*Proceedings of the International FrameNet Workshop 2020: Towards a Global, Multilingual FrameNet*, pages 63–69
Language Resources and Evaluation Conference (LREC 2020), Marseille, 11–16 May 2020


# Deriving a PropBank Corpus from Parallel FrameNet and UD Corpora

**Normunds Gruzitis, Roberts Dargis, Laura Rituma, Gunta Nespore-Berzkalne, Baiba Saulite**
Institute of Mathematics and Computer Science, University of Latvia
Raina bulv. 29, LV-1459, Riga, Latvia
{normunds.gruzitis, roberts.dargis, laura.rituma, gunta.nespore, baiba.valkovska}@lumii.lv

**Abstract**

We propose an approach for generating an accurate and consistent PropBank-annotated corpus, given a FrameNet-annotated corpus which has an underlying dependency annotation layer, namely, a parallel Universal Dependencies (UD) treebank. The PropBank annotation layer of such a multi-layer corpus can be semi-automatically derived from the existing FrameNet and UD annotation layers, by providing a mapping configuration from lexical units in [a non-English language] FrameNet to [English language] PropBank predicates, and a mapping configuration from FrameNet frame elements to PropBank semantic arguments for the given pair of a FrameNet frame and a PropBank predicate. The latter mapping generally depends on the underlying UD syntactic relations. To demonstrate our approach, we use Latvian FrameNet, annotated on top of Latvian UD Treebank, for generating Latvian PropBank in compliance with the Universal Propositions approach.

**Keywords:** PropBank, FrameNet, Universal Dependencies, Universal Propositions, Latvian

## 1. Introduction and Related Work

Proposition Bank or PropBank (Palmer et al., 2005) is (i) a shallow semantic representation for the annotation of predicate-argument structures, (ii) a lexicon of English verbs and their semantic predicates (frames) and semantic arguments (roles), and (iii) a large annotated text corpus of English, where the semantic roles of each predicate instance are added to the syntactic structures of the underlying treebank.

Since PropBank uses a small set of semantic roles which are defined on a verb-by-verb basis, and the annotated corpus provides broad-coverage training data, it is an attractive approach for robust automatic semantic role labelling, SRL (Cai and Lapata, 2019). This has also encouraged extensive use of PropBank framesets (coarse-grained verb senses each having a specific set of semantic arguments or roles) in the whole-sentence Abstract Meaning Representation (AMR) approach (Banarescu et al., 2013).

Following the work on English, PropBank-style corpora have been created for a number of languages. Apart from other aspects, creation of a propbank depends on a fundamental decision: whether to define language-specific framesets or to re-use the English PropBank framesets.

In most projects, language-specific framesets have been defined and used in the manual or semi-automatic corpus annotation workflow, e.g. for Chinese (Xue, 2008), Hindi/Urdu (Bhatt et al., 2009) and Finnish (Haverinen et al., 2015). Few attempts have been made to create a non-English propbank by reusing the English PropBank framesets. An example to the latter approach is Brazilian Portuguese PropBank (Duran and Aluisio, 2012), although the use of English framesets was intended only as an intermediate step on the way to define language-specific framesets.

Another consideration is the underlying syntactic representation – syntactic structures to which the semantic roles are added. In the case of phrase structure trees (e.g. the English and Chinese treebanks), semantic roles are added to constituents (phrases). In the case of dependency trees (e.g. the Finnish treebank), semantic roles are added to dependencies (syntactic roles of the root tokens of the respective subtrees). For some languages (e.g. Hindi/Urdu and Brazilian Portuguese) both kinds of syntactic representations and both kinds of PropBank-treebank mappings are available.

While dependency trees are often considered a more convenient and straightforward intermediate representation for robust automatic SRL, as it has been proved by state-of-the-art SRL parsers (Cai and Lapata, 2019), the use of a common inventory of PropBank framesets would facilitate cross-lingual SRL and the downstream applications like cross-lingual information extraction.

The Universal Propositions (UP) project[1] proposes to use the English PropBank framesets for universal SRL, on top of the Universal Dependencies (UD) syntax trees. The underlying UD representation (Nivre et al., 2016) facilitates cross-lingual semantic parsing even more.

