Baltic J. Modern Computing, Vol. 12 (2024), No. 4, pp. 513–525
https://doi.org/10.22364/bjmc.2024.12.4.13

Tēzaurs as a Digital Multifunctional Lexical Resource

Agute KLINTS*, Mikus GRASMANIS**, Gunta NEŠPORE-BĒRZKALNE*,
Lauma PRETKALNIŅA***, Madara STĀDE*, Normunds GRŪZĪTIS***,
Ilze LOKMANE***, Pēteris PAIKENS***, Laura RITUMA***,
Andrejs SPEKTORS*

Institute of Mathematics and Computer Science University of Latvia,
Raiņa blvd. 29, Riga, Latvia

agute.klints@lumii.lv, mikus.grasmanis@lumii.lv, gunta@ailab.lv,
lauma@ailab.lv, madara.stade@lumii.lv, normunds@ailab.lv,
ilze.lokmane@lumii.lv, peteris@ailab.lv, laura@ailab.lv, aspekt@ailab.lv

ORCID 0009-0001-8532-7250, ORCID 0000-0002-0668-0970, ORCID 0009-0002-3496-6455,
ORCID 0000-0002-6444-5581, ORCID 0009-0003-9500-8928, ORCID 0000-0003-0511-1829,
ORCID 0000-0002-5842-4522, ORCID 0000-0002-5939-5436, ORCID 0000-0002-0408-7700,
ORCID 0009-0005-4695-2239

Abstract. In this paper we describe the implementation of the latest solutions and improvements in the largest Latvian online dictionary Tēzaurs. The solutions include the continuously expanding Latvian WordNet, a new derivation resource within the WordNet, an updated system of in-gloss lexeme references, and improved guidelines for virtual links for better navigation between different elements of the dictionary. The derivation resource, which is currently underway, includes morphological and semantic information and reflects the complex derivation process in the Latvian language; we expect that it will prove especially useful for studying the theoretical and practical aspects of Latvian morphology and semantics. Virtual and in-gloss links, on the other hand, improve the usability and navigability of Tēzaurs through the use of, e.g., anchor links and automatically generated links to corpora examples, providing the users with a broader overview of their queries. Thus, improvements are being made on various levels of the Tēzaurs structure to meet the needs of an increasingly digitized research and study environment.

Keywords: wordnet, derivational semantics, heuristics, lexicography, Latvian

* Base funding of IMCS
** The State Research Programme project (VPP-LETONIKA-2021/1-0006)
*** Latvian Council of Science project (LZP2022/1-0443)

1 Introduction

Tēzaurs[1] is the largest electronic explanatory Latvian dictionary (Grasmanis et al., 2023) with more than 400,000 entries. It has started as an electronic compilation of multiple printed dictionaries, but in this paper we want to focus on the unique features Tēzaurs has acquired as a digital resource greatly transcending its paper dictionary origin. The digital world provides many capabilities, including quick and varied search possibilities, huge language material availability and unique backup copy storage challenges. Digitally, Tēzaurs is also treated as a large collection of objects (entries, lexemes, senses, examples etc.), which can be linked in various interesting ways. In the last 5 years Tēzaurs had served as a basis for a multitude of research initiatives increasing its inter- and intraconnectedness.

The first step was the deduplication of multi-word expressions (MWE). For ease of searching, paper dictionaries like LLVV tend to duplicate some information, such as common MWE explanations, in all related single-word entries, thus, also providing space for inconsistencies. Digital dictionaries like Tēzaurs do not need such duplicates as each MWE can be explained once and linked to all relevant single-word entries.

This was followed by the creation of Latvian WordNet (Paikens et al., 2023), which is based on Tēzaurs and built according to the principles of Princeton WordNet (Fellbaum, 1998). In it, the word senses are included in sets of synonyms among which other types of semantic relations are also created. This lexical resource also contains links between Latvian WordNet and Princeton WordNet.

The experience gained in creating Latvian WordNet later helped us in derivation annotation with a quite ambitious goal to provide derivational links on both morphological and semantical levels. All this research has been greatly influenced and aided by Tēzaurs' digital nature, and the results of this research, in turn, help develop Tēzaurs as a truly multifunctional, one of a kind resource, as can be seen in the example entry provided in Figure 1.

Chapter 2 of this paper gives insight into Latvian WordNet elements and its structure. Chapter 3 describes derivational links integrated in Tēzaurs. Chapter 4 and 5 outlines types of virtual links featured in Tēzaurs – links between senses and entries, and links to other dictionaries. Finally, the last chapter of the article consists of conclusions and plans for Tezaurs' future developement.

