

Applying Word Embeddings for Lithuanian Morphology: The Case of Adjectival Participles

Laima JANCAITĖ-SKARBALĖ, Erika RIMKUTĖ,
Justina MANDRAVICKAITĖ

Vytautas Magnus University, Kaunas, Lithuania

`laima.jancaite-skarbale@vdu.lt, erika.rimkute@vdu.lt,
justina.mandravickaite@vdu.lt`

ORCID 0009-0005-9210-672X, ORCID 0000-0003-0858-8593, ORCID 0000-0001-9426-6165

Abstract. This paper presents how word embeddings were used to identify adjectival Lithuanian participles. Although traditionally considered to be a form of a verb, participles in the Lithuanian language also have the characteristics of adjectives. The paper describes a study on how one of the criteria for the identification of adjectival participles was applied using the *fastText* word embedding model. This criterion involves the recognition of adjectives and pronouns that are semantically similar to participles (e.g., these adjectives and pronouns can be synonyms or antonyms of participles). The paper assesses the extent to which word embeddings can help to identify adjectival Lithuanian participles and summarises the advantages and disadvantages of this method. Out of 289 analysed participles, 48 participles (16.61%) that are semantically similar to adjectives and pronouns were identified using word embeddings.

Keywords: Lithuanian language, word embeddings, fastText, grammar, adjectival participles

1. Introduction

Lithuanian is a morphologically complex, inflected language (Ambrasas, 2006, p. 89). The most complex and distinctive part of speech are verbs, as some forms are declined (participles, e.g., *einantis žmogus* “a person **walking**”, *keptas pyragas* “a **baked** pie”), other forms are conjugated (moods and tenses, e.g., *žmogus eina* “a man **is walking**”, *mama keptų pyragą* “mother **would bake** a pie”), some forms are minimally inflected (half participles and gerunds, e.g., *eidamas į universitetą nukritau* “I fell **while going** to the university”), and others are not inflected at all (infinitives, verbal adverbs, e.g., *eiti* “to go”, *kepti* “to bake”; *bėgte bėgo į universitetą* “ran **in a hurry** to the university”).

Participles have many similarities to adjectives: they have categories of gender, number, and cases, can have degrees, a neuter gender, and often can have an attributive function. On the other hand, they retain the most important categories of verbs: transitivity,

aspect, and reflexivity, and they have similar valency to verbs. Thus, participles have the properties of both adjectives and verbs.

However, some participles are similar to adjectives not only in syntactic functions and grammatical categories but also semantically. In this paper, these participles are referred to as adjectival participles. They may even lose verbal properties - they may change their lexical meaning from the verb and lose the grammatical categories of participles that are not characteristic of adjectives (voice, tense)¹, e.g., *patyręs asmuo* “an experienced person” – *patirti* “to experience”; *papildomas* “additional” – *papildyti* “to add”; *prieinamos kainos* “affordable prices” – *prieiti* “to approach”; *teigiamas dalykas* “a positive thing” – *teigti* “to state”; *niekuo dėtas* “innocent” – *dėti* “to put”. It is essential to identify adjectival participles as they not only differ from other participles in meaning but also can begin to function as separate lexemes compared to other verb forms and form certain exceptions to usage. Since these participles begin to function as a separate lexeme, they can have separate lemmas from the corresponding verbs in dictionaries; identifying them is also vital in teaching Lithuanian as a foreign language; moreover, they are important in the morphological and syntactic annotation of corpora, in conducting various linguistic studies, and in the writing of grammars.

A study is being carried out, aiming to identify adjectival participles in an objective and partially automated way. It attempts to identify them based on various criteria: grammatical criteria (frequent attributive function; no verbal arguments and words indicating place and time; combination with adverbs of measure/degree; gradation (*mylimas* “beloved” – *mylimiausias* “most beloved”); frequent pronominal (definite) form; semantic criteria (change in lexical meaning from the corresponding verb; synonyms and antonyms with adjectives, the same semantic classes with adjectives); derivational criteria (tendency to form adverbs with the suffix *-ai* and abstract nouns with the suffixes *-umas* and *-ybė* from participles); and quantitative criteria (frequent use of participles in corpora compared to other forms of verbs). The criteria of adjectival participles are also mentioned in Jancaitė (2023).

Some of the mentioned criteria for adjectival participles (especially the semantic criteria) appear to be very subjective, time-consuming, so it was important to find a way to apply these criteria in a more objective manner (using corpus data). The aim of this paper is to present how one semantic criterion for adjectival participles (participles that have synonyms and antonyms similar to adjectives and share the same semantic classes as adjectives) was applied in a more objective way, based on word embeddings methodology.

The analysis was not limited to adjectives, as it is assumed that some participles can be used like pronouns (e.g., *praeitą / aną savaitę* “last week”: *praeitą* is a participle, *aną* is a pronoun). They are also different from other participles, so it was decided to analyse which participles are similar to adjectives and pronouns.

¹ Participles that lose their verbal properties but gain adjectival properties can be called adjectivalised participles (cf. Petrunina, 2021, Paulauskienė, 1994, Kustova, 2021). However, some participles that are semantically similar to adjectives still retain the meaning of a verb (e.g., *miręs* “dead” - *mirti* “to die”). In this paper, participles with adjectival meaning (whether they have lost verbal properties or are related to verbs) are referred to as adjectival participles.

The sources of the study are Pedagogic Corpus of Lithuanian² (hereinafter referred to as Pedagogic Corpus), as well as in the Lexical Database of Lithuanian Language Usage³ (hereinafter referred to as Lexical Database) and pre-trained *fastText* model, freely available on the internet⁴. The Pedagogic Corpus is a corpus designed for learners of Lithuanian at different levels of proficiency; it contains 699,000 words, comprising written (619,000 words) and spoken (35,000 words) texts. For more details on the Lexical Database, see Kovalevskaitė et al. (2022). The *fastText* model was trained on a corpus with social media comments (comments taken from Delfi.lt and lrytas.lt, written between 18-09-2014 and 30-05-2020, totalling 2 million comments). This corpus consists of 58,042,082 words, including 2,306,198 unique words.

Chapter 2 describes the word embedding approach that was crucial for identifying adjectival participles. Chapter 3 outlines the research methodology and presents the research results. The conclusions are then provided, followed by appendices containing analysed participles with and without semantically similar adjectives or pronouns.

2. Word Embeddings in Linguistic Analysis

Word embeddings show semantic relationships between words based on distributional (syntactic and lexical) information. Using word embeddings, words can be grouped by distribution into groups – word classes (see Kutuzov et al., 2016), word synonyms can be identified, and other studies can be carried out (cf. Gutiérrez and Keith, 2019).

Word embeddings are related to the research area of distributional semantics, which was founded on the ideas of Firth (1962), Harris (1954), and other researchers. The distributional hypothesis was derived, which states that words that are used and occur in the same contexts tend to have similar meanings (Firth, 1962). For instance, Harris (1954, p. 156) noticed that synonyms (like *oculist* and *eye-doctor*) tend to occur in the same environments. However, words that are not semantically similar will not share the same environments; for instance, *oculist* and *eye-doctor* may occur near words like *eye*, *examined*, but *lawyer* probably will not.

Thus, the idea that the meaning of a word is determined by context was already discussed in the first half of the 20th century. As technology has advanced, it has become possible to automatically analyse large volumes of texts, allowing for the automatic analysis of word distribution on a large scale.

