Enriching an Explanatory Dictionary with FrameNet and PropBank Corpus Examples

Pēteris Paikens[1], Normunds Grūzītis[2], Laura Rituma[2], Gunta Nešpore[2], Viktors Lipskis[2], Lauma Pretkalniņa[2], Andrejs Spektors[2]

[1] Latvian Information Agency LETA, Marijas street 2, Riga, Latvia
[2] Institute of Mathematics and Computer Science, University of Latvia, Raina blvd. 29, Riga, Latvia
E-mail: peteris.paikens@leta.lv, normunds.gruzitis@ailab.lv

Abstract

This paper describes ongoing work to extend an online dictionary of Latvian – Tezaurs.lv – with representative semantically annotated corpus examples according to the FrameNet and PropBank methodologies and word sense inventories. Tezaurs.lv is one of the largest open lexical resources for Latvian, combining information from more than 300 legacy dictionaries and other sources. The corpus examples are extracted from Latvian FrameNet and PropBank corpora, which are manually annotated parallel subsets of a balanced text corpus of contemporary Latvian. The proposed approach augments traditional lexicographic information with modern cross-lingually interpretable information and enables analysis of word senses from the perspective of frame semantics, which is substantially different from (complementary to) the traditional approach applied in Latvian lexicography. In cases where FrameNet and PropBank corpus evidence aligns well with the word sense split in legacy dictionaries, the frame-semantically annotated corpus examples supplement the word sense information with clarifying usage examples and commonly used semantic valence patterns. However, the annotated corpus examples often provide evidence of a different sense split, which is often more coarse-grained and, thus, may help dictionary users to cluster and comprehend a fine-grained sense split suggested by the legacy sources. This is particularly relevant in case of frequently used polysemous verbs.

Keywords: explanatory dictionary; FrameNet; PropBank; semantic annotation; Latvian

1. Introduction

A major function of an explanatory dictionary is to describe the word senses and illustrate their usage with examples. The separation of word senses is usually done by a lexicographer, based on linguistic intuition and corpus evidence. For less-resourced languages, however, modern corpus-based dictionaries are often missing or works in progress, and the established dictionaries and their senses are not based on corpus evidence. As a consequence, the word sense split is often too fine-grained, which can make it difficult even for a native speaker to grasp the difference, while certain contemporary word senses tend to be missing.

These issues are particularly salient when working on semantic resources for the needs of computational linguistics. Word sense inventories used for automatic word sense disambiguation and semantic parsing tasks need to be formal, well-defined and exhaustive, while the existing dictionaries leave much to the reader's interpretation and rely on illustrative examples of various word usages.

The current work is aimed to extend Tezaurs.lv,[1] the largest Latvian online reference dictionary (Spektors et al., 2016). Tezaurs.lv is structured as an explanatory dictionary which has been compiled from approximately 300 dictionaries and other sources, and contains more than 310,000 entries. In addition to common dictionary content, Tezaurs.lv has been extended with structured data for various natural language processing needs – inflectional paradigm and inflection tables, phonetic transcriptions, domains of usage, stylistic markers and usage restrictions.

Currently the dictionary entries contain usage examples – citations automatically selected from a balanced text corpus of modern Latvian (Levane-Petrova, 2019). These corpus examples tend to illustrate the most common senses and not represent the whole variety of word usage.

However, semantically annotated corpora have sufficient information to separate substantially different uses of the same word, and thus provide examples for each such subsense. In this work we describe the process and results of adding this information to Tezaurs.lv. Section 2 describes the semantically annotated datasets used for this task, Section 3 contains the implementation details, and Section 4 illustrates the resulting changes to the online dictionary.

