‹1›, viliojantis‹1›,
šaunus‹1›, žavus‹1›

2. giedras (apie orą) (tr. serene (weather))

No correspondences from SM

3. riebus, tarpus, vešlus (tr. thick, luxuriant, lush)

Correspondences from SM: stambus

4. didelis, gausingas, apstus (tr. large, abundant, numerous)

No correspondences from SM

5. geras, tinkamas, vertingas (tr. good, suitable, valuable)

No correspondences from SM

6. doras; teisingas; mandagus(tr. honest, fair, courteous)

Correspondences from SM: doras‹1›, taktiškas‹1›

7. malonus, smagus (tr. enjoyable, funny)

Correspondences from SM: nuostabus‹1›, puikus‹1›, malonus‹1›, mielas‹1›, nepakartojamas, simpatingas

8. tyras, švarus (tr. pure, clean)

Correspondences from SM: taurus

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 18.

All the rest of the adjectives that were returned by SM and named noise are 23. 23 + 18 = 41.

P = 18 / 41 = **0.44**

The number of the results that should have been returned is 21. And the new correct results are added: 21 + 18 = 39.

R = 18 / 39 = **0.46**

F = 0.45

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:
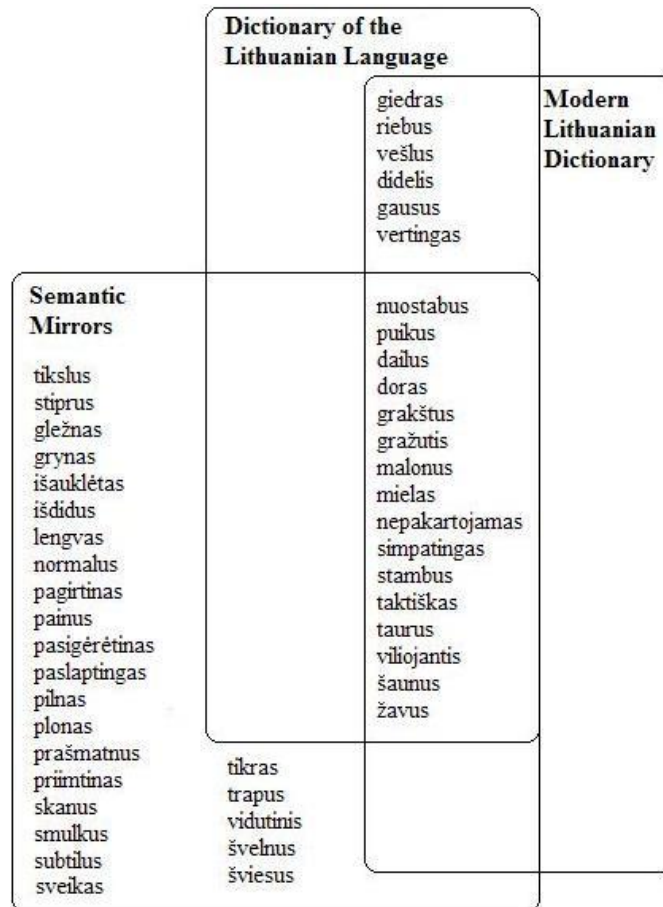
Figure 6.2.9 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

From the figure 6.2.9 we can see that even though a lot of senses fell into some of the sense group, there are still many adjectives that are excluded as noise.

X.   Meilus

The Semantic Mirrors

(*Translation:* affectionate, loving, sweet, nice.)

Hyperonyms: tikslus‹2›.

Synonyms: brangus‹1›, gaivus‹1›, gardus‹1›, grynas, išauklėtas‹1›, meilingas, mielas‹1›, nuoširdus‹1›, pasigėrėtinas, saldus‹2›, simpatingas, švelnus‹1›, šventas‹1›, šviežias‹2›.

Related words: ištikimas‹2›, sėkmingas.

DLKZ (Modern Lithuanian Dictionary)

meilus

1. kuris meilę rodo, lipšnus (tr. who shows love, sweet)

Correspondences from SM: meilingas, nuoširdus‹1›, saldus‹2›, simpatingas, švelnus‹1›

2. mielas, mylimas, brangus (tr. dear, beloved, precious)

Correspondences from SM: brangus‹1›, mielas‹1›, pasigėrėtinas

LKZ (the Dictionary of the Lithuanian Language)

meilus adj.

1. kuris moka meilintis, rodo meilę, švelnumą, lipšnus, glosnus (tr. who knows how to be romantic, demonstrates love, gentleness, is sweet, smooth)

Correspondences from SM: meilingas, nuoširdus‹1›, saldus‹2›, simpatingas, švelnus‹1›

2. malonus, mielas (tr. nice, cute)

Correspondences from SM: mielas‹1›, pasigėrėtinas

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 8.

All the rest of the adjectives that were returned by SM and named noise are 9. 9 + 8 = 17.

P = 8 / 17 = **0.47**

The number of the results that should have been returned is 6. And the new correct results are added: 6 + 8 = 14.

R = 8 / 14 = **0.57**

F = 0.516

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:
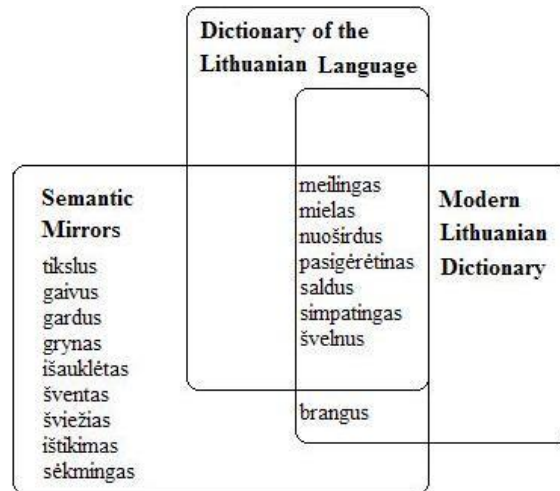
Figure 6.2.10 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

From the figure 6.2.10 we can see that even though a lot of senses fell into some of the sense group, there are still many adjectives that are excluded as noise.

XI.   Kuklus

The Semantic Mirrors

Subsense (i)
(*Translation:* plain.)

Synonyms: akivaizdus‹1›, atviras‹1›, elementarus‹1›, lygus, neįdomus‹1›, tuščias‹2›, varginantis‹2›, žemiškas‹2›.

Subsense (ii)
(*Translation:* light, small, little.)
Synonyms: mažas‹2›, miglotas‹1›, nepakankamas‹1›, nereikšmingas‹1›, nesudėtingas‹1›, purus‹1›, siauras ‹1›, smulkus‹1›, tylus‹2›, žemas‹2›.

Related
words: giedras, jaunas‹2›, lengvas‹1›, miklus‹1›, neryškus‹2›, plonas‹1›, silpnas‹1›, taktiškas‹1›, trumpas‹ 2›, švelnus‹1›.