Akbik et al. (2015) present a method for automatic projection of English framesets to a target language, and they have applied this method to generate UP propbanks for multiple languages. In this paper, we present our work which contributes to the UP initiative. We propose an alternative approach for generating accurate and consistent UP propbanks for languages that have a FrameNet-annotated corpus where FrameNet annotations are specified on top of a UD treebank, or a dependency treebank in general.

To some extent, our approach is similar to the one applied to convert the SALSA Corpus for German into a PropBank-like corpus for the CoNLL 2009 shared task (Hajič et al., 2009). The SALSA corpus (Burchardt et al., 2006) uses semantic roles in the FrameNet paradigm (Fillmore et al., 2003), annotated on top of a treebank, which were semi-automatically converted to the respective PropBank arguments. The semantic predicates, however, remain German-specific in the converted SALSA corpus. In contrast, we reuse semantic predicates from the English PropBank (following the UP approach), which was the most challenging part in the Latvian FrameNet-to-PropBank conversion. The

[1] https://github.com/System-T/UniversalPropositions

| LEMMA | UPOS | $PRED_{FN}$ | $PRED_{PB}$ |
|---|---|---|---|
| mācīt | VERB | Education_teaching | teach.01 |
| mācīties | VERB | Education_teaching | study.01 |
| mācīties | VERB | Memorization | learn.01 |
| dzīvot | VERB | Residence | reside.01 |
| dzīvot | VERB | Dead_or_alive | live.01 |
| dzīvot | VERB | Living_conditions | live.02 |

Table 1: Sample mapping from lexical units (verb-frame pairs) in Latvian FrameNet (FN) to English PropBank (PB) predicates (verb sense-specific translation equivalents).

consecutive conversion of FrameNet roles into PropBank roles is rather straightforward, although it depends on the underlying UD roles.

On the one hand, FrameNet defines a set of more abstract semantic frames (compared to PropBank predicates) that can be evoked by different target words. On the other hand, FrameNet uses more fine-grained semantic roles (frame elements), some of which are often not expressed in a sentence as direct syntactic arguments of the predicate. Therefore our proposed FrameNet-to-PropBank conversion approach is unidirectional, i.e., a rather complete PropBank corpus can be derived from an existing FrameNet corpus (with parallel dependency annotations), however, it would not be possible to derive a complete FrameNet corpus from an existing PropBank corpus without additional annotation work.

To demonstrate our approach, we use Latvian UD Treebank (Gruzitis et al., 2018b) and Latvian FrameNet (Gruzitis et al., 2018a) for generating Latvian PropBank, compliant to the Universal Propositions approach.

## 2. Mapping Configuration

Semantic roles in PropBank are much more robust compared to FrameNet frame elements, and the overall PropBank annotation systematically follows the syntactic verb-argument structure. Therefore the PropBank layer of such a multi-layer text corpus can be semi-automatically derived from the existing FrameNet and UD layers of the corpus, by providing (i) a mapping configuration from lexical units (LU) in [a non-English language] FrameNet to [English language] PropBank predicates (see Table 1), and (ii) a mapping configuration from FrameNet frame elements to PropBank semantic arguments for the given pair of a FrameNet frame and a PropBank predicate, i.e., independently from LUs (see Table 2).

We are building on the previous work on SemLink (Palmer, 2009) and Predicate Matrix (Lopez de Lacalle et al., 2016), although none of the two data sets provide complete mapping suggestions, especially for less frequently used lexical units, since the suggestions are corpus-driven. We use the suggested mapping alternatives between English FrameNet and English PropBank as a draft configuration. The manual task for a linguist is to map the LUs from Latvian FrameNet to the semantic predicates of English PropBank, and to verify the mapping between FrameNet frame elements (FE) and PropBank semantic roles, which generally depends on the underlying syntactic relations. The successive generation of a PropBank annotation layer is a straightforward automation.