2. Semantically annotated Latvian corpora

A dataset of semantically annotated Latvian text units is being created within a larger research project "Full Stack of Language Resources for Natural Language Understanding and Generation in Latvian" (Gruzitis et al., 2018b). The goal of that project is to build a balanced multilayer corpus based on cross-lingually oriented syntactic and semantic representations: Universal Dependencies (Nivre et al., 2016), FrameNet (Fillmore et al., 2003), PropBank (Palmer et al., 2005), Abstract Meaning Representation (Banarescu et al., 2013), as well as auxiliary layers for named entity and coreference annotation.[2]

The data is selected to provide a balanced and representative medium-sized corpus of Latvian: around 13,000 sentences annotated at all the above mentioned layers, including FrameNet. To ensure that the corpus is balanced not only in terms of text genres and writing styles but also in terms of lexical units, the text unit of this corpus is an isolated

[1] Open access at www.tezaurs.lv

[2] Available at https://github.com/LUMII-AILab/FullStack.

paragraph. Paragraphs are manually selected from a balanced 10-million-word text corpus (Levane-Petrova, 2019): 60% news, 20% fiction, 7% academic texts, 6% legal texts, 5% spoken language, 2% miscellaneous. The corpus is considered a representative selection of contemporary literary Latvian, including diverse sources starting from year 1991 but excluding translations and genres such as user-generated comments and chat.

The paragraph selection is done with the goal to ensure good coverage for the 1,000 most frequently used Latvian verbs and each of their coarse-grained word senses. We assume that the corpus will prove to be balanced also with respect to nominal lexical units, as the source data is well balanced in terms of genres and frequency distribution. The corpus is not large but has good coverage of the most frequently used verbs, which also tend to be the most ambiguous ones, and there is ongoing work to increase this corpus.

2.1 FrameNet annotations

The annotation of the general-purpose Latvian FrameNet is based on the latest Berkeley FrameNet (BFN) frame inventory (v1.7). The choice to rely on the English BFN frames was made in order to reuse the BFN frame hierarchy and other inter-frame relations, as well as semantic types of frame elements (FE), and the definitions of frames and FEs in general. Another reason for BFN compatibility is to facilitate use cases that require cross-lingual semantic parsing.

In annotating the Latvian FrameNet a concordance approach was followed: frame instances are annotated separately for each target word instead of going through all documents and sentences. Such an approach increases the annotation consistency. In the current annotations only core FEs (which characterize and define the frame) and two non-core FEs (Time and Place) are systematically annotated.

The annotations follow a corpus-driven approach: lexical units in Latvian FrameNet are created only based on the annotated corpus examples. Moreover, the FrameNet annotation is done on the top of the underlying Universal Dependency treebank layer (Pretkalnina et al., 2018), so the annotation of frames and frame elements is thus guided by the dependency structure of a sentence. The currently annotated dataset contains approximately 1,600 distinct lexical units.

2.2 PropBank annotations

The Latvian PropBank corpus is derived from the Latvian FrameNet corpus, thus, this is a parallel dataset. The same original sentences are used, however, the annotations at times are substantially different. The initial draft configuration is automatically generated using the suggested mapping alternatives between English FrameNet and English PropBank. This was followed by linguists mapping the lexical units from

Latvian FrameNet annotation to the semantic frames of English PropBank, and verification of the mapping between FrameNet and PropBank semantic roles, which generally depends on the underlying sentence syntax.

The reason for integrating both FrameNet and PropBank corpus examples into Tezaurs.lv entries is that PropBank tends to provide even more robust and fine-grained sense splits. The semantic roles of the PropBank semantic predicates follow the syntactic argument structure of the target verb, while FrameNet frame elements are often annotated beyond the syntactic argument structure of the target verb. The totality of PropBank annotations for a particular verb essentially constitute a valency dictionary, describing the syntactic relations possible (and used in corpus) for every semantic argument of that verb. Another benefit is that PropBank predicates are lexical compared to the highly abstract FrameNet frames. Therefore both representations are complementary from the Tezaurs.lv user perspective.

A particular source of difficulty is the alignment of Latvian verbs with the English PropBank – unlike some other languages (Haverinen et al., 2015; Xue, 2008), the annotation project chose to use the English PropBank sense inventory instead of native Latvian senses so that the results are immediately aligned and usable for multilingual processing tools. This requires upfront work with translation dictionaries to appropriately map the intended meaning of each verb to its English equivalent. If multiple PropBank verbs match the intended meaning, then extra attention is paid to verb argument structures, however in some cases the choice between multiple options is mostly subjective.

It's worth noting that the sentences are not fully annotated with PropBank roles – only the verbs expressing the FrameNet annotation are targeted, and only the arguments matching the FrameNet core roles are annotated.

3. Technical implementation

For a given lexical entry of Tezaurs.lv, illustrative annotated examples from the Latvian FrameNet and PropBank corpora are selected and visualized as follows.