DLKZ (Modern Lithuanian Dictionary)

---

kuklus

1. neišdidus, neišpuikęs; drovus (tr. not arrogant, unpresuming; shy)

No correspondences from SM

---

2. paprastas, neprabangus (tr. simple, not luxurious)

Correspondences from SM: elementarus‹1›, žemiškas‹2›

LKZ (the Dictionary of the Lithuanian Language)

kuklus adj.

1. neišdidus, neišpuikęs; drovus, padorus (tr. not arrogant, unpresuming; shy, decent)

No correspondences from SM

2. paprastas, be prabangos (tr. simple, luxury-free)

Correspondences from SM: elementarus‹1›, žemiškas‹2›

3. nedidelis, menkas, negausus (tr. small, poor, sparse)

Correspondences from SM: mažas‹2›, nepakankamas‹1›, nereikšmingas‹1›, smulkus‹1›, žemas‹2›, plonas‹1›, silpnas‹1›

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 9.

All the rest of the adjectives that were returned by SM and named noise are 19. 19 + 9 = 28.

P = 9 / 28 = **0.32**

The number of the results that should have been returned is 6. And the new correct results are added: 6 + 9 = 15.

R = 9 / 15 = **0.6**

F = 0.417

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:
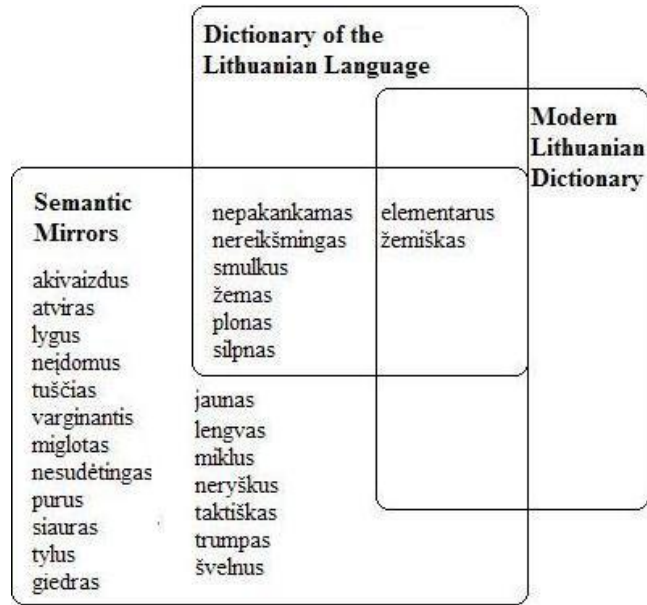
Figure 6.2.11 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

From the figure 6.2.11 we can see that even though a lot of senses fell into some of the sense group, there are still many adjectives that are excluded as noise.

XII. Sklandus

The Semantic Mirrors

Hyperonyms: nuostabus‹1›.

Subsense (i)
(*Translation:* proper, good.)

Synonyms: atskiras‹1›.

*Related
words:* atitinkamas, deramas, nepriekaištingas‹1›, normalus‹1›, padorus‹1›, pakankamas‹1›, pilnas‹2›, specialus‹1›, taisyklingas‹1›, teisingas, tinkamas‹1›, tvarkingas‹1›.

Subsense (ii)
(*Translation:* sound.)

Synonyms: tvirtas‹1›.

*Related
words:* blaivus‹1›, garsus‹1›, logiškas‹1›, nuodugnus, panašus‹1›, patikimas‹1›, perspektyvus‹1›, racionalus‹1›, svarus, sveikas‹1›.

Subsense (iii)

(*Translation:* smooth.)
Synonyms: švelnus‹1›.
Related
words: glotnus‹2›, grakštus, harmoningas‹1›, lengvas‹1›, lygus, nekliudomas‹1›, santūrus‹2›, tolygus‹1›.

DLKZ (Modern Lithuanian Dictionary)

---

sklandus

1. kuris be sutrikimų, darnus (tr. trouble-free, consistent)

Correspondences from SM: taisyklingas‹1›, tvarkingas‹1›, lygus, tolygus‹1›

2. darnus, sutariantis (tr. harmonious, concordant)

Correspondences from SM: harmoningas‹1›

3. kuriuo sklendžiama, slidus (tr. the one on which you can glide, slippery)

Correspondences from SM: glotnus‹2›

LKZ (the Dictionary of the Lithuanian Language)

---

sklandus adj.

1. kuriuo lengvai slystama (tr. easy to slide on) -

Correspondences from SM: glotnus‹2›

2. kuris lengvai išsmunka, išslysta, nuslysta (tr. who easily slips out or away)

No correspondences from SM

3. be kliūčių, lygus (tr. without obstacles, smooth)

Correspondences from SM: lygus, tolygus‹1›

4. rišlus, nuoseklus (tr. coherent, consistent)

Correspondences from SM: logiškas‹1›, nuodugnus, racionalus‹1›

5. be sutrikimų, darnus (tr. trouble-free, harmonious)

Correspondences from SM: taisyklingas‹1›, tvarkingas‹1›, harmoningas‹1›

6. gram.: Sklandíeji priebalsiai (pusbalsiai l ir r) (tr. gram.: Liquid consonants (semivowel l and r)

No correspondences from SM

---

7. gerai planiruojantis, sklendžiantis (tr. gliding well, floating)

No correspondences from SM

8. gerai prigulęs, sandarus (tr. well fitted, tight)

No correspondences from SM

9. kuris lygiai, darniai, lengvai ką daro (tr. someone that performs something smoothly)

No correspondences from SM

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 9.

All the rest of the adjectives that were returned by SM and named noise are 24. 24 + 9 = 33.

P = 9 / 33 = **0.27**

The number of the results that should have been returned is 6. And the new correct results are added: 6 + 9 = 15.

R = 9 / 15 = **0.6**

F = 0.375

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:
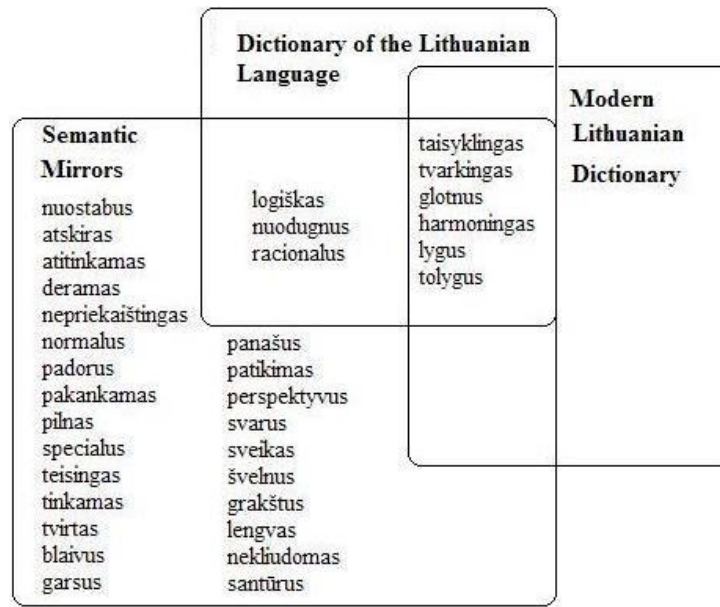
Figure 6.2.12 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

From the figure 6.2.12 we can see that even though a lot of senses fell into some of the sense group, there are still many adjectives that are excluded as noise.