Vector semantics is important to automatically estimate which words are used in similar contexts. In vector semantics, a word is represented as a point in a multidimensional semantic space that is derived from the distributions of neighbouring words. Vectors that represent words are called embeddings (Jurafsky and Martin, 2024, p. 105). These embeddings encode the characteristics of the words and the contexts in which they are used. This is also how similar words can be identified – similar words are represented by embeddings in a nearby space. Thus, words belonging to the same semantic classes, such

² <https://kalbu.vdu.lt/en/resources/pedagogic-corpus-of-lithuanian/>

³ <https://kalbu.vdu.lt/mokymosi-priemones/leksikonas/>

⁴ <http://fasttext.vdu.lt/>

as synonyms, antonyms, hyponyms, hyperonyms, and other related words can be identified (this is also reported in Kovalevskaitė et al., 2021; in this study, arbitrary collocations were identified using word embeddings).

Word embeddings are used for a diverse variety of tasks and cases, e.g., for domain-specific representations (Wang et al., 2021, Brandl et al., 2022), to discriminate inflection and derivation (Haley et al., 2023), detect semantic shifts in inflectional morphology (Gromann and Declerck, 2019), produce large analogical clusters (Hong and Lepage, 2018), distinguish denominal and root-derived verbs (Benbaji et al., 2022), predict semantic priming (Kastner, 2020), and evaluate gender bias in word representations (Altinok, 2024), to name just a few.

However, word embeddings have been used only to a limited extent in research of the Lithuanian language: examples include using *GloVe* embeddings for analysing arbitrary collocations (Kovalevskaitė et al., 2021) and identifying Lithuanian multiword expressions (Bumbulienė et al., 2018), developing a recogniser of hate/offensive speech in social media texts (hatespeech.vdu.lt). Word embeddings have also been applied in sentiment analysis of Lithuanian texts (Kapočiūtė-Dzikienė et al., 2019). Different types of word embeddings for Lithuanian have been used in deep learning cases such as news clustering (Stankevičius and Lukoševičius, 2020), automatic extraction of cybersecurity terms (Rokas et al., 2020), and the identification of semantic and syntactic similarity/relatedness (Petkevičius and Vitkutė-Adžgauskienė, 2021).

3. Word Embeddings for Identifying Adjectival Participles

Since one of the criteria for identifying adjectival participles is their semantic similarity with adjectives (including their synonymy, antonymy, and tendency to share the same semantic class), word embeddings were tested to automatically identify participles that are distributionally similar to adjectives and pronouns. Although distributional similarity is not the same as semantic similarity, it is believed to help detect semantic similarity in a more objective way than relying solely on introspection.

3.1. Methodology

As mentioned in the introduction, the sources of the study are the Pedagogic Corpus, the Lexical Database, and the pre-trained *fastText* model. The study includes 289 participle lemmas⁵. Lemmas were selected from the 200 most frequent verbs in the Lexical Database, e.g., *mylėti* “to love”; *žinoti* “to know”; *eiti* “to go”. Only participles with at least 5 occurrences in the Pedagogic Corpus in the simple (indefinite) form⁶ were

⁵ Usually, verb lemma also covers participles, e.g., *dirbti* “to work” also covers *dirbantis žmogus* “working person”. In this paper, nominative participle forms like *dirbantis* are called participle lemmas.

⁶ The pronominal (definite) forms of adjectives, participles, etc., indicate a specific, known object (e.g., *gražusis* “the beautiful one”), while the simple forms do not mark definiteness but only

analysed; in this study, pronominal (definite) participles were not included (e. g., *valgomas* “edible, eaten” was included, but not “valgomas”).

Initially, the idea of the research was to use a word embedding model trained with Pedagogic Corpus data (*fastText* and *word2vec* models were trained), as this corpus was used to select analysed words and analyse them. However, it turned out that the models trained with Pedagogic Corpus data had too little training data, resulting in a lack of accuracy (since this corpus is relatively small). In the end, the pre-trained *fastText* model, freely available on the internet⁷ was chosen.

FastText represents words as bags of character n-grams to capture sub-word information, making it robust to out-of-vocabulary words. *FastText* is based on *word2vec* as it uses skip-gram (it predicts context words when given a target word) and Continuous Bag of Words (CBOW; it predicts a target word when given context words) methods (Bojanowski et al., 2016; Naseem et al., 2021). It became popular due to being fast and efficient and due to models being available for a large number of languages (Joulin et al., 2017). Since *fastText* models do not ignore the morphology of words, i.e., they take into account the internal structure of words, the information about linguistic units smaller than the word, they are said to be more suitable for morphologically rich languages such as Lithuanian (other techniques associate each word with a distinct vector without parameter sharing (Bojanowski et al., 2017). However, since this model ranks words with similar morphology as more similar, it was decided in this research to search for the most frequent forms of words rather than their lemmas. Therefore, distributionally similar words for each word form were identified first (see Table 1). Later, the participles and the identified similar words were assigned to a single lemma. For each word form, 50 distributionally similar words were found. During this research, when comparing words, a certain estimate of word similarity was obtained – the closer it is to 1, the more similar the word is (see Table 1).

When analysing the participles, first of all, the 3 most frequent forms of the participles from the Pedagogic Corpus were entered into the search (for an example of a search with similar words, see Table 1), for instance, the forms of the participle *verdantis* “boiling”: *verdantį*, *verdančio*, *verdančiu*. If a participle in the Pedagogic Corpus has a comparative or superlative degree, the most frequent forms of the comparative/superlative degrees were also searched for separately (e.g., in addition to some forms of the participle *lankomas* “visited”, the superlative form *lankomiausias* “the most visited” was also searched for). However, if no distributionally similar adjectives or pronouns were found among these most frequent forms, all forms of the participles used in the Pedagogic Corpus were searched.

Thus, the initial search focused on adjectives and pronouns (see Table 1).

describe the characteristic of an object, without indicating definiteness (e.g., *gražus* “beautiful”). In Lithuanian, definite adjectives are formed by adding a suffix such as *-is* (masculine), *-oji* (feminine), etc. (Valeckienė, n.d.).

⁷ <https://fasttext.vdu.lt/>. During the pilot study, another *fastText* model—a Lithuanian model trained on a corpus from *Wikipedia* and Common Crawl, available as pre-trained word vectors from the *fastText* project (Grave et al., 2018)—was applied. However, due to uncertainty regarding the number of trained tokens and the nature of the texts in Lithuanian, this model was later disregarded.

Table 1. Similar words analysis using word embeddings.A search for the participle *praėjusi* (“past”, “passed”) (highest score is 1)

Nr.	Word	Score
1	Praėjusi	0.89438444
2	praėjusią	0.88527215
3	patį.Praėjusi	0.8692309
4	praeitą	0.8532955
5	“Praėjusi	0.8373966
6	užpraeitą	0.8287578
7	praeitą,	0.8278644
8	pirmąjį	0.81250143
9	Praėjusią	0.81066614
10	Įvykusį	0.805664
11	prasadėjusį	0.80437165
12	įvyksiantį	0.7967169
13	praeitą.	0.79365396
14	neįvykusį	0.7915372
15	būsiantį	0.79064226
16	praitą	0.7906226
17	pastarąjį	0.7896867
18	vykusį	0.78891814
19	besitęsiantį	0.78527635
20	šiųmetinį	0.78449905
21	“Praėjusią	0.7789983
22	užsitęsusių	0.7773801
23	(paskutinį	0.77316064

As can be seen in Table 1, there was also a lot of noise in the analysis – the results include different forms of the same participle (e.g., *praėjusią*, *Praėjusią*) and the participles were very often recognised as similar to other participles (e.g., *užpraeitą* “last”/“previous”, *įvykusį* “occured”, *prasadėjusį* “started”, *įvyksiantį* “that will happen”). Also, letters were not converted to lower case, and punctuation marks were not separated from words. Thus, at this stage of the research, linguistic editing was necessary, i.e.,

adjectives and pronouns were selected from all the similar words found. In this case (see Table 1), the adjectives selected are the following: *pastarąjį* “last”, *šiųmetinį* “this year’s”, *paskutinį* “last”.