From the Latvian FrameNet dataset (Section 2.1), we first select all annotation sets where the headword is the target word. Each annotation set represents a single frame, together with its core elements, evoked by the target word. If the same sentence contains more than one frame instance, each instance is encoded in a separate annotation set.

Latvian FrameNet annotation sets are encoded in an extended CoNLL-U format,[3] since

[3] https://universaldependencies.org/format.html

FrameNet annotations are added on top of dependency trees of Latvian UD Treebank[4] (Gruzitis et al., 2018a). The extension follows the CoNLL-2009 format.[5] Figure 1 illustrates an annotation set from the Latvian FrameNet corpus for the sentence "as soon as Sophie had closed [the] garden gate behind her she opened [the] envelope" with the Closure frame evoked by the verb 'to close' (*"aizvērt"*), and its elements (semantic roles) Agent and Container_portal filled by the subject (nsubj – *"Sofija"*/'Sophie') and object (obj – *"vārtiņus"*/'gate') arguments of the verb respectively.

```
# sent_id      = a-d199-p12s1
# text         = Tiklīdz Sofija bija aizvērusi aiz sevis dārza vārtiņus, viņa atvēra aploksni.
# word-by-word = As-soon-as Sophie had closed behind her garden gate , she opened envelope.
1 Tiklīdz     tiklīdz   SCONJ cs                _ 4 mark    _ _ _ _ _          _
2 Sofija      Sofija    PROPN npfsn4           _ 4 nsubj _ _ _ _ _            Agent
3 bija        būt       AUX   vcnisii30an      _ 4 aux     _ _ _ _ _          _
4 aizvērusi   aizvērt   VERB  vmnpdfsnasnpn _ 11 advcl _ _ Y Closure _
5 aiz         aiz       ADP   spsg             _ 6 case    _ _ _ _ _          _
6 sevis       sevis     PRON  px000gn          _ 4 obl     _ _ _ _ _          _
7 dārza       dārzs     NOUN  ncmsg1           _ 8 nmod    _ _ _ _ _          _
8 vārtiņus    vārtiņi   NOUN  ncmpa1          _ 4 obj     _ _ _ _ _          Container_portal
9 ,           ,         PUNCT zc               _ 4 punct _ _ _ _ _           _
10 viņa       viņa      PRON  pp3fsnn          _ 11 nsubj _ _ _ _ _          _
11 atvēra     atvērt    VERB  vmnist130an     _ 0 root    _ _ _ _ _          _
12 aploksni   aploksne  NOUN  ncfsa5          _ 11 obj _ _ _ _ _ _
13 .          .         PUNCT zs               _ 11 punct _ _ _ _ _          _
```

Figure 1: Sample FrameNet annotation set. Fields 1–10 correspond to the CoNLL-U fields: ID, FORM, LEMMA, UPOS, XPOS, FEATS, HEAD, DEPREL, DEPS, MISC; fields 11–13 correspond to the CoNLL-2009 fields: FILLPRED, PRED, APRED. To save space, values of FEATS, DEPS and MISC are excluded from the sample. The word-by-word English translation is added for clarity.

Since the Latvian PropBank corpus is derived from the Latvian FrameNet corpus (Section 2.2), PropBank annotation sets are available as parallel data in the same extended CoNLL-U format (see Figure 2). The initial CONLL-U columns of both datasets are identical, containing the Universal Dependencies (UD) syntactic representation, but the final columns contain the relevant semantic annotation.

For each lexical unit in Latvian FrameNet and Latvian PropBank, there are seven annotation sets on average. To automatically select concise sets of annotated examples to be included in Tezaurs.lv entries of the corresponding verbs, the following selection criteria are applied (in this order):