XIII.   Dailus

The Semantic Mirrors

(*Translation:* perfect, bright, beautiful, nice, fine, smart, elaborate, good.)

Hyperonyms: šaunus‹1›, paprastas‹2›, tikslus‹2›, gražus‹1›, geras‹1›, nuostabus‹1›.

Synonyms: malonus‹1›.

Related words: giedras, gražutis, išauklėtas‹1›, išdidus‹1›, kerintis‹1›, kilnus, mielas‹1›, pasigėrėtinas, paslaptingas‹1›, patogus‹1›, saulėtas‹1›, simpatingas, tvirtas‹1›, vaizdingas‹1›, įdomus‹1›, į stabus, žavingas‹1›, žavus‹1›.

Hyponyms: aiškus‹1›, akinantis, charakteringas, grynas, idealus‹1›, laimingas‹1›, lengvas‹1›, logiškas‹1›, lygut lygutėlis‹1›, nepriekaištingas‹1›, panašus‹1›, patrauklus, taisyklingas‹1›, tikras, tinkamas‹1›, tobulas‹1›,tv arkingas‹1›, tyras, visiškas, šventas‹1›.

DLKZ (Modern Lithuanian Dictionary)

---

dailus

> 1. malonus pažiūrėti, labai gražus (tr. nice to look at, very beautiful)
>
> **Correspondences from SM:** gražus‹1›, gražutis, kerintis‹1›, malonus‹1›, mielas‹1›, pasigėrėtinas, simpatingas, žavingas‹1›, žavus‹1›, patrauklus
>
> 2. Lygus (tr. smooth)
>
> **Correspondences from SM:** lygut lygutėlis‹1›

LKZ (the Dictionary of the Lithuanian Language)

> Dailus adj.
>
> 1. gražus; puikus (tr. beautiful, excellent)
>
> **Correspondences from SM:** šaunus‹1›, gražus‹1›, geras‹1›, nuostabus‹1›, gražutis, kerintis‹1›, pasigėrėtinas, simpatingas, žavingas‹1›, žavus‹1›, idealus‹1›, nepriekaištingas‹1›, patrauklus, tobulas‹1›
>
> 2. lygus (tr. smooth)
>
> **Correspondences from SM:** lygut lygutėlis‹1›
>
> 3. atsiganęs, neliesas (tr. fed a lot, not thin)
>
> **No correspondences from SM**
>
> 4. malonus, giedras, nešaltas (tr. pleasant, serene, mild)
>
> **Correspondences from SM:** giedras, malonus‹1›, mielas‹1›, saulėtas‹1›

<u>The calculation of the idealized recall, precision and F-score in this case is:</u>

The amount of correct results (senses that were matched with adjectives from SM) is 19.

All the rest of the adjectives that were returned by SM and named noise are 26. 26 + 19 = 45.

P = 19 / 45 = **0.42**

The number of the results that should have been returned is 8. And the new correct results are added: 8 + 19 = 27.

R = 19 / 27 = **0.7**

F = 0.528

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:

Figure 6.2.13 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

There are few adjectives among the ones that are "noise", which do mean "*dailus*" (tr. nice) in some way, and represent its semantics, even though they were not listed as a separate sense in the dictionaries. Here are the adjectives that can be connected with the meaning of "*dailus*": "*malonus*" (tr. sweet, nice), "*vaizdingas*" (tr. scenic), "*įstabus*" (tr. amazing). For example:

Sodas buvo išpuoselėtas ir malonus pažiūrėti. (tr. The garden was sleek and <u>nice</u>.).

Paryžius yra vaizdingas miestas. (tr. Paris is a <u>nice/ beautiful</u> city.).

Prieš akis atsivėrė įstabus reginys. (tr. In fromt of (us) the amazing/ <u>beautiful</u> view appeared.).

The adjective "*dailus*" in DLKZ is described through two senses; in LKZ - through two more senses (four in total). The first sense in both golden standards is slightly different, because LKZ next to the meaning "*gražus*" (tr. beautiful) it gives the meaning "*puikus*" (tr. excellent). This makes the sense wider than the sense in DLKZ. It is interesting that SM manages to determine the senses in both golden standards even when LKZ contains more specific information. This suggests the idea that SM, in some cases, is capable of capturing more precise semantic information than any one of two golden standards separately.

## XIV.	Tikras

The Semantic Mirrors

(*Translation:* fair, reasonable, perfect, fine, good, substantial.)
Hyperonyms: dailus‹1›, konkretus‹2›, gražus‹1›, nuostabus‹1›.
Synonyms: deramas, doras‹1›,
garbingas‹1›, lygus, neblogas, nemenkas, nesuterštas‹1›, pagrįstas‹1›, pakankamas‹1›, protingas‹1›, realus
‹1›, sąžiningas‹1›, teisingas, vidutinis, vidutiniškas‹1›, švelnus‹1›, šviesus‹1›.
Related
words: atitinkamas, brangus‹1›, išmintingas‹1›, logiškas‹1›, nemažas‹1›, padorus‹1›, prasmingas‹1›, priei
namas‹1›, priimtinas‹1›, racionalus‹1›, santūrus‹2›, supratingas‹1›, tinkamas‹1›, įtikinamas‹2›.

DLKZ (Modern Lithuanian Dictionary)

---

tikras

1. koks yra pagal prigimtį (tr. the way one is by nature)

No correspondences from SM

2. turintis būdingas ypatybes (tr. with characteristic features)

Correspondences from SM: konkretus‹2›

3. Neabejojamas (tr. indubitable)

No correspondences from SM

4. Įsitikinęs (tr. sure)

No correspondences from SM

5. vedantis į tikslą (tr. leading to the goal)

No correspondences from SM

6. grynas, nuoširdus (tr. pure, sincere)

Correspondences from SM: doras‹1›, garbingas‹1›

7. labai panašus (tr. very similar)

No correspondences from SM

LKZ (the Dictionary of the Lithuanian Language)

Tìkras adj.