Once the adjectives and pronouns were selected, it was important to check whether they are really semantically similar to the participle under analysis (As mentioned earlier, although word embeddings can help identify semantically similar words, this is not always the case, because word embeddings are based on statistics – they reflect distributional, rather than semantic, similarity, and this does not always align). For the purposes of this study, semantically similar words are assumed to be those that are used synonymously and antonymously and belong to the same semantic class. For instance, in this case (see Table 1), it has been recognised that the participle *praėjęs* “last” is used synonymously with the adjective *pastarasis* “last” and that this participle with the adjectives *pastarasis* “last” and *šiųmetinis* “of this year” belong to the same semantic class denoting time (for example, *praėjusią / pastarąją savaitę* “last week”). In another example, the participle *verdantis* “boiling” is synonymous to the adjective *karštas* “hot”, both of which belong to the semantic class denoting temperature. As mentioned before, the analysis was not limited to adjectives, as it is assumed that some participles can be used like pronouns, e.g., *praeitą / aną savaitę* “last week” (*praeitą* is a participle, *aną* is a pronoun).

Linguistic editing was also important at this stage of the study, as it was observed that the selection of the adjectives and pronouns resulted in some not being semantically similar to the participles (e.g., *nepamirštamas* “unforgettable” – *simboliškas* “symbolic”; *papildomas* “additional” – *periodinis* “periodic”). It is likely that these participles and adjectives were identified as similar because they are used in similar contexts, e.g., in similar terms. As mentioned before, distributional similarity does not always reflect semantic similarity.

However, without the context of usage, it is sometimes difficult to determine whether words are semantically similar or not. Therefore, during the linguistic editing stage, it was important to check the context of usage as well. The difficulty is that using this method, the context of words could not be seen, only similar words (as shown in Table 1). Therefore, the Pedagogic Corpus was used for this task. Moreover, participles can have multiple meanings. When analysing word embeddings data, it is not always clear which meaning is being represented by the similar words, as the wider context is not visible. For this reason, the concordances of the analysed participles in the Pedagogic Corpus were checked to determine the semantic context in which the participles are semantically similar to the adjectives or pronouns detected by the word embeddings (and whether such a context really exists). For example, the participle *praėjęs* (see Table 1) is semantically similar to the adjectives *pastarasis* and *šiųmetinis* in the following case:

*Čempionate neliko **praėjusį** sezoną paskutinės vietos užėmusios Elektrėnų „ESSM-2000” komandos.*

“**Last** season’s last-place team, Elektrėnai ESSM-2000 did not participate in the championship.”

In this case, the participle *praėjusį* can be replaced by adjectives *pastarąjį* “last” and *šiųmetinį* “this year’s”.

However, there are cases where the participle *praėjęs* cannot be considered semantically similar to the adjectives *pastarasis* and *šiųmetinis* because it conveys a

different meaning of the participle – not “last”, but “after passing” (this participle is semantically similar to the verb *praeiti* “to pass”):

Praėjęs pro gerinimo įrenginius, vanduo pasidaro švarus.

“After passing through the water treatment plant, the water becomes clean.”

Another example is a participle *patyręs* “experienced”. It can be considered an antonym of the adjective *jaunas* “young” in the following case:

Kovą iškilis finansinių sunkumų, bet dėl jų su niekuo nesitarkite – net su patyrusiais specialistais.

“There will be financial difficulties in March, but don’t discuss them with anyone – even experienced professionals.”

However, there are cases where the participle *patyręs*, used in a different context, cannot be considered an antonym of the adjective *jaunas* “young”:

Dar niekada, net ir vienišavimo laikais, nebuvau patyrusi tokios vienatvės.

“I have never experienced such loneliness, not even when I was single.”

In conclusion, the search for adjectives and pronouns that are semantically similar to participles consisted of two phases:

1. Adjectives and pronouns were selected from the lists of distributionally similar words found using word embeddings (on *fasttext.vdu.lt*).
2. The linguist checked whether these selected adjectives and pronouns are semantically similar to the participles (i.e., used synonymously, anonymously, belonging to the same semantic classes). This was checked by analysing the concordances of participles in the Pedagogic Corpus in order to see the context of their usage.

3.2. Results

The analysis of the participles revealed that 48 of the 289 participles (16.61%) are semantically similar to adjectives or pronouns. Analysed participles with and without semantically similar adjectives or pronouns can be found in the appendices.

We observed that some participles are semantically similar to adjectives only when used with other words, e.g., *atrodantis* “looking” is used synonymously with the adjectives *estetiškas* “aesthetic”, *simpatiškas* “handsome”, *išvaizdus* “handsome”, etc. only when used with certain adverbs of manner:

Esu kilnios širdies, liekna, gerai atrodanti, nuoširdi, paprasta, draugiška.

“I am noble-hearted, slender, good-looking, sincere, simple, friendly.”

The expression *gerai atrodanti* can be replaced by adjectives *simpatiška* “attractive” and *išvaizdi* “good-looking”.

Another example is the participle *dėtas* “put”, used synonymously with the adjective *nekaltas* “innocent”/“accidental” when combined with the pronoun *niekuo* “by nobody”:

*Ponas Filypas iš peties vožtelėjo kumščiu per **niekuo dėta** kavos staliuką.*
 “Mr Philip tapped his fist strongly on the **accidental** coffee table.”

The expression *niekuo dėta* can be replaced by the adjective *nekalta* “innocent”/“accidental”.

The participle *tikęs* “suitable”/“good” can also be used synonymously with the participles *beviltiškas* “hopeless”, *prasčiausias* “worst”, *nevertas* “unworthy”, etc., when used with the pronoun *niekam* “for nobody”, for example:

<...> *atrodo, jog dangus griūva, jog nieko nespėsite ir apskritai esate **niekam tikę** šeimininkai.*

“<...> it seems that the sky is falling, that you won’t be able to do anything in time, and that you are generally **useless** householders.”

The expression *niekam tikę* can be replaced by adjectives *beviltiški* “hopeless”, *prasčiausi* “the worst”, and *neverti* “unworthy”.

Although during this research 48 Lithuanian participles were identified as semantically similar to adjectives, only 23 of them are mentioned as synonyms or antonyms of Lithuanian adjectives in the Dictionary of Synonyms⁸ and Dictionary of Antonyms⁹.

The participles in this study, identified both by using word embeddings as semantically similar to adjectives and described in the Dictionary of Synonyms and Dictionary of Antonyms as having synonyms and antonyms of adjectives, are as follows:

- *atidarytas* “opened”/“open”,
- *atliekamas* “leftover”/“spare”,
- *gimtas* “native”,
- *išgėręs* “having drunk”,
- *keptas* “baked”/“fried”,
- *matomas* “visible”,
- *miręs* “dead”,
- *nematomas* “invisible”,
- *nematytas* “unseen”,
- *nepamirštamas* “unforgettable”,
- *nesuprantamas* “incomprehensible”,
- *pastebimas* “noticeable”,
- *praeitas* “last”/“previous”,
- *prieinamas* “accessible”/“available”/“affordable”,
- *priklausomas* “dependent”,
- *privalomas* “mandatory”,
- *suprantamas* “understandable”,
- *teigiamas* “positive”,
- *tikęs* “suitable”/“good”,

⁸ <https://ekalba.lt/sinonimu-zodynas/>

⁹ <https://ekalba.lt/antonimu-zodynas/>

- *tinkamas* “suitable”/“appropriate”,
- *valgomas* “edible”,
- *virtas* “cooked”/“boiled”,
- *vykęs* “good”/“successful”.