2 Latvian WordNet

WordNet is an extensive lexical and semantic resource that provides structured information about the word senses of nouns, verbs, adverbs, and adjectives, as well as the relations between these senses (Fellbaum, 1998). The creation of such a linguistic resource is significant for digital processing of the Latvian language and its further study in an increasingly digitized environment.

Latvian WordNet has been manually developed since 2019 (Paikens et al., 2022, 2023). As of Summer 2024, Latvian WordNet contains 8648 synsets which cover

[1] Publicly available for viewing at `https://tezaurs.lv` and for downloading at `http://hdl.handle.net/20.500.12574/107`

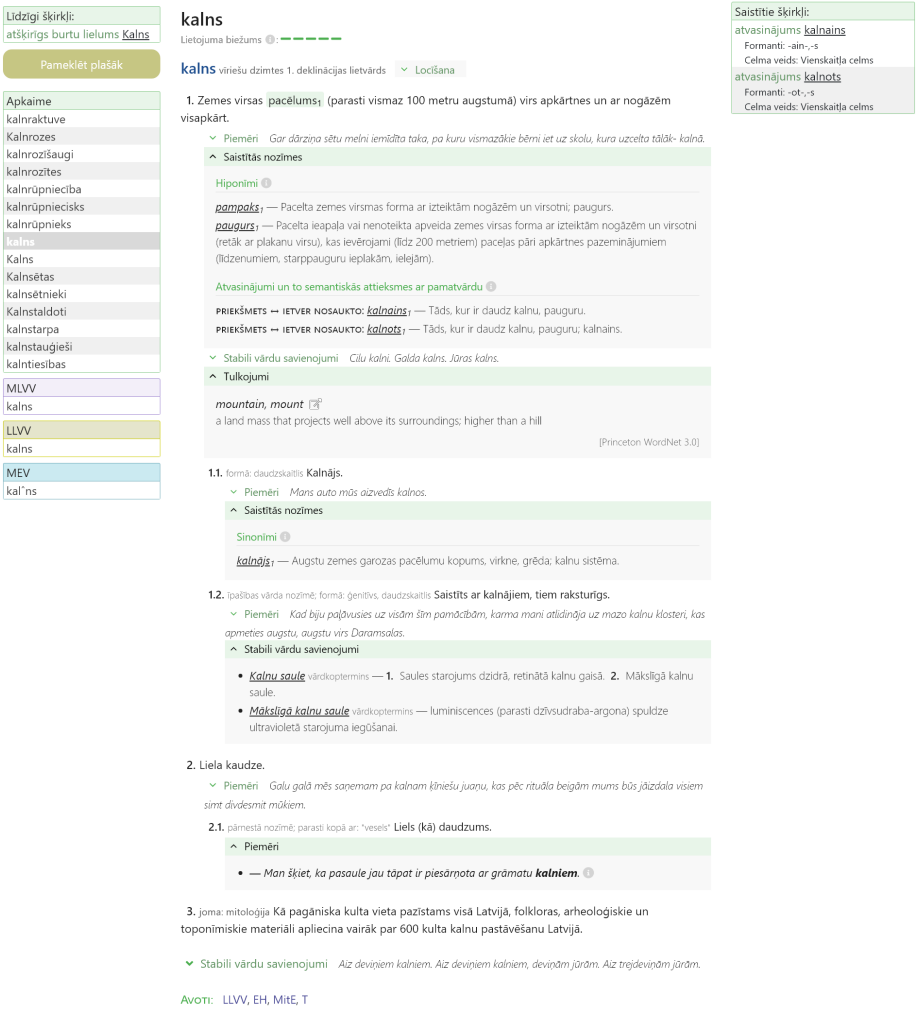

Fig. 1. Example from Tēzaurs: an entry ‘kalns’ (mountain, hill) as shown in `https://tezaurs.lv`

15626 word senses of the 2000 most frequently used words in The Balanced Corpus of Modern Latvian (Levāne-Petrova, 2019) and their related synsets and derivations (about derivations see more in Section 3). The inventory of words and senses is based on the Tēzaurs online dictionary (Grasmanis et al., 2023). The goal of the Latvian WordNet development is to create the highest quality material possible, so that based on manually created data afterwards the automatically expanded network would be as accurate as possible.