1.realus, apčiuopiamas, neapgaulingas (tr. real, tangible, not deceptive)

Correspondences from SM: realus‹1›,

autentiškas (apie pinigus, dokumentus) (tr. authentic (about money, documents))

No correspondences from SM

2.natūralus, grynas (ppr. su daiktavardžiais, reiškiančiais medžiagas) (tr. natural, pure (ppr. with nouns meaning materials))

No correspondences from SM

 atitinkantis ką, grynas (su daiktavardžiais, reiškiančiais daiktų rūšis) (tr. corresponding to something, pure (with nouns, meaning the kind of things)

Correspondences from SM: atitinkamas

 grynas, švarus (tr. pure, clean)

No correspondences from SM

grynas, be pagražinimų (sustiprinant žodžių tiesa, teisybė reikšmę) (tr. pure, without embellishment (enhancing the meaning of word "truth"))

No correspondences from SM

3. turintis visus būtinus daikto, reiškinio ar ypatybės požymius; toks, koks turėtų būti (tr. with all necessary phenomenon or characteristic features of some object;exactly as it should be)

Correspondences from SM: konkretus‹2›

4. savas, pačiam priklausantis, nuosavas (tr. own, of itself, its own)

 prigimtas kam, susijęs kraujo giminyste (tr. innate, related by blood kinship)

No correspondences from SM

5. teisingas, atitinkantis tiesą, nemelagingas (tr. fair, corresponding to the truth, not spurious)

Correspondences from SM: doras‹1›, garbingas‹1›, sąžiningas‹1›, teisingas

teisingas, deramas, toks, kokio reikėtų (tr. fair, reasonable, exactly as it should be)

Correspondences from SM: deramas

 tinkamas kam, geras, patogus (tr. suitable for something, comfortable)

Correspondences from SM: tinkamas‹1›

taisyklingas, teisingas (tr. accurate, correct)

No correspondences from SM

6. neabejotinas, patikimas; akivaizdus (tr. undoubted, reliable, clear)

No correspondences from SM

7. ppr. emph. Faktiškas (tr. factual)

No correspondences from SM

8. teisingas, doras, sąžiningas, pasitikimas(tr. fair, honest, conscientious, trustable)

Correspondences from SM: doras‹1›, deramas

nuoširdus, neapsimestinis (tr. sincere, genuine)

ištikimas(tr. loyal)

No correspondences from SM

9. tinkamas, tinkantis (tr. appropriate, suitable)

Correspondences from SM: priimtinas‹1›, tinkamas‹1›

 Atitinkamas (tr. corresponding)

Correspondences from SM: atitinkamas

10. įsitikinęs (tr. certain)

No correspondences from SM

11. dešinysis (tr. the right one)

No correspondences from SM

12. nuoširdus (tr. frank)

No correspondences from SM

The calculation of the idealized recall, precision and F-score in this case is:

The amount of correct results (senses that were matched with adjectives from SM) is 10.

All the rest of the adjectives that were returned by SM and named noise are 25. 25 + 10 = 35.

P = 10 / 35 = **0.29**

The number of the results that should have been returned is 30. And the new correct results are added: 30 + 10 = 27.

R = 10 / 30 = **0.33**

F = 0.307

Further down, we can see the overlapping intersections of the adjectives that appear to be some of the senses distinguished by SM, LKZ and DLKZ:
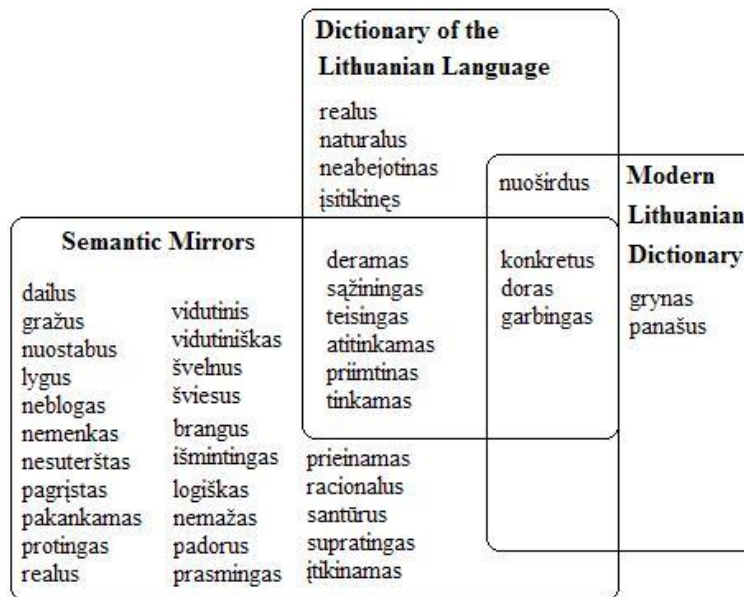
Figure 6.2.14 *Example of overlapping intersections of the adjectives distinguished by SM, LKZ and DLKZ (The translations might be seen in the Glossary (see Appendix))*

There is one adjective among the ones that are listed as "noise" in SM, which do mean "*tikras*" (tr. real) in some way, and represent its semantics, even though it was not listed as a separate sense in the dictionaries. Here is the adjective that can be connected with the meaning of "*tikras*": "*realus*" (tr. real). For example:

Mums grėsė realus pavojus. (tr. We were in real/ factual danger.).

And again the question of golden standard's coverage may be raised.

7 General Conclusion

Many adjectives from my random selection list seem to appear in each others' results generated by SM and/or given in the dictionaries. This is a little unfortunate, because I was not able to investigate a proper random selection of results. The reason why the adjectives in my random selection list happened to be semantically closely related is based on the method that I have used for the random selection: because I selected adjectives from one book, one chapter and I selected them one after another while I was reading, due to the tie in the context they happened to be quite related with one another. However, this enables us to study how a certain semantic field is treated by SM in more depth.

Recall, precision and F-score

In the chart below, the idealized recall scores are presented:
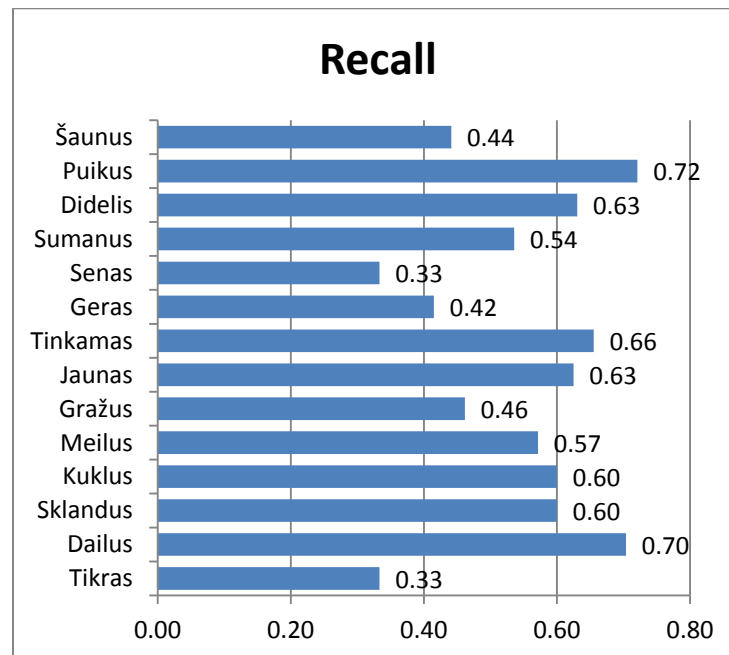
Figure 7.1 *The recall of the cases*