[4] https://github.com/UniversalDependencies/UD_Latvian-LVTB

[5] http://ufal.mff.cuni.cz/conll2009-st/task-description.html

1. The annotation sets corresponding to each separate frame using this word are selected.
2. If an annotation set is a subset of another annotation set in terms of the evoked frame and its frame elements, it is excluded from the selection, i.e. example sentences representing more frame elements are preferred over example sentences representing less frame elements for the same frame.
3. For each group of so far selected example sentences, shorter examples (containing less tokens) are preferred over longer examples.

```
# sent_id          = a-d199-p12s1
# text             = Tiklīdz Sofija bija aizvērusi aiz sevis dārza vārtiņus, viņa atvēra aploksni.
# word-by-word = As-soon-as Sophie had closed behind her garden gate , she opened envelope.
1 Tiklīdz     tiklīdz    SCONJ cs                  _ 4 mark     _ _ _ _    _
2 Sofija      Sofija     PROPN npfsn4              _ 4 nsubj    _ _ _ _    ARG0-PAG
3 bija        būt        AUX   vcnisii30an         _ 4 aux      _ _ _ _    _
4 aizvērusi   aizvērt    VERB  vmnpdfsnasnpn       _ 11 advcl   _ _ Y close.01 _
5 aiz         aiz        ADP   spsg                _ 6 case     _ _ _ _    _
6 sevis       sevis      PRON  px000gn             _ 4 obl      _ _ _ _    _
7 dārza       dārzs      NOUN  ncmsg1              _ 8 nmod     _ _ _ _    _
8 vārtiņus    vārtiņi    NOUN  ncmpa1              _ 4 obj      _ _ _ _    ARG1-PPT
9 ,           ,          PUNCT zc                  _ 4 punct    _ _ _ _    _
10 viņa       viņa       PRON  pp3fsnn             _ 11 nsubj   _ _ _ _    _
11 atvēra     atvērt     VERB  vmnist130an         _ 0 root     _ _ _ _    _
12 aploksni   aploksne   NOUN  ncfsa5              _ 11 obj     _ _ _ _    _
13 .  .       PUNCT zs   _ 11 punct _ _ _ _ _
```

Figure 2: Sample PropBank annotation set – a complementary semantic annotation to FrameNet (cf. Figure 1).

Additionally, frequency counts are summarized for each lexical unit and are used to sort the selected FrameNet- and PropBank-annotated example sentences (for each Tezaurs.lv entry). In the Tezaurs.lv user interface, the selected annotated examples are visualized using the *brat* JavaScript library[6] (Stenetorp et al., 2012). To generate annotation visualizations in SVG and PNG formats, two kinds of data structures (JSON objects) are generated form the FrameNet- and PropBank-annotated corpus examples.

First, a common stylesheet object is generated (as illustrated in Figure 3) from the FrameNet and PropBank frame inventories, listing all frames (predicates) and frame elements (semantic roles) and their visualization properties. Second, a *brat* annotation object (Figure 4) is generated from the corresponding FrameNet annotation set (Figure 3) for each selected corpus example. Similarly, a *brat* annotation object is generated from the corresponding PropBank annotation set. Note that frame elements (semantic roles) in the Latvian FrameNet and PropBank corpora are added to the root nodes of

[6] http://brat.nlplab.org

the respective syntactic subtree, instead of whole text spans (syntactic phrases). The text spans are calculated while generating the brat annotation objects, based on the dependency links encoded in the underlying UD annotations (the HEAD column in the CoNLL-U data structures; see Figure 1).

```
{
  "entity_types": [{"type": "FE", "bgColor": "yellow", "borderColor": "darken"}],
  "event_types": [
    {...},
    {"type": "Closure", "bgColor": "lightgreen", "borderColor": "darken", "arcs": [
      {"type": "Agent", "color": "blue"},
      {"type": "Time", "color": "blue"},
      {"type": "Place", "color": "blue"},
      {"type": "Containing_object", "color": "blue"},
      {"type": "Result", "color": "blue"},
      {"type": "Container_portal", "color": "blue"}
    ]},
    {...},
    {"type": "Body_movement", "bgColor": "lightgreen", "borderColor": "darken", "arcs": [
      {"type": "Agent", "color": "blue"},
      {"type": "Place", "color": "blue"},
      {"type": "Path", "color": "blue"},
      {"type": "Body_part", "color": "blue"},
      {"type": "Addressee", "color": "blue"}
    ]},
    {...}
  ]
}
```

Figure 3: An incomplete example stylesheet for the FrameNet frames and frame elements. A similar *brat* stylesheet is generated also for PropBank predicates and semantic roles.

```
{
  "text": "Tiklīdz Sofija bija aizvērusi aiz sevis dārza vārtiņus, viņa atvēra aploksni.",
  "triggers": [["T0", "Closure", [[20, 29]]]],
  "events": [["E1", "T0", [["Agent", "T1"], ["Container_portal", "T2"]]]],
  "entities": [
    ["T1", "FE", [[8, 14]]],
    ["T2", "FE", [[40, 54]]]
  ]
}
```

Figure 4: Example sentence with the *brat* annotation, corresponding to the FrameNet annotation given in Figure 1. A similar annotation object is also generated for the corresponding PropBank-annotated corpus example.