Although Latvian WordNet is still a comparatively small resource, we want to expand it with derivational links, similar to what Princeton WordNet (Mititelu et al., 2021) and others have implemented (e.g. Turkish (Bilgin et al., 2004), Bulgarian (Koeva, 2021), Romanian (Tufis et al., 2004), Czech (Rambousek et al., 2018)).

The information compiled in a wordnet is much broader than that in synonym dictionaries, as it provides insight into a wider linguistic context: in wordnet, specific word meanings (not just lemmas) are linked through semantic relations; the structure is a network in which word senses are linked to their synonyms, and these synsets are connected to other senses by other semantic relations; the meanings of the words involved in the network are also provided with references to additional meaning components; the problem of determining the dominant synonym in a synonym set is not addressed (see more in Veidemane (1970)), all components are considered equal; hyponyms are not listed within the synonym set, as it sometimes is in a synonym dictionary.

The Latvian WordNet is being developed based on the explanatory dictionary Tēzaurs, and the structure of the network is also influenced by the division of meanings in this particular dictionary (see more about criteria for distinguishing word senses in Tēzaurs in (Lokmane et al., 2021)).

WordNet links in Tezaurs.lv can be viewed in the word entry next to the corresponding sense in the section “saistītās nozīmes” (related senses) (see Fig. 1).

2.1 Semantic relations

The main semantic relations used in the creation of a wordnet are synonymy, antonymy, hyponymy, and meronymy (Jurafsky and Martin, 2024, pp. 4-5). In addition to the mentioned relations, the Latvian WordNet also includes gradation, which have been used also, for example, in the creation of the Polish WordNet (plWordNet) (Rudnicka et al., 2012), as well as similarity and conceptual connection links.

As mentioned above, WordNet’s structure is based on the relations that form between the synsets, therefore, **synonymy** is the primary semantic relation of the network. The synonym category in the Latvian WordNet includes both absolute and near synonyms (similarity relation) (see more in e.g. Veidemane (1970); Löbner (2002)), regardless of additional meaning components – a synset may include both neutral and expressive senses. Synset (*ēst$_1$, ammāt$_1$, štopēt$_4$, tiesāt$_1$, klemzt$_1$*) ‘to eat’ contains many word senses (not all are shown here), some of which are neutral, some contain emotional expressivity (they are used when talking to babies, i.e. they have a positive connotation), some are slang words, used only in particular social groups, some contain dialectic meaning components.

Usually synsets consist of multiple elements, but they can also consist of just one word sense. For example, if a sense is included in the WordNet (it is linked with another

sense by some other semantic relation links), but has no synonyms, this single network element is still considered a synset because new elements can potentially be added to this set as the network develops.

Hyponymy occurs when one meaning (the hyponym) refers to a narrower, more specific concept than the other, broader meaning (the hypernym). Hyponymy typically occurs between nouns and verbs. Here are some examples: hypernym (*darbarīks$_1$*) 'tool' and its hyponyms (*āmurs$_1$*) 'hammer', (*lāpsta$_1$*) 'shovel', (*sirpis$_1$*) 'sickle, reaping-hook'; hypernym (*radinieks$_1$, radagabals$_1$*) 'relative, relation' and its hyponyms (*sieva$_1$, laulene$_1$, vecene$_2$*) 'wife, married woman', (*mamma$_1$, māte$_1$, mammucīte$_1$, māmīte$_1$*) 'mother, mom, ma', (*vedekla$_1$, vedene$_1$, ļaudava$_2$, dēlasieva$_1$*) 'daughter-in-law'.

Antonymy in the Latvian WordNet includes all types of contrast relationships: binary oppositions (*woman/man*), the extreme points of a gradation scale (*cold/hot*), opposites derived from the same root (*alive/dead*) (in Latvian *dzīvs/ne-dzīvs*), and others (see more details, e. g., Löbner (2002); Griffiths and Cummins (2017)). Some examples of antonymy is (*mākoņains$_{1.1}$*) and (*dzidrs$_{1.3}$, skaidrs$_{1.1}$*) 'cloudy' and 'clear', and (*melnbalts$_1$*) and (*krāsains$_{1.2}$*) 'black and white' and 'color (adj) (having colors)'.

Meronymy refers to the relationship between a part and a whole. It is characteristic mainly of concepts related to physical objects, so in the lexical network, it is applicable only to nouns, and these connections are rarely drawn. Meronymy is drawn, e.g., between the synset (*mašīna$_2$, auto$_1$, automobilis$_1$*) 'car, auto, automobile' and its parts - (*motors$_1$*) 'motor', (*virsbūve$_1$*) 'bodywork', (*salons$_4$*) 'inside, interior'.