Participles identified using word embeddings as semantically similar to adjectives, but not described in the Dictionary of Synonyms and Dictionary of Antonyms as having synonyms and antonyms of adjectives are as follows:

- *ateinantis* (*ateinančią savaitę* “next week”),
- *atrodantis* (*gerai atrodantis* “handsome”),
- *derantis* “matching”/“fitting”/“good”,
- *dėtas* (*niekuo dėtas* “innocent”/“accidental”),
- *dirbantis* “working”,
- *lankomas* (*lankomiausios vietos* “the most visited places”),
- *likęs* “remaining”,
- *mėgstamas* “likeable”,
- *mylimas* “dear”/“beloved”,
- *mylintis* “loving”,
- *naudojamas* “used”/“utilised”/“useful”,
- *naudotas* (*naudotas automobilis* “used car”),
- *nurodytas* “indicated”,
- *papildomas* “additional”/“extra”,
- *pasirinktas* “chosen”,
- *pastebėtas* “noticed”,
- *patyręs* “experienced”,
- *pažįstamas* “familiar”/“acquainted”/“known”,
- *praėjęs* “past”,
- *skirtas* (*kepinui skirtas indas* “a dish intended for baking”),
- *tikintis* “religious”,
- *tinkantis* “fitting”,
- *vadinamas* “called”/“referred to as”,
- *veikiantis* “functioning”/“working”,
- *verdantis* “boiling”.

On the one hand, this may be because these dictionaries include only the most characteristic synonyms and antonyms, whereas we identified not only synonyms and antonyms but also words that belong to the same semantic classes. On the other hand, these dictionaries omit some fairly typical synonyms and antonyms (e.g., *naudotas* “used” – *naujas* “new”; *papildomas* “additional” – *bazinis* “basic”, *būtinai* “necessary”).

It is also worth noting that the participle *priėjęs* is listed as a synonym of the adjective *brandus* (“ripe”) in the Dictionary of Synonyms, but it was not identified as synonymous with this adjective when using word embeddings. For example, in the phrase:

Priėję grūdai, vaisiai, uogos.
 “Ripe grains, fruits, berries.”

However, this meaning of *priėjęs* is outdated and rarely used today, which may explain why it was not identified through word embeddings. The most common meaning of *priėjęs* is “having approached.”

Thus, semi-automated, word-embedding-based methods can, in theory, identify more synonyms and antonyms than a linguist might detect through introspection or by consulting certain Lithuanian language resources.

4. Conclusions

The word embeddings approach can help identify adjectival participles by identifying adjectives and pronouns that are semantically similar to participles. Of the 289 analysed participles, 48 (16.61%) were found to be semantically similar to adjectives and pronouns.

This method is useful because it can help identify which participles are semantically similar to adjectives and pronouns, even though sometimes it can be difficult to identify these participles by introspection or by analysing other Lithuanian language resources. For instance, in the Dictionary of Synonyms and Dictionary of Antonyms, only 23 out of 48 participles are described as synonymous or anonymous to adjectives. This method is quite simple and provides more objective data, which would be time-consuming to obtain directly from a corpus. However, even though this method identified adjectival participles, it did not determine which of them had changed lexical meaning from the verb and lost the grammatical categories of participles that are not characteristic of adjectives (voice and tense). For this purpose, additional criteria for adjectival participles need to be applied.

It is important to note that analysis based on word embeddings requires a linguist’s review of the results, which introduces an element of subjectivity at this stage of the research. In addition, this paper describes a study using the *fastText* model – it is likely that a different model may result in less noise in the analysis. Also, the training data should be of as high a quality as possible, for example, letters should be converted to lower case, and punctuation marks should be separated from words (this was not done in the present study). The results of the analysis also depend largely on the corpus – it is important to choose a sufficiently large and representative corpus in order to make the analysis as accurate as possible.

References

- Altinok, D. (2024). Gender Bias in Turkish Word Embeddings: A Comprehensive Study of Syntax, Semantics and Morphology Across Domains, in Faleńska, A., Basta, C., Costa-jussà, M., Goldfarb-Tarrant, S., Nozza, D. (eds), *Proceedings of the 5th Workshop on Gender Bias in Natural Language Processing (GeBNLP)*, Association for Computational Linguistics, Bangkok, Thailand, pp. 203–218, available at <https://aclanthology.org/2024.gebnlp-1.13.pdf>.
- Ambrazas, V. (ed.) (2006). *Lithuanian grammar*, second revised edition, Institute of the Lithuanian Language, Baltų lankų leidyba, Vilnius.
- Benbaji, I., Doron, O., Hénot-Mortier, A. (2022). Word-Embeddings Distinguish Denominal and Root-Derived Verbs in Semitic, in Moortgat, M., Wijnholds, G. (eds), *Proceedings End-to-*