Finally, an SVG or a PNG image is generated for each FrameNet and PropBank corpus example (as illustrated in Figure 5) from the common *brat* stylesheet object and the example-specific *brat* annotation objects.

Figure 5: A corpus example ("as soon as Sophie had closed the garden gate behind her [she opened the envelope]") with parallel FrameNet and PropBank annotation, illustrating the sense and use of the headword "aizvērt" ('to close').

4. Enriched online dictionary

The currently intended use case for the FrameNet- and PropBank-annotated corpus examples is to provide separate yet complementary information to the relevant dictionary entries. A set of concise and representative annotated corpus examples is shown to the dictionary user.

Figure 6 illustrates how such frame-semantic information would be displayed in the Tezaurs.lv interface. The original Tezaurs.lv entry contains:

1. the headword: "aizvērt" ('to close');
2. shorthand grammatical information in the Latvian lexicographic tradition, in this particular case showing some key inflectional forms and indicating that the verb is transitive: "-veru, -ver, -ver, pag. ('past') -vēru; trans.";
3. definitions of word senses: (1) "verot aizdarīt" ~ 'to become closed, shut', (2) "verotaizvirzīt aiz kā, kam cauri" ~ 'to move behind something, through something' (the marker "apv." indicates that this sense is used only in some regions);
4. definitions of subsenses: e.g. the first sense has a subsense for closing body parts like eyes and lips – "aizdarīt (acis, plakstus, lūpas, muti)";
5. idioms ("frazeologismi"): collapsed in this example;
6. references to source dictionaries ("avoti");
7. inflection table ("morfologija") automatically provided by a complementary web-service: collapsed in this example;

8. plain-text corpus examples (“korpusa piemēri”) automatically selected by a complementary web-service: it is not certain that the provided corpus examples cover all common senses of the headword, and the examples are selected by the lemma, without explicitly linking them to word senses.