Gradation links are not created between just two concepts but instead form a network within a group of concepts. A group of gradation concepts consists of words that share a common semantic element but differ in intensity, speed, size, or another quantitative factor, creating a gradual transition between the words. For example (*dievināt$_1$, pielūgt$_2$*) 'to worship', (*mīlēt$_2$, mīļot$_3$*) 'to love', (*pieciest$_1$*) 'bear, stand, tolerate', (*ienīst$_1$, nīdēt$_1$*) 'to hate' is a gradation because the intensity of the semantic element 'to love' changes gradually in these concepts.

Similarity links are created in cases where the meanings are semantically similar, but cannot be considered synonyms due to semantic or grammatical reasons.

In addition to the semantic links mentioned above, we also annotate **conceptual connections**, as a category for words that are semantically related, but not by any of the mentioned semantic relations.

For a more detailed description of semantic links see Paikens et al. (2023).

2.2 Interlingual links

To identify English equivalents for Latvian word meanings, Latvian WordNet to Princeton WordNet sense mapping is carried out; currently, Latvian word senses are being mapped to Princeton WordNet 3.0 (Strankale and Stāde, 2022).

In Latvian WordNet, interlingual links indicate a **complete correspondence** of synonym sets in Latvian and English, as well as cases when a Latvian synset is **narrower** or **broader** in meaning (Paikens et al., 2023). The links may not be equivalent if not all relevant semantic elements are included in the sets (Miller et al. 1993). The example of complete correspondence is Latvian synset (*piemērs$_1$, ilustrācija$_2$*) and its corresponding Princeton WordNet synset ('example, illustration, instance, representative'). But,

for instance, Latvian synset (*pilsēta*$_1$) corresponds to both synsets ('city, metropolis, urban center') and ('town'), so in this case the Latvian synset is marked as broader in meaning than both of these English synsets.

It should also be considered that each language system differs from others, so it is not always possible to denote something in English with a single word when it can be named with one word in Latvian (Löbner, 2002, pp. 156), for example, in Latvian a concept 'the day before yesterday' can be expressed in one word (*aizvakar*$_1$), but in English there is no such entry in a dictionary. In such cases, a semantic link is either not established or is marked as a broader meaning, forming interlingual hyponymy.

Interlingual links can be viewed in Tezaurs.lv in the word entry next to the corresponding sense in the section "tulkojumi" (translation) (see Fig. 1).

3 Derivation resource

Expanding the Latvian WordNet with derivational links is one of the ongoing processes currently being carried out within the Tēzaurs dictionary. Up until now, according to the traditions of lexicography, the regular derivatives listed in Tēzaurs had their own entries only if they had a specific sense which had become distant enough from the meanings of the base word. In order to represent the diversity of derivational relations, new entries for regular derivatives are also being created. As mentioned in Section 2 derivative word senses are also being included in the WordNet.

So far in Latvian linguistics, little or no attention has been paid to semantic relations between the senses of a polysemous base word and the senses of its derivatives; only general semantics of derivational formatives has been studied and described referring to the basic sense of the base word (Kalnača and Lokmane, 2021; Soida, 2009). Therefore, we have chosen to employ two kinds of derivational links in our derivation resource: **semantic derivation links** between word senses and **morphological derivation links** between lexemes.

At the moment, the semantics of derivatives is a network parallel to Latvian WordNet, the word sense inventory being the unifying element that is involved in both networks. In order to ensure a reliable resource for future research, the data is created manually. However, we assume that some semi-automatic methods could also be applied to unambiguous words in the future.

3.1 Morphological linking

The Tēzaurs.lv entries include relevant lexemes in a formal morphological structure (Paikens et al., 2024) and also support annotation of formal morphological derivational links that connect motivating word entry with the derived word (Grasmanis et al., 2023). This link is supplemented with two attributes: the derivational stem base and the derivational formative. The stem indicates which part and form of the motivating word the derivative is formed from; a formative is the means by which a new word is made, and it can be a single morpheme, such as a prefix or a suffix, or a combination of morphemes, such as a suffix and an ending, that together form a complex formative. For example, the adjective *kalnains* 'mountainous' is formed by adding formatives *-ain-*

and *-s* to the stem of the noun *kalns* 'mountain' (see upper-right corner of Fig.1). However, our aim is not to divide the entire word into morphemes; the internal composition of Latvian words is the objective of another project, "Database of Latvian Morphemes and Derivational Models" (see `https://www.dlmdm.lu.lv`). Instead, we only indicate the morphemes involved in the derivative process.