- End Compositional Models of Vector-Based Semantics*, 33rd European Summer School in Logic, Language and Information (15-16 Aug. 2022, NUI Galway, Galway, Ireland), EPTCS 366, 2022, pp. 35–49, available at <https://arxiv.org/abs/2208.05721>.
- Bojanowski, P., Grave, E., Joulin, A., Mikolov, T. (2017). Enriching Word Vectors with Subword Information, *Transactions of the Association for Computational Linguistics* **5**, 135–146, available at <https://aclanthology.org/Q17-1010.pdf>.
- Bumbulienė, I., Mandravickaitė, J., Bielinskienė, A., Boizou, L., Kovalevskaitė, J., Rimkutė, E., Vilkaitė-Lozdienė, L., Man, K. L., Krilavičius, T. (2018). RNNs for Lithuanian Multiword Expressions Identification, *International Journal of Design, Analysis and Tools for Integrated Circuits and Systems (IJDATICS)*, **7**(1), 44–47.
- Firth, J. P. (1962). A synopsis of linguistic theory, 1930-1955, in Firth, J. et al. (eds), *Studies in Linguistic Analysis*, Blackwell, Oxford, pp. 1–32, available at <https://cs.brown.edu/courses/csci2952d/readings/lecture1-firth.pdf>.
- Grave, E., Bojanowski, P., Gupta, P., Joulin, A., Mikolov, T. (2018). Learning Word Vectors for 157 Languages, in Calzolari, N., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S. (eds), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, European Language Resources Association (ELRA), Miyazaki, Japan, available at <https://aclanthology.org/L18-1550/>.
- Gromann, D., Declerck, T. (2019). Towards the Detection and Formal Representation of Semantic Shifts in Inflectional Morphology, in Eskevich, M., de Melo, G., Fäth, Ch., McCrae, J. P., Buitelaar, P., Chiarcos, Ch., Klimek, B., Dojchinovski, M. (eds), *2nd Conference on Language, Data and Knowledge (LDK 2019)*, LDK 2019 (10-23 May 2019, Leipzig, Germany), *Open Access Series in Informatics (OASIs)*, Vol. 70, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany, pp. 21:1-21:15, available at <https://doi.org/10.4230/OASIs.LDK.2019.21>.
- Gutiérrez, L., Keith, B. (2019). A systematic literature review on word embeddings, in Mejia, J., Muñoz, M., Rocha, Á., Peña, A., & Pérez-Cisneros, M. (eds), *Trends and Applications in Software Engineering, CIMPS 2018* (Oct 17-18 2018, Guadalajara, Jalisco, México), *Advances in Intelligent Systems and Computing*, Vol. 865, Springer, Cham, pp. 132–141, available at https://doi.org/10.1007/978-3-030-01171-0_12.
- Haley, C., Ponti, E. M., Goldwater, S. (2023). Language-Agnostic Measures Discriminate Inflection and Derivation, in Beinborn, L., Goswami, K., Muradoğlu, S., Sorokin, A., Kumar, R., Shcherbakov, A., Ponti, E. M., Cotterell, R., Vylomova, E. (eds), *Proceedings of the 5th Workshop on Research in Computational Linguistic Typology and Multilingual NLP, EACL 2023* (May 2-6, 2023, Dubrovnik, Croatia), Association for Computational Linguistics, Dubrovnik, Croatia, pp. 150-152, available at <https://aclanthology.org/2023.sigtyp-1.18/>.
- Harris, Z. S. (1954). Distributional Structure, *Word* **10**(2-3), 146-162, available at <https://doi.org/10.1080/00437956.1954.11659520>.
- Hong, Y., Lepage, Y. (2018). Production of Large Analogical Clusters from Smaller Example Seed Clusters Using Word Embeddings, in Cox, M., Funk, P., Begum, S. (eds), *Case-Based Reasoning Research and Development International, ICCBR 2018* (10-12 Jul. 2018, Stockholm, Sweden), *Lecture Notes in Computer Science()*, Vol. 11156, Springer, Cham, pp. 548-562, available at https://doi.org/10.1007/978-3-030-01081-2_36.
- Jancaitė, L. (2023). Subūdvardėjusių lietuvių kalbos dalyvių atpažinimo kriterijai (Criteria for Identifying Adjectival Participles in Lithuanian), *Taikomoji kalbotyra* **20**, 183-207, available at <https://doi.org/10.15388/Taikalbot.2023.20.14>.
- Joulin, A., Grave, E., Bojanowski, P., Mikolov, T. (2017). Bag of Tricks for Efficient Text Classification, in Lapata, M., Blunsom, P., Koller, A. (eds), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, EACL 2017* (April 3-7, 2017, Valencia, Spain), Association for

- Computational Linguistics, Valencia, Spain, pp. 427-431, available at <https://aclanthology.org/E17-2068/>.
- Jurafsky, D., Martin, J. H. (2024). *Speech and Language Processing*, third Edition draft of January 12, 2025, available at https://web.stanford.edu/~jurafsky/slp3/ed3book_Jan25.pdf.
- Kapočiūtė-Dzikienė, J., Damaševičius, R., Woźniak, M. (2019). Sentiment Analysis of Lithuanian Texts Using Traditional and Deep Learning Approaches, *Computers* **8**, 4, available at <https://doi.org/10.3390/computers8010004>.
- Kastner, I. (2020). Predicting semantic priming in Hebrew morphology using word embeddings, in Poster presented at *AMLAP (Architectures and Mechanisms for Language Processing)*, Sep. 3-5, 2020, online conference).
- Kovalevskaitė, J., Boizou, L., Bielinskienė, A., Jancaitė, L., Rimkutė, E. (2022). The First Corpus-Driven Lexical Database of Lithuanian as L2, in Utkā, A., Vaičėnonienė, J., Kovalevskaitė, J., Kalinauskaitė, D. (eds), *Human Language Technologies – The Baltic Perspective*, Proceedings of the Ninth International Conference Baltic HLT 2020 (Sep. 22-23, 2020, Kaunas, Lithuania), IOS Press, Amsterdam, Berlin, Washington, DC, pp. 245-252, available at <https://ebooks.iospress.nl/doi/10.3233/FAIA200630>.
- Kovalevskaitė, J., Rimkutė, E., Vaičėnonienė, J. (2021). Arbitraliųjų lietuvių kalbos kolokacijų nustatymas (Identification of Lithuanian Arbitrary Collocations), *Bendrinė kalba* **94**, available at <https://etalpykla.lituanistika.lt/fedora/objects/LT-LDB-0001:J.04~2021~1662992740098/datastreams/DS.002.0.01.ARTIC/content>.
- Kutuzov, A., Veldal, E., Øvreid, L. (2016). Redefining part-of-speech classes with distributional semantic models, in Riezler, S., Goldberg, Y. (eds), *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning* (Aug. 11-12, 2016 Berlin, Germany), The Association for Computational Linguistics, Berlin, Germany, pp. 115-125, available at <https://aclanthology.org/K16-1012.pdf>.
- Lassner, D., Brandl, S., Baillot, A., Nakajima, S. (2023). Domain-Specific Word Embeddings with Structure Prediction, *Transactions of the Association for Computational Linguistics* **11**, 320-335, available at https://doi.org/10.1162/tacl_a_00538.
- Naseem, U., Razzak, I., Khan, S. K., Prasad, M. (2021). A Comprehensive Survey on Word Representation Models: From Classical to State-of-the-Art Word Representation Language Models, *ACM Transactions on Asian and Low-Resource Language Information Processing* **20**(5), 1-35, available at <https://doi.org/10.1145/3434237>.
- Paulauskienė, A. (1994). *Lietuvių kalbos morfologija: paskaitos lituanistams (The morphology of the Lithuanian language: Lectures to specialists of Lithuanian)*, Mokslo ir enciklopedijų leidykla, Vilnius.
- Petkevičius, M., Vitkutė-Adžgauskienė, D. (2021). Intrinsic Word Embedding Model Evaluation for Lithuanian Language Using Adapted Similarity and Relatedness Benchmark Datasets, in Veitaitė, I., Lopata, A., Krilavičius, T., Woźniak, M. (eds), *CEUR workshop proceedings: IVUS 2021: proceedings of the 26th international conference on information society and university studies* (April 23, 2021, Kaunas, Lithuania), vol. 2915, CEUR-WS, Kaunas, Lithuania, available at <https://ceur-ws.org/Vol-2915/paper14.pdf>.
- Petrūnina, U. (2021). *Adjectivization in Russian: Analyzing participles by means of lexical frequency and constraint grammar*, PhD thesis, The Arctic University of Norway, Norway, available at <https://munin.uit.no/bitstream/handle/10037/20757/thesis.pdf?sequence=2&isAllowed=y>.
- Rokas, A., Rackevičienė, S., Utkā, A. (2020). Automatic extraction of Lithuanian cybersecurity terms using deep learning approaches, in Utkā, A., Vaičėnonienė, J., Kovalevskaitė, J., Kalinauskaitė, D. (eds), *Human Language Technologies – The Baltic Perspective*, Proceedings of the Ninth International Conference Baltic HLT 2020 (Sep. 22-23, 2020, Kaunas, Lithuania), IOS Press, Amsterdam, Berlin, Washington, DC, pp. 39-46, available at <https://ebooks.iospress.nl/volumearticle/55521>.