From the chart above we see which adjectives received highest or lowest results of the recall. There are two adjectives which are clasified as representing the 'high' recall ("*puikus*" - 0.72 and "*dailus*" - 0.7). All the other of the results fell into the range of the 'average' recall score. None of the adjectives were clasified to be of a 'low' recall.
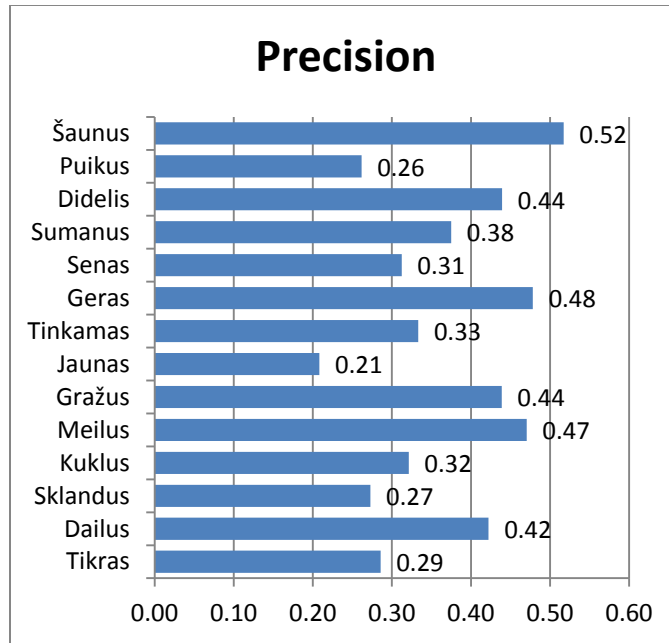
Figure 7.2 *The precision of the cases*

According to the precision, there were 8 adjectives that are included into the range of the 'average' scores; 6 other adjectives are included into the range of the 'low' scores ('bad' performance). The precision is quite low in many cases because of the high number of noise in some cases of adjectives. The F-score variable will give us more general impression of the automatic generation of the thesaurus entries using the Semantic Mirrors method.
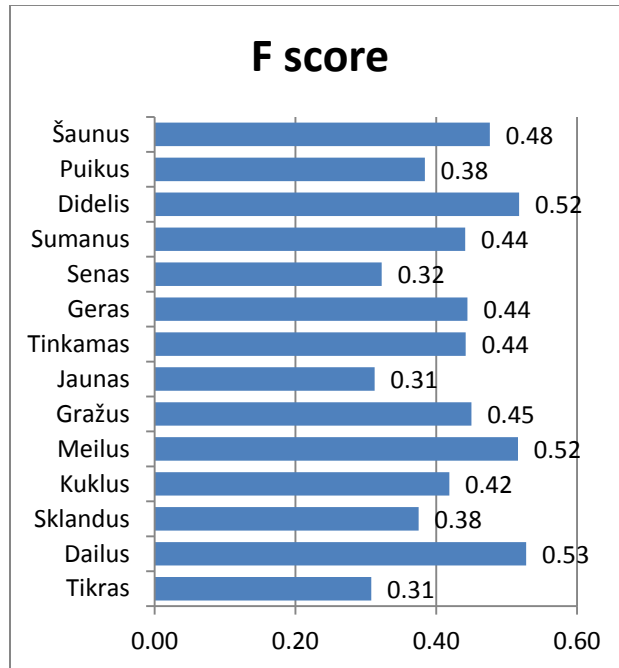
Fig 7.3 *The F-score of the cases*

According to the F-score, there are 11 cases with 'averag' performance, 3 with 'bad' and 0 with 'good' performance of automatic derivation of thesaurus entries. Averagely, the adjectives get the 0.42 score. Most of the cases of the F-score are 'average'. Few reasons of why this happened can be speculated:

a) High noise

Many cases ("*puikus*", "*jaunas*", "*senas*", "*kuklus*", "*sklandus*") suffered from the big amount of noise. However, in some cases this can be modified using the variables SynsetLimit and OverlapThreshold. Since in this research we were using the automatic variables, later on the modification of SL and OT was tested. In two cases, it was possible to reduce the noise: in the case of the adjective "*tinkamas*" the automatic settings of SL (20) and OT (0.05) returned 57 adjectives, out of which 5 was accurately matching the synonyms in the GS. When SL was decreased to 15 and the OT to 0.01 the number of returned results became 43 and within it all the 5 accurate matches stayed. In the case of "*sklandus*" the best result was reached when the SL was decreased to 10 and the OT to 0.01. Then, from the previous 34 words on the output, only 12 remained, and they still comprised the correct result (it was only one in this case). However, the manipulation of SL and the OT did not show a significant improvement on the other cases.

b) Limitations of gold standards

There were cases ("*šaunus*", "*puikus*", "*didelis*", "*tinkamas*", "*dailus*", "*tikras*") where the coverage of the GS was noticeably limited compared to the results from the SM. Occasionally, the SM showed being capable of sorting out the different senses listed in different dictionaries. Also it is capable of suggesting senses which do not appear in any of the dictionary but might as well be concluded to be an unclassified sense rather than noise (in the cases of the adjective "*šaunus*", "*puikus*").

The SM performance tested on Lithuanian adjectives

In many cases the most common problem in the thesaurus entries for Lithuanian adjectives generated by SM was the big amount of noise. In some cases, we concluded that the noise might have been further reduced given more complete gold standards, as the adjectives excluded as noise (based on the data in the dictionaries) were still intuitively falling into the semantic field of the target word. Nevertheless, none of the two dictionaries seemed to grasp the full coverage of the semantic information about the adjectives in my investigation. The Semantic showed ability to give more relevant semantic information than either of the dictionaries (or even both of them merged together) used as a gold standard. For example, in the case of the adjective "*šaunus*", the SM, in addition to recognizing some senses in the golden standards, contained one further sense which could plausibly to be included in the semantic information about the adjective "*šaunus*" (the sense of *mentall brightness*).

The semantic information contained in the translational correspondences

Professional translators are people with extensive bilingual knowledge. The choices that they make while choosing one or another correspondence in the target language reflect the knowledge of the latest language usage trends, linguistic competence, and sensibility of semantic ranges in the two languages. These abilities provide the great amount of the semantic information to be extracted.

The Thesaurus of Lithuanian Adjectives

The thesaurus that was generated during this research is an explorative version. Still, based on the research results, we can see that it contains a great amount of valid semantic relations among adjectives. The thesaurus automatically generated in this thesis could be the basis for developing a quality checked thesaurus that could be used by translators as a tool providing deep semantic information about adjectives. The thesis itself could be usefull in paving the way for such a development. Combined with manual human interaction or further automatic improvements, this explorative study might lay the foundation for a reliable thesaurus of Lithuanian adjectives. *The translation-based, Automatically Derived Thesaurus of Lithuanian Adjectives* is available on-line (one has to select the word base '*agne*': http://maximos.aksis.uib.no/mirrors/wn-entry.xml?listing=true.

Bibliography

Apidianaki, M. (2008). Acquisition automatique de sens pour la désambiguïsation et la sélection lexicale en traduction. Doctoral dissertation, Université Paris Diderot.

Apidianaki, Marianna and Sagot (2012), Benoit. Applying cross-lingual WSD to wordnet development. LREC - Eighth International Conference on Language Resources and Evolution, Istanbul: Turque.

Armalytė, O., Pažūsis L., (1990). Vertimo teorijos pagrindai (Basics of Translation Theory), Vilnius: Vilniaus universiteto leidykla, p.249.