| $PRED_{FN}$ | $APRED_{FN}$ | DEP | $PRED_{PB}$ | $APRED_{PB}$ |
|---|---|---|---|---|
| Education_teaching | Student | nsubj | study.01 | A0 |
| Education_teaching | Student | obj | teach.01 | A2 |
| Education_teaching | Student | iobj | teach.01 | A2 |
| Education_teaching | Subject | obj | study.01 | A1 |
| Education_teaching | Subject | obj | teach.01 | A1 |
| Education_teaching | Teacher | obl | study.01 | A2 |
| Education_teaching | Teacher | nsubj | teach.01 | A0 |
| Education_teaching | Institution | obl | study.01 | AM-LOC |
| Education_teaching | Institution | obl | teach.01 | AM-LOC |
| Education_teaching | Level | obl | study.01 | AM-LOC |
| Education_teaching | Time | obl | study.01 | AM-TMP |
| Education_teaching | Time | obl | teach.01 | AM-TMP |

Table 2: Mapping from FrameNet (FN) frame elements to PropBank (PB) semantic roles, taking UD dependency relations (syntactic roles) into account.

Since the FrameNet annotation is semantically richer, and it can be non-projective w.r.t. the underlying dependency tree, some FrameNet frame elements are not transferred to the PropBank layer, if they are not syntactic arguments of the target verb.

To ensure productive work on defining the language-specific mapping configuration (Latvian FrameNet to English PropBank via Latvian UD Treebank), we have developed a convenient and predictive user interface that exploits a simple but efficient method for sorting candidate suggestions for LU-to-predicate mapping (Section 2.1) and for FE-to-argument mapping (Section 2.2). Note that both kinds of mapping are done on the type level, i.e., no individual occurrences are mapped. Affected corpus examples, however, are dynamically selected and displayed, which helps the annotator to verify the choices made.

### 2.1. LU-to-predicate mapping

Figure 1 partially illustrates the interface for mapping FrameNet lexical units (verb-frame pairs) to the corresponding PropBank predicates.

In total, there are nearly 11,000 English PropBank framesets, therefore an efficient method to narrow down the LU-to-predicate mapping candidates is necessary.

Mapping suggestions are extracted from two existing data sets. First, the SemLink data set was parsed to extract suggested FrameNet frame candidates (if any) for each PropBank predicate. Second, additional mapping alternatives between FrameNet frames and PropBank predicates were similarly extracted from the Predicate Matrix data set. Overall, the two data sets provide suggestions for about 90% of Berkeley FrameNet frames reused in Latvian FrameNet. Although the ultimate mapping must be provided between the language-specific LUs (verb-frame pairs) and the PropBank predicates, not just between FrameNet frames and PropBank predicates, the candidate predicates are proposed based on the FrameNet frame.