Since the Latvian language has an extremely rich inflectional and derivational morphology (Kalnača and Lokmane, 2021), new words can be made from various stems, e.g., the present, past, infinitive or participle stems of the verbs, singular or plural stems of the nouns, using prefixes, suffixes and interfixes. Thus recording information about the derivational stem seems to be crucial in describing Latvian derivational morphology.

Currently, the morphology links in Tēzaurs are only directional, i.e., indicating which is the motivating word and which is the derivative. However, there are cases in the language material, when it is unclear, which word is derived from which and whether they are derived from one another at all, even though both words appear to be derivationally connected in some way; an example for such a case would be the noun *spēle* 'a game' and the verb *spēlēt* 'to play' or noun *fizika* 'physics' and adjective *fizisks* 'physical'. Currently there is no complete solution for displaying such cases, but we expect that in the future there will be a possibility to display words sharing the same root in nests. This will enable users to browse related words even if the derivative relations between them are not possible to be clearly defined.

Morphological links between lexemes in Tezaurs can be viewed in the upper-right corner of the entry, in the "saistītie šķirkļi" (related entries) section (see Fig. 1).

3.2 Semantic linking

Even within the boundaries of one word and its derivatives, there can be a large variety of semantic relations, especially when all the senses of a word are considered; the derivation resource aims to illustrate and document these relations. Semantic links between senses are formed as a pair of semantic labels, which are given to both ends of the link. For example, in the derivational combination "noun-adjective", a noun *brīnums*$_1$ 'wonder' (semantic label: ABSTRACTNOTION) is paired with an adjective *brīnumains*$_1$ 'wondrous' (semantic label: SIMILARTO), whereas *brīnums*$_2$ 'miracle' is paired with *brīnumains*$_2$ 'miraculous' with the same semantic labels as in the former example. Thus, two semantic label pairs (ABSTRACTNOTION–SIMILARTO) were formed. It is important to note that not all such derivative sense pairs are symmetrical (e.g. "sense 1" – "sense 1") and there are instances, when certain senses have no corresponding derivative sense at all.

The list of semantic labels is not yet fixed and final, and will be enriched as we proceed with other lexical groups. Current efforts concentrate on semantic links between nouns, verbs and adjectives, thus, the current sets of labels represent the semantic links between these word groups specifically. There are currently 32 labels for three word group combinations (noun-verb, verb-noun and noun-adjective). For example, verbs are classified into three basic types – actions, states and processes – with respect to semantic properties relevant for describing semantic roles of the arguments. The semantic label of the derivative noun characterizes the semantic (thematic) role of the participant involved in the event.

Assignment of a single semantic label to a sense is not always indisputable, as clear-cut boundaries in semantics do not exist. Thus, at this point, the semantic labels also include the category of 'other' for ambiguous cases that will be revised in the future.

Only after refining the sense inventory of the derivative, the semantic derivation links are formed between the senses of the base and the derivative. Semantic labels are assigned to each end of the link based on the possible semantic roles for a specific word class and category. When assigning a role, it must be taken into account that different senses of one word can correspond to different roles, both for the base word and the derivative. If necessary, the list of semantic labels is supplemented. One example of such supplementing is the label RELATEDTO, which is used when a derivative expresses a connection to the base word (and not similarity or being a part of something), e.g. *zinātne* 'science' and *zinātnisks* 'scientific'.

Semantic links in Tēzaurs can be viewed next to the corresponding sense in the section's "saistītās nozīmes" (related senses) subsection "atvasinājumi un to semantiskās attieksmes ar pamatvārdu" (derrivations and their semantic labels) (see Fig. 1).

4 Lexeme references from inside the sense gloss

Most of the links featured in Tēzaurs connect atomic objects, i.e, they are between full senses, lexemes or entries. However paper dictionaries serving as an initial basis of the Tēzaurs often contained references inside the text of the gloss of the sense to disambiguate homonym or otherwise ambiguous word used in the gloss explanation. This information could not be easily discarded without reformulating the gloss text, thus, we needed to introduce a mechanism to force these references automatically update whenever the target homonym index or sense number changes. This lead us to introduce what Tezaurs.lv platform informally calls *anchor links* – an asymmetric link between sense and either sense or entry, where the first end is anchored to a specific word in the gloss. Publicly these anchor links are displayed as an element in sense gloss that the end-user can click on and be forwarded to the referenced entry. For example anchor link in the verb *zvanīt*$_1$ 'to ring' explanation 'radīt dzidras skaņas (par ZVANU$_1$)' ('to produce clear sounds (referring to a BELL') (anchor link here shown in small capitals) leads to the noun entry *zvans* 'bell'.