- Stankevičius, L., Lukoševičius, M. (2020). Testing pre-trained transformer models for Lithuanian news clustering, in Lopata, A., Sukackė, V., Krilavičius, T., Veitaitė, I., Woźniak, M. (eds), *CEUR workshop proceedings: IVUS 2020: proceedings of the information society and university studies* (April 23, 2020, Kaunas, Lithuania), vol. 2698, CEUR-WS, Kaunas, Lithuania, pp. pp. 46-53, available at <https://arxiv.org/abs/2004.03461>.
- Valeckienė, A. (n.d.). Įvardžiutinės formos (Pronominal forms), *Visuotinė lietuvių enciklopedija* (Universal Lithuanian Encyclopedia), Mokslo ir enciklopedijų leidybos centras, available at <https://www.vle.lt/straipsnis/ivardziuotines-formos/>.
- Wang, Y., Huang, G., Li, J., Li, H., Zhou, Y., Jiang, H. (2021). Refined Global Word Embeddings Based on Sentiment Concept for Sentiment Analysis, *IEEE Access* **9**, 37075-37085, available at <https://doi.org/10.1109/ACCESS.2021.3062654>.

Received January 31, 2025, revised March 9, 2025, accepted March 10, 2025

Appendix A

Analysed participles without semantically similar adjectives or pronouns

Aplankęs (“having visited”), *atėjęs* (“having come”), *atidaręs* (“having opened”), *atidaromas* (“being opened”), *atiduotas* (“given away”), *atliekantis* (“performing”), *atlikęs* (“having performed”), *atliktas* (“performed”), *atsakęs* (“having answered”), *atsakomas* (“being answered”), *atsiradęs* (“having appeared”), *atvažiuavęs* (“having arrived (by vehicle)”), *atvažiuojantis* (“arriving (by vehicle)”), *atvykęs* (“having arrived”), *atvykstantis* (“arriving”), *augantis* (“growing”), *augęs* (“having grown”), *auginamas* (“being grown”), *auginantis* (“growing (something)”), *baigęs* (“having finished”), *baigiamas* (“being finished”), *baigtas* (“finished”), *bandomas* (“being tried/tested”), *dainuojantis* (“singing”), *dalyvaujantis* (“participating”), *dalyvavęs* (“having participated”), *darytas* (“made”), *daromas* (“being made”), *dedamas* (“being put”), *dirbamas* (“being worked”), *dirbęs* (“having worked”), *duotas* (“given”), *einantis* (“going”), *ėjęs* (“having gone”), *gaminamas* (“being made/produced”), *gaminantis* (“making/producing”), *gaunamas* (“being received”), *gautas* (“received”), *gavęs* (“having received”), *geriamas* (“being drunk”), *gimęs* (“born”), *girdėjęs* (“having heard”), *gyvenamas* (“inhabited/lived in”), *gyvenantis* (“living”), *gyvenęs* (“having lived”), *gyventas* (“lived”), *grįžęs* (“having returned”), *ieškantis* (“searching”), *ieškomas* (“being searched for”), *imamas* (“being taken”), *imtas* (“taken”), *įrengtas* (“installed”), *įsigijęs* (“having acquired”), *įsigytas* (“acquired”), *įsikūręs* (“having settled”), *išeinantis* (“leaving”), *išėjęs* (“having left”), *išgirdęs* (“having heard”), *išlikęs* (“having remained”), *išvykęs* (“having departed”), *jaučiamas* (“being felt”), *kalbamas* (“spoken”), *kalbantis* (“speaking”), *keičiamas* (“being changed”), *keliamas* (“being raised”), *keliantis* (“raising”), *keliaujantis* (“traveling”), *kepamas* (“being baked”), *ketinamas* (“intended”), *ketinantis* (“intending”), *kylantis* (“rising”), *kilęs* (“having risen/originated”), *kuriamas* (“being created”), *kuriantis* (“creating”), *kurtas* (“created”), *kviečiamas* (“being invited”), *laikantis* (“holding”), *laikytas* (“held”), *laikomas* (“being held/considered”), *lankantis* (“visiting”), *laukiamas* (“awaited”), *laukiantis* (“waiting”), *lauktas* (“waited for”), *leidžiamas* (“being allowed”), *leidžiantis* (“allowing/descending”), *manytas* (“thought”), *manomas* (“believed/assumed”), *matęs* (“having seen”), *matytas* (“seen”), *mėgstantis* (“liking”), *miegantis* (“sleeping”), *miegamas* (“being slept in”), *mokamas* (“paid”), *mokantis* (“knowing/able to”), *nematęs* (“not having seen”), *nešantis* (“carrying”), *nuėjęs* (“having gone (on foot)”), *nurodantis* (“indicating”), *nurodomas* (“being indicated”), *nusipirkęs* (“having bought (for oneself)”), *nusprendęs* (“having decided”), *nuspręstas* (“decided”), *nustatytas* (“set”, “established”), *nustatomas* (“being determined”), *padaręs* (“having done”), *padarytas* (“done”), *padaromas* (“being done”), *padedamas* (“being helped/placed”), *padedantis* (“helping”), *padėjęs* (“having placed/helped”), *padėtas* (“placed”), *paėmęs* (“having taken”), *pakeistas* (“changed”), *pakeitęs* (“having changed”), *pakėlęs* (“having raised”), *pakeliamas* (“being raised”), *pakeltas* (“raised”), *pakilęs* (“having risen”), *pakviestas* (“invited”), *pakvietęs* (“having invited”), *palikęs* (“having left”), *paliktas* (“left”), *pamatęs* (“having seen”), *pamiršęs* (“having forgotten”), *parašęs* (“having written”), *parašytas* (“written”), *parduodamas* (“being sold”), *parduotas* (“sold”), *parodytas* (“shown”), *paruoštas* (“prepared”), *pasakęs* (“having said”), *pasakytas* (“said”), *pasakojamus* (“being told”), *pasiekęs* (“having achieved”), *pasiekiamas* (“reachable”), *pasiejęs* (“having taken (for oneself)”), *pasirenkamas* (“being chosen”), *pasirinkęs* (“having chosen”), *pasirinktas* (“chosen”), *pasirodęs* (“having appeared”), *pasiūlytas* (“offered”), *paskambinęs* (“having called”), *pastatęs* (“having built”), *pastatytas* (“built”), *pastebėjęs* (“having noticed”), *pastebėtas* (“noticed”), *patariamus* (“being advised”), *pateikęs* (“having presented”), *pateikiamas* (“being presented”), *pateiktas* (“presented”), *patekęs* (“having reached”), *patiekiamas* (“being served”), *patiektas* (“served”), *patirtas* (“experienced”), *perkamas* (“being bought”), *pirktas* (“bought”), *planuojamas* (“being planned”), *pradedamas* (“being started”), *pradedantis* (“starting”), *pradėjęs* (“having started”), *pradėtas* (“started”), *praleidžiantis* (“spending (time)”), *pranešamas* (“being announced”), *praneštas* (“announced”), *prasisėjęs* (“having begun”), *priėjęs* (“having