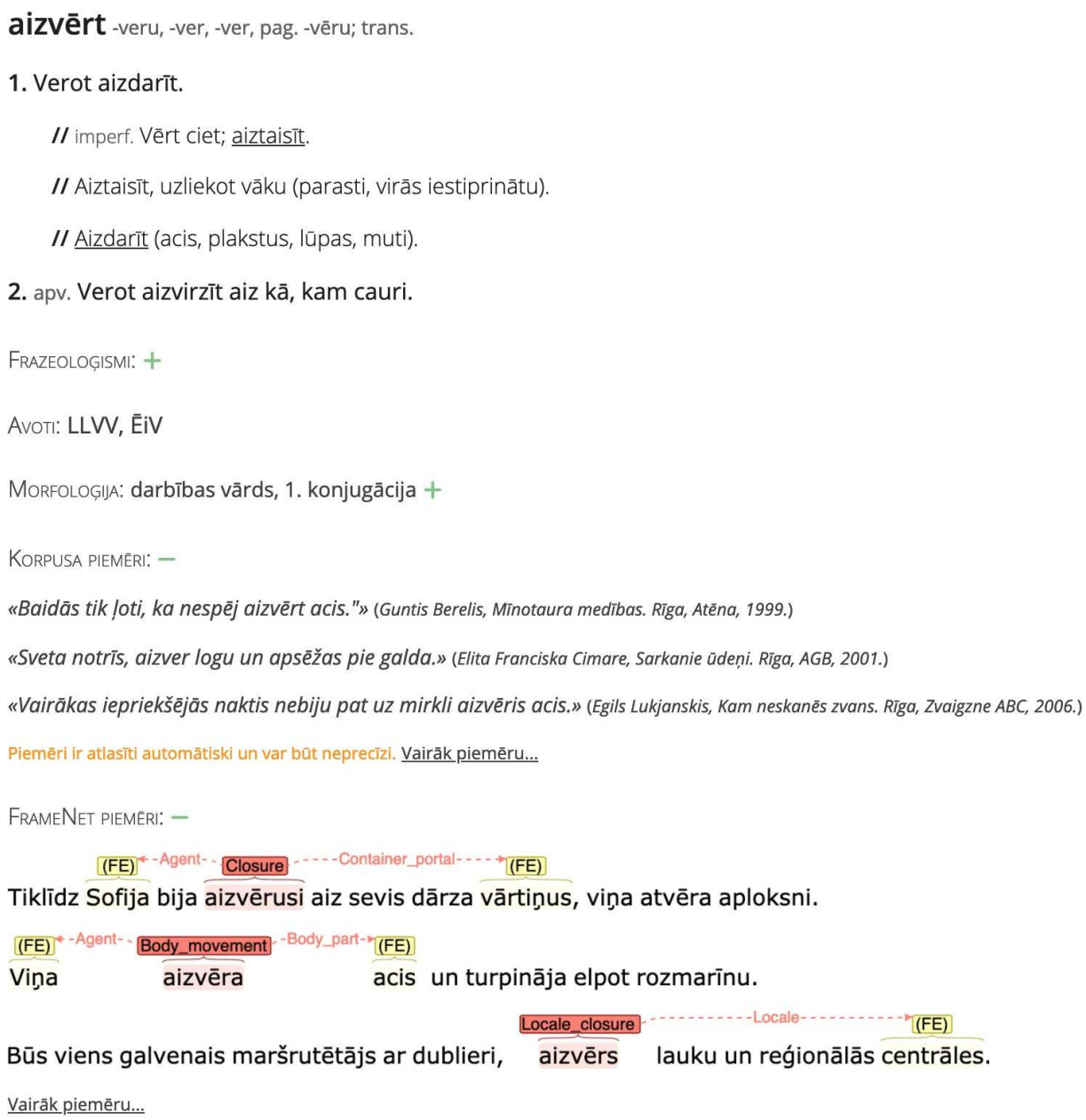

Figure 6: Tezaurs.lv entry for the verb ‘aizvērt’ (to close): https://tezaurs.lv/#/sv/aizvērt. The original entry, consolidated from two source dictionaries (LLVV and EiV), is enriched with automatically extracted usage examples from (i) a balanced text corpus (‘Korpusa piemēri’), and (ii) a FrameNet-annotated corpus (‘FrameNet piemēri’). FrameNet annotations can be switched to the parallel PropBank annotations.

In the supplementary section of FrameNet corpus examples (“FrameNet piemēri”), a concise annotated example is given for each of the different semantic frames evoked by the headword to illustrate its sense split and semantic valency according to FrameNet.

In the above example, two of the FrameNet frames – Closure and Body_movement – align with the first sense (and its third subsense) of the headword, and it is debatable whether Body_movement is a subsense of Closure or not (for this particular verb). However, the third FrameNet example which evokes Locale_closure, illustrates a distinct meaning of the verb *'aizvērt'*, which is missing in the original Tezaurs.lv entry despite being a commonly used word sense for already a long time. Also note that the second word sense provided by Tezaurs.lv is rare and possibly obsolete, and therefore is not represented in the balanced FrameNet-annotated corpus.

5. Conclusions and future work

In summary, we propose to extend online dictionaries by adding frame-semantically annotated corpus examples. Such examples enable complementary analysis of word senses and word valence patterns from the perspective of frame semantics, which is substantially different from the traditional lexicographic approach.

In our opinion, the major benefit of the suggested approach for everyday dictionary users is the following: it often provides an alternative and more coarse-grained split of word senses based on semantically annotated corpus evidence according to the robust FrameNet and PropBank methodologies.

Since Latvian FrameNet uses the abstract frame inventory of Berkeley FrameNet and the more concrete semantic predicate inventory of English PropBank, it also makes it easier for language learners to understand the differences between particular word senses, assuming that they know English better than Latvian.

Another benefit is the modernization of legacy dictionaries. A large portion of Tezaurs.lv entries and word sense splits originate from Latvian dictionaries of 1970s, but the semantically annotated corpus represents contemporary usage of the language. Because of this, corpus examples illustrate usage and sense split of words in more contemporary contexts, some of which were not identified in the earlier dictionaries.