Centre of the Computational Linguistics; Vytautas Magnus University : http://donelaitis.vdu.lt : http://tekstynas.vdu.lt

Centre of the Computational Linguistics; Vytautas Magnus University; Parallel Corpus : http://tekstynas.vdu.lt/page.xhtml?id=parallelCorpus

Centre of the Computational Linguistics; Vytautas Magnus University; About : http://tekstynas.vdu.lt/tekstynas/menu?page=about

Centre of the Computational Linguistics; Vytautas Magnus University; Statitics : http://tekstynas.vdu.lt/tekstynas/menu?page=statistics

CorAlit: the Corpus of Accademic Lithuanian : http://coralit.lt/en/node/18

Corpuscle http://iness.uib.no/korpuskel/main-page

Crouch, C.J. (1988). Construction of a Dynamic Thesaurus and its Use for Associated information retrieval. In proceedings of the 11th international conference on Research and Development in Information Retrieval. France, Grenoble: p. 309-320.

Crouch, C.J. (1990). An approach to the automatic construction of global thesauri. Information Processing & Management; Volume 26, Issue 5, p. 629–640.

Curran, J. R. and M. Moens (2002). Improvements in Automatic Thesaurus Extraction. Institute for Communicating and Collaborative Systems, University of Edinburgh. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.21.6422&rep=rep1&type=pdf (On-line)

Dictionary of the Lithuanian Language : http://www.lkz.lt/en/dze.htm

Dyvik H., and P. Meurer. A Demonstrator for Automatic Thesaurus Derivation; User's Guide : http://maximos.aksis.uib.no/~helge/mirrwebguide.html

Dyvik, H. (1998a). A translational basis for semantics. In Stig Johansson and Signe Oksefjell (eds.): Corpora and Crosslinguistic Research: Theory, Method, and Case Studies. Amsterdam: Rodopi, p. 51-86.

Dyvik, H. (1998b). Translations as semantic mirrors. In Proceedings of Workshop W13: Multilinguality in the lexicon II, pp. 24.44, Brighton, UK. The 13th Biennial European Conference on Artificial Intelligence ECAI.

Dyvik, H. (2002). Translations as semantic mirrors: from parallel corpus to wordnet. Language and Computers. Rodopi: vol. 49, iss. 1, (1 April 2004), p. 311-326 (16).

Dyvik, H. (2005). Translations as a semantic knowledge source. In Proceedings of the Second Baltic Conference on Human Language Technologies, Institute of Cybernetics, Tallinn University of Technology, and Institute of the Estonian Language, Tallinn, p. 27-38.

Dyvik, H. (2008). Semantic Mirrors. Unpublished paper summarizing the method.

Euro Wordnet http://www.illc.uva.nl/EuroWordNet/

Fellbaum, C. (1998). WordNet: An electronic lexical database. MIT Press, Cambridge, MA.

Freksa, C., M. Jantzen, M. Rüdiger Valk (1997). Foundations of Computer Science Potential-Theory-Cognition; Lecture Notes in Computer Science, Vol. 1337; 524 p.

Grefenstette, G. (1993). Automatic thesaurus generation from raw text using knowledge-poor techniques. Rank Xerox Research Centre, Grenoble Laboratory, 6 chemin de maupertuis, 38240 Meylan, France.

Gudavičius, A. (2007). Gretinamoji semantika (Contrastive Semantics). Šiauliai: Šiaulių universiteto leidykla, p. 244.

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In proceedings of the 14th international conference on Computational Linguistics, Nantes, France: p. 539-545.

Hindle, D. (1990). Noun classification from predicate-argument structures; In Proceedings of ACL-90; Pennsylvania, Pittsburg: (June) p. 268-275.

Jakaitienė, E. (1988). Leksinė semantika (Lexical semantics). Vilnius: Mokslas, p. 174.

Lithuanian Dictionary of Synonyms : http://www.lietuviuzodynas.lt/sinonimai/Geras

Lithuanian Scientific Corpus CorALit : http://coralit.lt

Lithuanian WordNet : http://metashare.dfki.de/repository/browse/lithuanian-wordnet

Lyse, G. I. (2010). A Translation-based 'Wordnet" for Word Sense Disambiguation. Masters thesis, University of Bergen.

Mandala, R., T. Tokunaga, and H. Tanaka (1999). Complementing WordNet with Roget's and Corpus-based Thesauri for Information Retrieval. Department of Computer Science; Tokyo Institute of Technology. Japan, Tokyo, 152-8522.

Miller, G. A., R. Beckwith, C. Fellbaum, D. Gross, and K. J. Miller. (1993). Introduction to WordNet: An On-line Lexical Database.

Milne, D., O. Medelyan, and I. H.Witten (2006). Mining Domain-Specific Thesauri from Wikipedia: A case study. Department of Computer Science. University of Waikato.

Mirrors Web: Thesaurus Listing : http://maximos.aksis.uib.no/mirrors/wn-entry.xml?listing=true

Modern Lithuanian Dictionary : http://dz.lki.lt

Navakauskienė, J., and L. Barštytė. (2005). "Keletas hierarchijos „būti" raiškos aspektų prancūzų ir lietuvių kalbose" (Several Aspects of Expressing BE Hierarchy in French and Lithuanian). Kalbotyra, Nr. 55 (3).

Park, Y. C. and K.-S. Choi (1996). Automatic Thesaurus Constructions Using Bayesian Networks. Computer Science Department, Korea Advanced Institute of Science and Technology. Taejon, Korea.

Pereira, F., N. Tishby, and L. Lee (1993). Distributional clustering of English words. In proceedings of the 31st annual meeting of the Association for Computational Linguistics. Columbus, Ohio, USA.

Priss, U., and J. Old (2005). Conceptual Exploration of Semantic Mirrors. In Ganter, Godin (eds.) Formal Concept Analysis: Third International Conference, ICFCA 2005, Springer Verlag.

Rimkutė, E., J. Kovalevskaitė, V. Daudaravičius. (2006). Daugiakalbių tekstynų naudojimas ir taikymas (The Usage and Application of Multilingual Corpora); ISSN 1392-0588 Darbai ir dienos, 45.

Ruge, G. (1997). Automatic Detection of Thesaurus Relations for Information Retrieval Applications. In Foundations of Computer Science: potential – theory – Cognition, Lecture Notes in Computer Science, Germany, Berlin: Springer Verlag, volume LNCS 1337, p. 499-506.

Ruzaitė, J. (2010). Translation Equivalents of Vague Language Items: A Study of General Extenders in a Parallel Corpus; Studies About Languages (Kalbų Studijos); ISSN 1648-2824: 16, p. 33-38.

Stumme, G. (1997). Concept Exploration - A Tool for Creating and Exploring Conceptual Hierarchies; In proceedings of the 5th international conference on conceptual structures.

Tajarobi, A. (1998). La reconnaisance automatique des hyponymes, Master's thesis, Département de langues et linguistique, Université Laval.