approached”), *priėmęs* (“having accepted”), *priimamas* (“being accepted”), *priimtas* (“accepted”), *priklausantis* (“belonging”), *priklausęs* (“having belonged”), *primenantis* (“reminding”), *prisiminęs* (“having remembered”), *radęs* (“having found”), *randamas* (“being found”), *rastas* (“found”), *rašęs* (“having written”), *rašomas* (“being written”), *reiškiantis* (“meaning/expressing”), *renģiamas* (“being prepared”), *renkamas* (“being collected/elected”), *rodantis* (“showing”), *rodomas* (“being shown”), *sakęs* (“having said”), *sakomas* (“being said”), *saugantis* (“protecting”), *saugomas* (“being protected”), *sėdintis* (“sitting”), *siekiamas* (“being aimed for”), *siekiantis* (“aiming for”), *siūlantis* (“offering”), *siūlomas* (“being offered”), *skaitęs* (“having read”), *skiriamas* (“being assigned”), *skiriantis* (“distinguishing”), *spaudžiamas* (“being pressed”), *statytas* (“built”), *statomas* (“being built”), *stebintis* (“observing”), *stovėjęs* (“having stood”), *stovintis* (“standing”), *sudarantis* (“forming/constituting”), *sudarytas* (“formed”), *sudaromas* (“being formed”), *sudedamas* (“being put together”), *sudėtas* (“put together”), *sukėjęs* (“having caused”), *sukeliantis* (“causing”), *sukeltas* (“caused”), *sukūręs* (“having created”), *sukurtas* (“created”), *sulaukęs* (“having received/awaited”), *supjaustytas* (“sliced”), *supjaustomas* (“being sliced”), *supratęs* (“having understood”), *suradęs* (“having found”), *susipažinęs* (“having become acquainted”), *susitikęs* (“having met”), *suteikiamas* (“being granted”), *suteikiantis* (“granting”), *suteiktas* (“granted”), *sutikęs* (“having agreed/met”), *sutiktas* (“agreed/met”), *sužinojęs* (“having learned”), *tapęs* (“having become”), *tekęs* (“having fallen to (one's lot)”), *tikimas* (“being believed”), *tikimasi* (“it is expected”), *vadintas* (“called”), *vaikščiojantis* (“walking around”), *vaikštantis* (“walking”), *vartojamas* (“being used”), *vartojantis* (“using”), *važiuavęs* (“having traveled (by vehicle)”), *važiuojantis* (“traveling (by vehicle)”), *veikęs* (“having worked/functioned”), *veikiamas* (“being influenced”), *verdamas* (“being boiled”), *vertinamas* (“being evaluated”), *vertinantis* (“evaluating”), *vykstantis* (“happening”).

Appendix B

Analysed participles (in alphabetical order) with semantically similar adjectives or pronouns

Participles	Semantically similar adjectives and pronouns
<i>ateinantis</i> (“upcoming”, “approaching”, “next”)	<i>dabartinis</i> (“current”), <i>artimiausias</i> (“closest,” “nearest”), <i>paskutinis</i> (“last”, “final”), <i>šis</i> (“this” - pronoun), <i>šitas</i> (“this one” - pronoun), <i>anasis</i> (“that one (farther)” - pronoun), <i>tas</i> (“that” - pronoun)
<i>atidarytas</i> (“opened”, “open”)	<i>atdaras</i> (“open”)
<i>atliekamas</i> (“leftover”, “spare”)	<i>laisvas</i> (“free”, “available”), <i>papildomas</i> (“additional”, “extra”)
<i>atrodantis</i> (“appearing”, “seeming”), <i>gerai atrodantis</i> (“handsome”)	<i>elegantiškas</i> (“elegant”), <i>simpatiškas</i> (“sympathetic,” “likeable”), <i>išvaizdus</i> (“handsome”, “presentable”), <i>žavingas</i> (“charming”), <i>nesimpatiškas</i> (“unlikeable”), <i>išraiškingas</i> (“expressive”), <i>neišvaizdus</i> (“plain”, “unattractive”), <i>estetiškas</i> (“aesthetic”), <i>stilingas</i> (“stylish”), <i>dailus</i> (“pretty,” “neat”)
<i>derantis</i> (“matching”, “fitting”, “good”)	<i>panašus</i> (“similar”), <i>identiškas</i> (“identical”)
<i>dėtas</i> (“placed”, “blamed”), <i>niekuo dėtas</i> (“innocent”, “accidental”)	<i>nekaltas</i> (“innocent”), <i>kaltas</i> (“guilty”)
<i>dirbantis</i> (“working”)	<i>bedarbis</i> (“unemployed”), <i>kvalifikuotas</i> (“qualified”), <i>darbingas</i> (“able to work”)
<i>gimtas</i> (“native”), <i>gimtas miestas</i> “hometown”	<i>svetimas</i> (“foreign”), <i>tolimas</i> (“distant”)
<i>išgėręs</i> (“having drunk”)	<i>blaivus</i> (“sober”), <i>neblaivus</i> (“intoxicated”), <i>girtas</i> (“drunk”), <i>girtutėlis</i> (“very drunk”)
<i>keptas</i> (“baked”, “fried”)	<i>šviežias</i> (“fresh”)
<i>lankomas</i> (“visited”), <i>lankomiausios vietos</i> (“the most visited places”)	<i>įdomiausias</i> (“most interesting”), <i>smagiausias</i> (“most fun”), <i>linksmiausias</i> (“funniest”), <i>mieliausias</i> (“most charming”), <i>populiariausias</i> (“most popular”), <i>įspūdingiausias</i> (“most impressive”), <i>garsiausias</i> (“most famous”), <i>vaizdingiausias</i> (“most picturesque”), <i>nykiausias</i> (“most boring/dull”)
<i>likęs</i> (“remaining”)	<i>didysis</i> (“major”), <i>mažytis</i> (“minor”, “tiny”), <i>žymus</i> (“significant”)
<i>matomas</i> (“visible”)	<i>tamsus</i> (“dark”), <i>akivaizdus</i> (“evident”, “obvious”), <i>raiškus</i> (“clear”, “distinct”), <i>vaizdus</i> (“vivid”, “graphic”)
<i>mėgstamas</i> (“liked”, “favorite”, “likeable”)	<i>geras</i> (“good”), <i>prastas</i> (“poor”), <i>kokybiškas</i> (“high-quality”), <i>žymiausias</i> (“most notable”), <i>populiariausias</i> (“most popular”), <i>įžymiausias</i> (“most famous”), <i>šūdiniausias</i> ¹⁰ (“worst”, “most awful”), <i>įdomiausias</i> (“most interesting”), <i>žaviausias</i> (“most charming”), <i>smagiausias</i> (“most fun”)

¹⁰ The lexicon of informal language was analysed, e.g., *šūdinas* “shitty”, *fainas* “cool”, as such words were found when analysing word embeddings.