Currently we still lack strict guidelines when such references must be used, their use is more or less left to the discretion of the linguist to improve the readability and understanding of the gloss. However, when moving towards an increasingly interconnected Tēzaurs, more and more uses arise for this mechanism.

One of the first things that was added to the public view of Tēzaurs was crude algorithm to deduce where senses are explained with synonyms and to add anchor links there (more in Section 5). However, this algorithm often makes mistakes, therefore, we plan to validate the links generated by it eventually.

Another future aim is to greatly extend the use for such anchor links for any situation whenever an ambiguous word is used in a gloss. Currently, given the size of the Tēzaurs and resources we have (team is small and a wide-coverage word sense disambiguation tool for Latvian has not been made yet to the best of our knowledge), this goal is clearly not obtainable, however nearing to it as much as possible would greatly increase the

usability of dictionary for the end-user. Currently when there is a word that the end-user doesn't understand used in a gloss, they need to copy the word into the dictionary's search bar, find all possible meanings and then guess by themselves what is the most probable meaning. Extensive use of anchor links would reduce such guessing and tell directly, which was the meaning intended by linguist who wrote the gloss.

5 Virtual links

The evolution of digital dictionaries has clearly shown that navigable linking between different elements of a dictionary brings a great gain in usability. However, good linking is a creative process that is best done by a professional lexicographer. Since Tēzaurs was created as a compilation of many printed dictionaries, it has not inherited links within one source, let alone links between entries from different sources. Given that Tēzaurs is a large dictionary with over 403 thousand headwords and over 521 thousand definitions (the Summer 2024 edition), manual enrichment of all entries is hardly conceivable, especially with limited resources. This is also shown by the number of manually created links: 307 definitions have manual anchor links set in this edition (less than 0.06 percent of all definitions), and 3843 entries are enriched with manually selected corpora examples (less than 1 percent of all entries).

In order to still enjoy the benefits of linking, Tēzaurs uses **heuristic linking** for these tasks. The two most important areas where links are created using heuristic methods are the anchor links mentioned above and the corpora examples linked in the entries. These two types differ from other links in that a manually set link is also possible for them.

The general principle for both types of links is that the heuristic (virtual) links are only generated if the entry does not yet have any materialized (manually set) links of the same type. This ensures that the professional opinion of a lexicographer always takes precedence. And in parallel, it also provides a way to correct heuristically incorrectly set links by inserting manual links into the database.

Before a sense definition is displayed to the user, a search is made for possible candidates for **anchor links**. The basis for the candidate status is the syntactical structure of the definition. Safe candidates are stand-alone words that are separated from the rest of the text with separators such as semicolons or the beginning or end of the definition, because such words are usually synonymous, hyponymous or meronymous with the entry word and, not unimportant, are usually in the base form. Other, more specific syntactical patterns are also involved.

For such link candidates, it is checked whether a corresponding target keyword actually exists in the dictionary.

In addition, several heuristic exclusion criteria such as the exclusion of proper names etc. are applied. A further improvement comes from part-of-speech alignment between the source and the target. It would also be possible to use a Universal Dependencies (UD) parser to select the central word in a longer text passage as a linking candidate.

An example of virtual links to other entries can be seen in Fig. 2.

To distinguish heuristic links from manually created ones, in editor's view the automatically created anchor links are displayed in a different color.

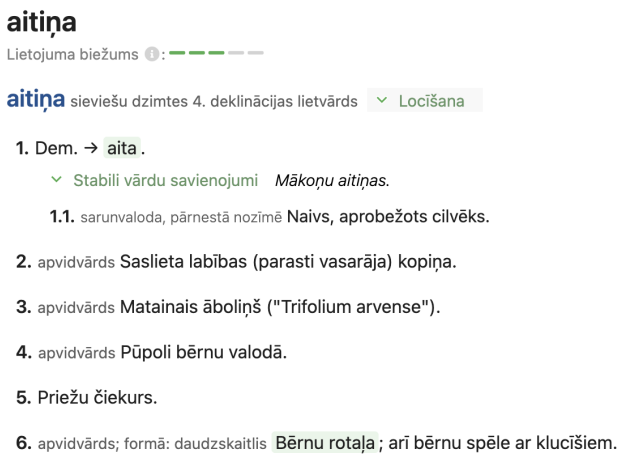

Fig. 2. Example from Tēzaurs: an entry ‘aitiņa’ (a little sheep) with virtual links visible in senses 1 and 6 (parts of the gloss with a greenish background)

The entries without manually selected examples are supplemented with **word usage examples** from 14 different corpora (Saulīte et al., 2022). When selecting the examples, certain corpora are prioritized, and examples of around 10 words in length are preferred. Since the corpora are lemmatized, examples can also be selected where the word is conjugated or declined. A hint text is displayed for automatically selected corpora examples.