Text Encoding Initiative : http://www.tei-c.org

Thesaurus on Wikipedia : http://en.wikipedia.org/wiki/Thesaurus

Thunes, M. (2003). Evaluating thesaurus entries derived from translational features. Paper presented at Reykjavík: Nodalida.

Uni Computing; About : http://www.computing.uni.no/?session-id=234905593953605

Appendix

Glossary

Absoliutus - Absolute

Aitrus - Bitter

Akivaizdus - Clear, Obvious

Aktyvus - Active,

Ankstesnis - Previous

Ankstyvas - Early

Apčiuopiamas - Tangible

Apdairus - Cautious

Apgalvotas - Considered

Apgriuvęs - Dilapidated, Crazy, Tumbledown, Decrepit

Apstus - Numerous

Apsukrus - Shifty

Aršus - Savage

Aštrus - Sharp

Atidus - Attentive

Atitinkamas - Corresponding

Atkaklus - Persistent

Atsakomas - Answerable

Atsiganęs - Fed A Lot

Atskiras - Separate

Atšiaurus - Stark

Atžarus - Offensive

Augalotas - Stalwart

Auksinis - Golden

Aukštas - Tall

Autentiškas - Authentic

Baisus - Scary

Baltas - White

Besaikis - Excessive, Inordinate

Blizgantis - Glossy

Blykčiojantis - Blinking

Brandus - Mature

Brangus - Precious, Expensive

Būtinas - Necessary

Čaižus - Switching

Dailus - Beautiful, Pretty, Handsome, Nice, Elegant

Darbštus - Industrious, Laborious, Diligent, Hard-Working

Darnus - Consistent, Harmonious

Dažnas - Frequent

Deramas - Appropriate, Reasonable

Derlingas - Fertile

Dešinysis - The Right One

Didelis - Large, Big, Great, Considerable

Didingas - Magnificent, Majestic

Didis - Great

Dieviškas - Divine

Doras - Honest

Dosnus - Generous

Dramatiškas - Dramatic

Drūtas - Thick

Duslus - Muted

Džiugus - Merry, Exuberant

Egzotiškas - Exotic

Ekstravagantiškas - Extravagant

Elegantiškas - Elegant

Energingas - Vigorous

Erdvus - Spacious

Faktiškas - Factual

Fantastiškas - Fantastic

Fenomenalus - Phenomenal

Gabus - Gifted

Gailestingas - Compassionate, Gracious

Galingas - Powerful

Galintis - Able, Capable

Galvotas - Brainy

Garbingas - Honorable, Respectable

Gardus - Palatable

Garsus - Famous

Gausingas - Abundant

Gausus – Abundant, Plentiful, Rich

Genialus - Genial

Geras - Good, Nice, Kindly, Gentle

Gerokas - Sizeable

Giedras - Cloudless, Serene

Giluminis - Abyssal, Deep

Gilus - Deep

Gyvas - Alive

Glaudus - Close

Gležnas - Delicate

Globalinis - Global

Globalus - Global

Glosnus - Smooth

Grakštus - Graceful

Gražus - Beautiful, Lovely, Pretty, Good

Greitas - Fast

Gremėzdiškas - Cumbersome, Unwieldy

Grėsmingas - Formidable, Sinister

Griežtas - Strict

Griozdiškas - Clumsy, Unwieldy, Cumbersome, Bulky

Gudrus - Clever, Sly

Guvus - Agile, Spry

Idealus - Ideal

Įdėmus - Intent, Staring

Įdomus - Interesting,

Įjunkęs - Practiced

Įkritęs - Hollow, Cavernous

Ilgas - Long

Intensyvus - Intensive

Intriguojantis - Gripping

Įsidėmėtinas - Notable, Observable

Įsitikinęs - Certain, Sure

Įspūdingas - Impressive

Įstabus - Great

Išauklėtas - Polite

Išdidus - Proud

Išimtinis - Exceptional

Iškalbingas - Eloquent

Iškilmingas - Solemn

Iškilus - Prominent

Išlaikytas - Maintained

Išmintingas - Wise

Išmoningas - Ingenious

Išplėstas - Extensive

Išpuikęs - Haughty

Išpūstas - Exaggerated, Inflated

Išradingas - Inventive

Išraiškingas - Expressive

Išsamus - Comprehensive

Išsilavinęs - Educated

Išsipūtęs - Swollen, Bloated

Išskaidytas - Resoluble

Ištvermingas - Persevering

Išvaizdus - Presentable

Įtaigus - Forcible, Suggestible

Įtakingas - Influential

Įtikimas - Cogent

Įtikinamas - Compelling

Įvairus - Various

Įžymus - Famous

Įžvalgus - Shrewd

Ypatingas - Special

Jaudinantis - Moving, Exciting

Jaukus - Cozy

Jaunas - Young, Green, Juvenile, Youthful

Judrus - Agile, Mobile

Kapitalinis - Capital

Kategoriškas - Peremptory

Kilnus - Noble

Kimus - Hoarse

Klestintis - Prosperous

Klusnus - Obedient, Humble

Kokybiškas - Qualitative

Kruopštus - Thorough

Kuklus - Modest, Humble, Conservative, Quiet

Labas - Good

Laimingas - Happy

Lemiamas - Critical, Decisive

Lemtas - Fateful

Lemtingas - Fatal

Lengvas - Light, Easy

Linksmas - Happy

Lipšnus - Sweet

Liūdnas - Sad

Lygus - Smooth

Logiškas - Logical

Madingas - Trendy

Magiškas - Magic

Maksimalus - Maximal

Malonus - Enjoyable, Kind, Pleasant

Mandagus - Courteous

Masyvus - Massive

Matomas - Apparent, Visible

Mažas - Small

Meilus - Affectionate, Loving, Lovely, Kind

Mielas - Cute

Miklus - Deft

Milžiniškas - Huge, Enormous

Mirguliuojantis - Iridescent

Mįslingas - Mysterious, Puzzling

Mylimas - Beloved

Narsus - Brave

Naturalus - Natural

Naudingas - Beneficial, Useful

Naujas - New

Neabejojamas - Indubitable

Neabejotinas - Undoubted

Neapgaulingas - Not Deceptive

Neaprėpiamas - Immense, Spanless

Neapsimestinis - Genuine

Neatidėliotinas - Urgent

Neatskiriamas - Indistinguishable

Nebartinas - Not Blamable

Nebegaliojantis - Expired

Neblogas - Not Bad

Neeilinis - Uncommon

Neįkainojamas - Invaluable

Neįsivaizduojamas - Inconceivable, Unthinkable

Neišdidus - Not Arrogant

Neišpuikęs - Unpresuming

Neįtikėtinas - Unbelievable

Neįveikiamas - Compulsive

Nekliudomas - Udisturbed

Neliesas - Not Thin

Nemalonus - Unpleasant

Nemažas - Not Small, Significant

Nenormalus - Abnormal

Nenumaldomas - Inexorable

Nenusakomas - Nondescript

Nepakankamas - Insufficient

Nepaprastas - Extraordinary

Nepeiktinas - Irreaproachable

Neperskiriamas - Inseparable