<i>mylimas</i> (“loved”, “dear”)	<i>mielas</i> (“dear”, “sweet”), <i>artimas</i> (“close”), <i>savasis</i> (“one’s own”), <i>nesavas</i> (“alien”, “not one’s own”), <i>meilus</i> (“affectionate”), <i>nuostabus</i> (“wonderful”), <i>gerasis</i> (“the good one”), <i>fainiausias</i> (“coolest”), <i>žavingiausias</i> (“most charming”), <i>šauniausias</i> (“nicest”)
<i>mylintis</i> (“loving”)	<i>rūpestingas</i> (“caring”), <i>egoistiškas</i> (“selfish”), <i>nedėkingas</i> (“ungrateful”), <i>motiniškas</i> (“motherly”), <i>ištikimas</i> (“loyal”), <i>doras</i> (“honest”), <i>nesavanaudis</i> (“selfless”), <i>dorovingas</i> (“virtuous”)
<i>miręs</i> (“dead”)	<i>gyvas</i> (“alive”), <i>negyvas</i> (“lifeless”), <i>nebegyvas</i> (“no longer alive”)
<i>naudojamas</i> (“used”, “utilized”, “useful”)	<i>reikalingas</i> (“necessary”)
<i>naudotas</i> (“used”, “second-hand”)	<i>senesnis</i> (“older”), <i>antrinis</i> (“secondary”), <i>antikvarinis</i> (“antique”), <i>naujesnis</i> (“newer”)
<i>nematomas</i> (“invisible”)	<i>tamsiausias</i> (“darkest”), <i>paslaptingas</i> (“mysterious”)
<i>nematytas</i> (“unseen”)	<i>neįtikėtinas</i> (“unbelievable”)
<i>nepamirštamasis</i> (“unforgettable”)	<i>įprastas</i> (“usual”), <i>neįprastas</i> (“unusual”), <i>savotiškas</i> (“peculiar”), <i>nesėkmingas</i> (“unsuccessful”), <i>džiugus</i> (“joyful”), <i>neįtikėtinas</i> (“unbelievable”), <i>reikšmingas</i> (“significant”), <i>įdomiausias</i> (“most interesting”)
<i>nesuprantamas</i> (“incomprehensible”)	<i>keistas</i> (“strange”), <i>banalus</i> (“banal”), <i>elementarus</i> (“elementary”), <i>neesminis</i> (“non-essential”), <i>niekinis</i> (“insignificant”), <i>kvailas</i> (“stupid”), <i>neprotingas</i> (“unreasonable”), <i>painus</i> (“confusing”), <i>idiotiškas</i> (“idiotic”), <i>beprasmiškas</i> (“meaningless”)
<i>nurodytas</i> (“indicated”)	<i>konkretus</i> (“specific”), <i>tikslus</i> (“precise”), <i>netikslus</i> (“imprecise”)
<i>papildomas</i> (“additional”, “extra”)	<i>bazinis</i> (“basic”), <i>būtinasis</i> (“necessary”), <i>specialus</i> (“special”)
<i>pasirinktas</i> (“chosen”)	<i>tikslingas</i> (“purposeful”)
<i>pastebimas</i> (“noticeable”)	<i>ryškesnis</i> (“more distinct”), <i>tamsiausias</i> (“darkest”), <i>geriausias</i> (“best”), <i>populiariausias</i> (“most popular”), <i>įdomiausias</i> (“most interesting”), <i>puikiausias</i> (“excellent”)
<i>pastebėtas</i> (“noticed”)	<i>įdomus</i> (“interesting”), <i>reikšmingas</i> (“significant”), <i>akivaizdus</i> (“evident”)
<i>patyręs</i> (“experienced”)	<i>jaunas</i> (“young”), <i>veiksnius</i> (“capable”), <i>stiprus</i> (“strong”), <i>neprofesionalus</i> (“unprofessional”), <i>pajėgus</i> (“able”)
<i>pažįstamas</i> (“familiar”, “acquainted”, “known”)	<i>artimas</i> (“close”), <i>draugiškas</i> (“friendly”), <i>šeimyniškas</i> (“family-like”), <i>nedraugiškas</i> (“unfriendly”), <i>neaiškus</i> (“unclear”)
<i>praeitas</i> (“passed”, “last”, “previous”)	<i>senesnis</i> (“older”), <i>paskutinis</i> (“last”), <i>anas</i> (“that one (past)”), <i>šis</i> (“this one” - pronoun), <i>kitas</i> (“another” - pronoun)
<i>praėjęs</i> (“past”)	<i>šiometinis</i> (“this year’s”), <i>ankstesnis</i> (“earlier”), <i>šiųmetis</i> (“of this year”), <i>pastarasis</i> (“recent”), <i>šiandieninis</i> (“today’s”), <i>paskutinis</i> (“last”), <i>šis</i> (“this” - pronoun), <i>kažkuris</i> (“some one” - pronoun), <i>tas</i> (“that” - pronoun)
<i>prieinamas</i> (“accessible”, “available”, “affordable”)	<i>paprastas</i> (“simple”), <i>pigus</i> (“cheap”), <i>patogus</i> (“convenient”, “comfortable”)
<i>priklausomas</i> (“dependent”)	<i>savarankiškas</i> (“independent”), <i>laisvas</i> (“free”)

<i>privalomas</i> (“mandatory”)	<i>privalus</i> (“obligatory”), <i>priverstinis</i> (“forced”), <i>būtinus</i> (“necessary”)
<i>skirtas</i> (“intended”, “meant”), <i>kepinui skirtas indas</i> (“a dish intended for baking”)	<i>reikalingas</i> (“necessary”), <i>naudingas</i> (“useful”)
<i>suprantamas</i> (“understandable”)	<i>keistas</i> (“strange”), <i>keistokas</i> (“somewhat strange”), <i>mistinis</i> (“mystical”), <i>painus</i> (“confusing”), <i>akivaizdus</i> (“obvious”), <i>apgaulingas</i> (“misleading”)
<i>teigiamas</i> (“positive”)	<i>reikšmingas</i> (“significant”), <i>negatyvus</i> (“negative”), <i>geras</i> (“good”), <i>pozityvus</i> (“positive”), <i>blogas</i> (“bad”), <i>pesimistinis</i> (“pessimistic”), <i>viltingas</i> (“hopeful”), <i>optimistinis</i> (“optimistic”), <i>žalingas</i> (“harmful”), <i>negeras</i> (“not good”)
<i>tikintis</i> (“believing”, “faithful”, “religious”)	<i>bedieviškas</i> (“godless”), <i>religingas</i> (“religious”), <i>bedievis</i> (“atheist”), <i>nereligingas</i> (“non-religious”), <i>krikščioniškas</i> (“Christian”)
<i>tikęs</i> (“suitable”, “good”), <i>niekam tikęs</i> (“useless”, “worthless”)	<i>nejdomus</i> (“uninteresting”), <i>beviltiškas</i> (“hopeless”), <i>bukas</i> (“dull”), <i>bevertis</i> (“worthless”), <i>banalus</i> (“banal”), <i>menkavertis</i> (“low-value”), <i>nepraktiškas</i> (“impractical”), <i>prasčiausias</i> (“worst”), <i>profesionalus</i> (“professional”), <i>kvailas</i> (“foolish”)
<i>tinkamas</i> (“suitable”, “appropriate”)	<i>geras</i> (“good”), <i>idealus</i> (“ideal”), <i>pusėtinus</i> (“average”), <i>praktiškas</i> (“practical”), <i>maloniausias</i> (“most pleasant”), <i>puikus</i> (“excellent”), <i>reikalingas</i> (“necessary”)
<i>tinkantis</i> (“fitting”), <i>jums tinkanti šukuosena</i> (“the hairstyle that suits you”)	<i>geriausias</i> (“best”), <i>idealus</i> (“ideal”), <i>maloniausias</i> (“most pleasant”), <i>patogus</i> (“comfortable”), <i>reikalingas</i> (“necessary”)
<i>veikiantis</i> (“functioning”, “working”)	<i>aktyvus</i> (“active”), <i>funkcionalus</i> (“functional”), <i>efektyvus</i> (“effective”)
<i>adinamas</i> (“called”, “referred to as”)	<i>tas</i> (“that”), <i>savotiškas</i> (“peculiar”), <i>kažkoks</i> (“some kind of” - pronoun)
<i>valgomas</i> (“edible”)	<i>nuodingas</i> (“poisonous”)
<i>verdantis</i> (“boiling”)	<i>karštas</i> (“hot”)
<i>vykęs</i> (“successful”, “good”)	<i>sėkmingas</i> (“successful”), <i>nesėkmingas</i> (“unsuccessful”), <i>įdomus</i> (“interesting”), <i>savotiškas</i> (“peculiar”), <i>geriausias</i> (“best”)
<i>virtas</i> (“boiled”, “cooked”)	<i>šviežias</i> (“fresh”), <i>žalias</i> (“raw”)