The **advantages of virtual linking**: it allows the linking of lexical data sets that would be too labor-intensive to link manually, thus achieving a significant improvement in usability; not storing virtual links allows for easier improvement of the algorithms between dictionary releases.

In addition to the link types already mentioned, other types of automatically created links are also used in Tēzaurs: 1) intra-dictionary links to the same entryword in other dictionaries in the Tēzaurs system (currently: MLVV[2], LLVV[3], MEV[4]), 2) links to the words in alphabetical neighborhood of the entryword, 3) links to similarly spelled words, 4) links to similarly pronounced words.

The **disadvantages of virtual linking**: the link targets for both anchor linking and corpora examples are currently only the entries, as opposed to manually created links that are usually defined at the sense level. It does not seem impossible to calculate automatic links also with the sense granularity using large language models (LLMs), which could be a direction for future research. Another disadvantage of the current implementation may seem that these links cannot be included in the data exports for

[2] Dictionary of Contemporary Latvian Language `https://mlvv.tezaurs.lv`
[3] Dictionary of Latvian Literary Language `https://llvv.tezaurs.lv`
[4] Lettisch-deutsches Wörterbuch by K. Mühlenbachs `https://mev.tezaurs.lv`

use in third-party tools; however, this is only a limited disadvantage because it ensures that only hand-curated data is exported.

6 Conclusions

Tēzaurs as a Latvian language resource is unique specifically due to its multifunctionality, as it serves not only as an explanatory dictionary, but also as the platform for the Latvian WordNet, derivation resource, and interlingual links to Princeton WordNet that also serve as translations for particular word senses. It contains word paradigms, word usage examples for each individual word sense, and lexeme references from inside the text of the sense gloss (the anchor links).

The goal of the Latvian WordNet development is to create the highest quality material possible, therefore, manual methods and data validation are preferred despite them being more time-consuming.

Since Tēzaurs already contains an extensive collection of MWEs, in future we plan to develop it as a more comprehensive MWE resource, adding both semantic and formal annotation to the MWEs. We also want to integrate derivational links into the Latvian WordNet as it has been done for certain other languages.

The ongoing work on all types of links in Tēzaurs is aimed at creating an increasingly large lexical network, where as many words as possible are linked to each other with different links.

Tēzaurs is a resource that is widely used by the general public, including schools; in 2024, from the period from 1 September to 1 October, the resource received 6.1 million queries (i.e., instances of search by its users), and the dictionary has approximately 298,000 active users. Thus, any content published in Tezaurs.lv has a great impact as it is seen and used by a significant number of users. We believe the use of Tezaurs.lv will only expand as it acquires more and more features. In addition, Tēzaurs has already become the central resource used by the Latvian language technologies tools and we expect that it will retain its importance in the future.

7 Acknowledgments

This research is funded by the the Base funding of IMCS in synergy with Latvian Council of Science project “Advancing Latvian computational lexical resources for natural language understanding and generation” (LZP2022/1-0443) and the State Research Programme project “Research on Modern Latvian Language and Development of Language Technology (LATE)” (VPP-LETONIKA-2021/1-0006).

References

Bilgin, O., Çetinoğlu, Ö., Oflazer, K. (2004). Building a wordnet for Turkish, *Romanian Journal of Information Science and Technology* **7**, 163–172.

Fellbaum, C. (1998). *WordNet: An Electronic Lexical Database*, Language, Speech and Communication, Mit Press.
`http://books.google.at/books?id=Rehu8OOzMIMC`

Grasmanis, M., Paikens, P., Pretkalniņa, L., Rituma, L., Strankale, L., Znotins, A., Grūzītis, N. (2023). Tēzaurs.lv – the experience of building a multifunctional lexical resource, *Electronic lexicography in the 21st century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 conference*, Lexical Computing CZ s.r.o., pp. 400–418.
https://elex.link/elex2023/wp-content/uploads/89.pdf

Griffiths, P., Cummins, C. (2017). *An Introduction to English Semantics and Pragmatics*, An Introduction to English Semantics and Pragmatics, Edinburgh University Press.
https://books.google.lv/books?id=N4eqAQAACAAJ

Jurafsky, D., Martin, J. (2024). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Vol. 2, Online manuscript released August 20, 2024. https://web.stanford.edu/ jurafsky/slp3.