Nepiktas - Not Angry

Nepriekaištingas - Perfect

Neprilygstamas - Unequalled

Nereikšmingas - Insignificant

Neryškus - Dim

Nesavas - not (my) own

Nesudėtingas - Simple

Nešaltas - Mild

Netikėtas - Unexpected

Neturtingas - Poor, Poverty-Stricken, Indigent, Penniless

Norimas - Desired

Normalus - Normal

Nuodugnus - Thorough

Nuoseklus - Consistent

Nuostabus - Admirable, Amazing

Nuoširdus - Frank, Sincere

Nuovokus - Perceptive, Sensible

Nusipelnęs - Worthy, Deserving

Nuskurdęs - Poverty-Stricken

Nutviekstas - Shot

Opus - Burning, Sore

Orus - Dignified

Padarytas - Made

Padidėjęs - Enlarged

Padorus - Decent

Pagarsėjęs - Notorious

Pagirtinas - Commendable

Pagrindinis - Basic, Major

Pagrįstas - Valid, Legitimate

Painus - Intricate, Confusing

Pajėgus - Capable, Able

Pakankamas - Sufficient

Pakilus - Elevated

Palaikis - Aged

Pamatinis - Fundamental

Panašus - Similar

Paplitęs -Common, Prevalent

Paprastas -Simple

Parankus - Handy

Pasenęs - Outdated

Pasibaisėtinas - Terrible, Appalling

Pasigėrėtinas - Admirable

Pasipūtęs - Arrogant

Pasirengęs - Ready

Pasitikimas - Trustable

Paslaptingas - Mysterious

Pastebimas - Noticeable, Significant

Pašėlęs - Furious

Patenkintas - Pleased

Patikimas- Reliable

Patinkamas - Pleasing

Patyręs - Experienced

Patobulintas - Improved

Patogus - Comfortable, Convenient

Patrauklus - Attractive

Pats - Same

Pavartotas - Used

Pavojingas - Dangerous

Pelningas - Profitable

Perspektyvus - Promising

Piktas - Angry

Plonas - Thin

Populiarus - Popular

Pozityvus - Positive

Prabangus - Luxurious

Pradinis - Initial, Original

Praeinantis - Finishing

Pilnas - Full

Prasmingas - Meaningful

Prašmatnus - Luxurious

Prideramas - Proper

Prieinamas - Available, Accessible

Priimtinas - Acceptable

Protingas - Clever

Puikus - Excellent, Superb, Splendid, Beautiful, Great,Amazing

Puošnus - Gorgeous, Ornate

Putnus - Plump

Racionalus - Rational

Radikalus - Radical, Drastic

Rafinuotas - Sophisticated

Realus - Real

Reikiamas - Necessary

Reikšmingas - Significant

Reikšminis - Meaningful

Retas - Rare

Riebus - Thick

Rinktinis - Exquisite, Select

Rišlus - Coherent

Ryškus - Bright

Ryžtingas - Determined

Rūstus - Severe

Sąmojingas - Witty, Ingenious

Sąmoningas - Conscious

Sandarus - Tight

Saulėtas - Sunny

Sausas - Dry

Savas - Own

Savotiškas - Oddish, Particular

Sąžiningas - Conscientious

Sėkmingas - Successful

Senas - Old, Aged, Used, Elderly

Senoviškas - Antediluvian

Sensacingas - Sensational

Siaubingas - Horrible

Simpatingas - Likable

Skaistus - Bright, Virgin

Skambus - Sonorous, Resonant

Skanus - Tasty

Skardus - Loud, Resounding

Skaudus - Painful

Skiriamasis - Distinctive

Sklandus - Fluent, Smooth, Round

Sklendžiantis - Floating

Skubus -Immediate, Urgent

Slaptas - Latent, Secret

Slidus - Slippery

Smagus - Funny

Smailus - Sharp

Smarkus - Intense, Vigorous

Smulkmeniškas - Meticulous

Smulkus - Small, Detailed

Sodrus - Lush

Solidus - Sedate, Solid

Sotus - Full

Spartus - Quick

Spindintis - Shining

Spindulingas - Radiant

Spindulingasis - Radiant

Stačiokiškas - Brusque

Stambus - Large

Stangrus - Resilient

Status - Vertical, Steep

Stebėtinas - Surprising

Stebinantis - Surprising, Starling

Stebuklingas - Wonderful, Miraculous

Stiprus - Strong

Storas - Fat

Stulbinamas - Striking

Stuomeningas - Handsome

Suaugęs - Grown-Up

Subrendęs - Mature

Subtilus - Subtle, Delicate

Sudėtingas - Complicated

Sumanytas - Devisable

Sumanus - Clever, Great, Intelligent, Bright,
Ingenious, Smart

Sunkus - Heavy

Supratingas - Understanding

Susikaupęs - Concentrated

Sustingęs - Stagnant, Numb

Sutariantis - Concordant

Suveltas - Tousled

Svaiginantis - Heady

Svarbiausias - Fundamental

Svarbus - Important

Svarus - Weighty

Sveikas - Healthy

Šaižus -Jarring

Šaunus - Cool, Dear, Dashing, Valiant, Decent,
Nice

Šiuolaikiškas - Modern

Šiurpus - Terrible, Horrific

Šlovingas - Glorious

Švarus - Clean, Pure

Šviečiantis - Shining

Šviesus - Lucid

Šviežias - Fresh

Švytintis - Luminous

Taiklus - Accurate, Pointed

Taisyklingas - Accurate

Taktiškas - Discrete, Considerate

Tamsus - Dark

Tankus - Dense

Tariamas - Supposed

Tarptautinis - International

Tarpus - Luxuriant

Taurus - Sublime

Teigiamas - Positive

Teisingas - Correct, Fair

Tikras - Real, Confident, Genuine, Sure, Certain,
Positive

Tikslus - Accurate

Tinkamas - Suitable, Appropriate, Happy,
Relevant

Tirštas - Dense

Tylus - Quiet,

Tyras - Pure

Tobulas - Perfect

Triuškinantis - Overwhelming

Trumpas - Short

Turtingas - Rich

Tuščias - Empty

Tvarkingas - Neat

Tviskantis - Fulgent

Unikalus - Unique

Užkimęs - Hoarse

Vaiskus - Transparent

Valyvas - Tidy

Valstybinis - National

Veiksmingas - Efficient

Vėlus - Late

Vertingas - Valuable

Vertingas - Valuable

Vešlus - Lush

Vibruojantis - Shaky, Vibrant

Vidutinis - Average

Visiškas - Complete

Vienintelis - Only, Unique

Viliojantis - Tempting

Vykęs - Successful, Felicitous

Vyraujantis - Prevailing, Dominant

Vyriausias - Supreme, Eldest

Žavus - Fascinating

Žemas - Low, Short

Žėrintis - Sparkling

Žėruojantis - Flaming, Glittering

Žibantis - Sparkling

Žilas - Gray, Grizzled, Hoary

Žinomas - Known

Žymus - Considerable

Žmoniškas - Humane

Žvilgantis - Brilliant, Glossy

Žvitrus - Sprightly