The differences in sense splitting between legacy dictionaries and examples from the semantically annotated corpora illustrate the need for future work on updating the Latvian word sense inventory based on corpus evidence, either as part of the traditional lexicographic workflow or as a separate lexical resource in the likeness of WordNet (Miller, 1995; Bond & Foster, 2013).

Another direction of future work is the handling of multi-word expressions (MWEs) such as phrasal verbs. For example, the verb *'aiziet'* ('to go away') has distinct senses invoked by *'aiziet bojā'* ('to perish'), *'aiziet mūžībā'* ('to die'). Such MWEs are explicitly annotated in the Latvian FrameNet dataset, but are currently not included in the CoNLL-style output format and, thus, are not included in the FrameNet example visualizations.

6. Acknowledgements

This work has received financial support from the European Regional Development Fund under grant agreement No 1.1.1.1/16/A/219 and from the Latvian State research program "Latvian language" (VPP-IZM-2018/2-0002).

7. References

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M. & Schneider, N. (2013). Abstract Meaning Representation for Sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*. Sofia, Bulgaria, pp. 178–186.

Bond, F. & Foster, R. (2013). Linking and Extending an Open Multilingual Wordnet. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Sofia, Bulgaria: Association for Computational Linguistics, pp. 1352–1362. https://www.aclweb.org/anthology/P13-1133.

Fillmore, C. J., Johnson, C. R. & Petruck, M. R. (2003). Background to FrameNet. *International Journal of Lexicography*, 16(3), pp. 235–250.

Gruzitis, N., Nespore-Berzkalne, G. & Saulite, B. (2018a). Creation of Latvian FrameNet based on Universal Dependencies. In *Proceedings of the International FrameNet Workshop (IFNW)*. Miyazaki, Japan, pp. 23–27.

Gruzitis, N., Pretkalnina, L., Saulite, B., Rituma, L., Nespore-Berzkalne, G., Znotins, A. & Paikens, P. (2018b). Creation of a Balanced State-of-the-Art Multilayer Corpus for NLU. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC)*. Miyazaki, Japan, pp. 4506–4513. http://www.lrec-conf.org/proceedings/lrec2018/pdf/935.pdf.

Haverinen, K., Kanerva, J., Kohonen, S., Missilä, A., Ojala, S., Viljanen, T., Laippala, V. & Ginter, F. (2015). The Finnish Proposition Bank. *Language Resources and Evaluation*, 49(4), pp. 907–926. https://doi.org/10.1007/s10579-015-9310-y.

Levane-Petrova, K. (2019). LVK2018: Līdzsvarotais mu¯sdienu latviešu valodas tekstu korpuss, tā nozīme gramatikas pētījumos. *Language: Meaning and Form*, 10.

Miller, G.A. (1995). WordNet: A Lexical Database for English. *Commun. ACM*, 38(11), pp. 39–41. URL http://doi.acm.org/10.1145/219717.219748.

Nivre, J., de Marneffe, M. C., Ginter, F., Goldberg, Y., Hajič, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R. & Zeman, D. (2016). Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*. pp. 1659–1666.

Palmer, M., Gildea, D. & Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1), pp. 71–106.

Pretkalnina, L., Rituma, L. & Saulite, B. (2018). Deriving Enhanced Universal Dependencies from a Hybrid Dependency-Constituency Treebank. In *Text,*

Speech, and Dialogue, volume 11107. Springer, pp. 95–105. https://www.researchgate.net/publication/327520269.

Spektors, A., Auzina, I., Dargis, R., Gruzitis, N., Paikens, P., Pretkalnina, L., Rituma, L. & Saulite, B. (2016). Tezaurs.lv: the largest open lexical database for Latvian. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*. Portoroz, Slovenia, pp. 2568–2571.

Stenetorp, P., Pyysalo, S., Topić, G., Ohta, T., Ananiadou, S. & Tsujii, J. (2012). brat: a Web-based Tool for NLP-Assisted Text Annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*. Avignon, France: Association for Computational Linguistics, pp. 102–107. https://www.aclweb.org/anthology/E12-2021.

Xue, N. (2008). Labeling Chinese Predicates with Semantic Roles. *Comput. Linguist.*, 34(2), pp. 225–255. http://dx.doi.org/10.1162/coli.2008.34.2.225.