Kalnača, A., Lokmane, I. (2021). *Latvian grammar*, University of Latvia Press.

Koeva, S. (2021). The Bulgarian WordNet: Structure and specific features, *Papers of BAS, 8, 1* pp. 47–70.

Levāne-Petrova, K. (2019). Līdzsvarotais mūsdienu latviešu valodas tekstu korpuss, tā nozīme gramatikas pētījumos, *Language: Meaning and Form* **10**, 131–146. The Balanced Corpus of Modern Latvian, its role in grammar studies.
https://www.apgads.lu.lv/fileadmin/user_upload/lu_portal/apgads/PDF/Valoda-nozime-forma/VNF-10/vnf_10-12_Levane_Petrova.pdf

Löbner, S. (2002). *Understanding Semantics*, UK: Hodder Arnold.

Lokmane, I., Rituma, L., Stade, M., Klints, A. (2021). The latvian wordnet and word sense disambiguation: Challenges and findings, *7th Biennial Conference on Electronic Lexicography (eLex)*, pp. 232–246.
https://elex.link/elex2021/wp-content/uploads/2021/08/eLex_2021_13_pp232-246.pdf

Mititelu, V., Leseva, S., Stoyanova, I. (2021). Semantic analysis of verb-noun derivation in Princeton WordNet, *in* Vossen, P., Fellbaum, C. (eds), *Proceedings of the 11th Global Wordnet Conference*, Global Wordnet Association, University of South Africa (UNISA), pp. 108–117.
https://aclanthology.org/2021.gwc-1.13

Paikens, P., Grasmanis, M., Klints, A., Lokmane, I., Pretkalniņa, L., Rituma, L., Stāde, M., Strankale, L. (2022). Towards Latvian WordNet, *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, European Language Resources Association, Marseille, France, pp. 2808–2815.
https://aclanthology.org/2022.lrec-1.300

Paikens, P., Klints, A., Lokmane, I., Pretkalniņa, L., Rituma, L., Stāde, M., Strankale, L. (2023). Latvian WordNet, *12th Global Wordnet Conference*, Global Wordnet Association, pp. 187–196.
https://aclanthology.org/2023.gwc-1.23.pdf

Paikens, P., Pretkalniņa, L., Rituma, L. (2024). A computational model of latvian morphology, *Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING)*, p. 221.
https://aclanthology.org/2024.lrec-main.20

Rambousek, A., Horák, A., Pala, K. (2018). Sustainable long-term wordnet development and maintenance: Case study of the czech wordnet, *Cognitive Studies* **2018**.

Rudnicka, E., Maziarz, M., Piasecki, M., Szpakowicz, S. (2012). A strategy of mapping Polish WordNet onto Princeton WordNet, *in* Kay, M., Boitet, C. (eds), *Proceedings of COLING 2012: Posters*, The COLING 2012 Organizing Committee, Mumbai, India, pp. 1039–1048.
https://aclanthology.org/C12-2101

Saulīte, B., Darģis, R., Grūzītis, N., Auziņa, I., Levāne-Petrova, K., Pretkalniņa, L., Rituma, L., Paikens, P., Znotiņš, A., Strankale, L., Pokratniece, K., Poikāns, I., Bārzdiņš, G., Skadiņa, I., Baklāne, A., Saulespurēns, V., Ziediņš, J. (2022). Latvian National Corpora Collection – Korpuss.lv, *13th Language Resources and Evaluation Conference (LREC)*, pp. 5123–5129. http://www.lrec-conf.org/proceedings/lrec2022/pdf/2022.lrec-1.548.pdf

Soida, E. (2009). *Vārddarināšana*, University of Latvia Press.

Strankale, L., Stāde, M. (2022). Automatic word sense mapping from princeton WordNet to Latvian WordNet, *14th International Conference on Agents and Artificial Intelligence*, Vol. 1, pp. 478–485.

Tufis, D., Barbu, E., Mititelu, V., Ion, R., Bozianu, L. (2004). The romanian wordnet, *Romanian Journal of Information Science and Technology* **7**.

Veidemane, R. (1970). *Latviešu valodas leksiskā sinonīmija*, Zinātne. https://books.google.lv/books?id=M2dgAAAAIAAJ

Received November 27, 2024 , accepted November 29, 2024