In addition to SemLink and Predicate Matrix, we also used

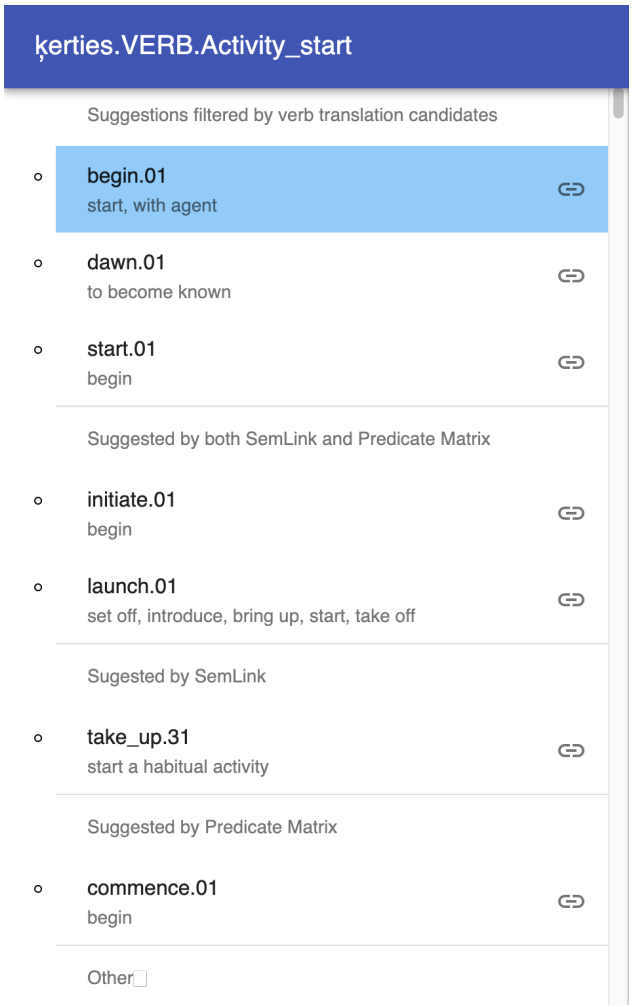


Figure 1: User interface for mapping a LU in Latvian FrameNet (the verb 'ķerties' in the *Activity_start* sense) to a PropBank predicate (*begin.01*). Candidate predicates are sorted depending on the sources that have suggested this candidate: translation candidates are the most probable, followed by suggestions by SemLink and Predicate Matrix (if not already proposed as translation candidates).

a state-of-the-art Latvian-English machine translation (MT) system[2] to acquire translation candidates for Latvian verbs (lexical units in Latvian FrameNet). Translation candidates that correspond to PropBank predicates or their aliases (alternative lexical units, or target verbs, suggested by Predicate Matrix) are used to reorder the final suggestions. As it has turned out, the MT-supported suggestions are the most useful and accurate ones.

To make the choice for a linguist easier, the PropBank predicate suggestions are split in the user interface into four priority groups (see Figure 1). Each lower priority group contains suggestions that are not contained by any of the higher priority groups.

The first group contains LU-to-predicate suggestions that are supported by Latvian-English verb translation candidates. This group is provided for 61% of all LUs in Latvian FrameNet, suggesting 2.6 PropBank predicates per LU on average. When this group of predicate candidates is available, a mapping suggestion proposed by this group is accurate in 79% of the cases.

The second group contains suggestions which are an intersection of suggestions proposed by both SemLink and Predicate Matrix. Although this group is provided for 71% of all LUs in Latvian FrameNet, it is the first group of suggestions only in 23% of the cases (when no MT-supported suggestions are available). On average, there are 5 suggestions per LU contained in this group. When this group is the first available one, a suggestion proposed by this group is accurate in 27% of the cases.

The remaining groups contain suggestions that are supported only by SemLink or by Predicate Matrix. One of these two groups is the first priority group only for 7% of all LUs.

Note that for 10% of all LUs, the ultimately selected PropBank predicate was not present in any of the suggestion groups, and the linguist had to find an appropriate predicate on its own.

Also note that each FrameNet frame has 15 PropBank predicate suggestions on average due to the highly abstract FrameNet frames that each can be evoked by different target verbs. Consequently, for all LUs of the same frame, the same 15 candidates (on average) are suggested for PropBank predicate mapping, except that these candidates are grouped differently, based on translation candidates of the target verb. For example, the FrameNet frame *Body_movement* can be evoked by many target verbs, and therefore it has 70 PropBank predicate suggestions, such as *clap.01*, *close.01*, *kneel.01*, *nod.01*, and *wave.01*. However, if we consider, for instance, the LU *aizvērt.VERB.Body_movement*, the predicate *close.01* is the top suggestion, while for *pamāt.VERB.Body_movement* the top suggestion is *wave.01*.

## 2.2. FE-to-argument mapping

When a lexical unit (LU) is mapped between FrameNet and PropBank at the frame-predicate level, the next step is to map FrameNet frame elements (FE) to PropBank semantic arguments.

Figures 2 and 3 partially illustrate the interface for mapping FrameNet frame elements to the corresponding PropBank arguments, depending on the underlying syntactic relations.

In case of FE-to-argument mapping, we consult only the Predicate Matrix data set (in addition to the PropBank data set itself) to extract and group FE-to-argument mapping suggestions, since Predicate Matrix is a more recent data set, and it provides mapping suggestions directly between FrameNet and PropBank, instead of the indirect SemLink mappings via VerbNet (Schuler et al., 2000).

For each PropBank frameset, core and non-core arguments are extracted and grouped separately. The group of core arguments is prioritised over the group of non-core arguments.

Suggestions supported by Predicate Matrix are separated in the highest priority group. Such priority suggestions are available for 51% of the required FE-to-argument mappings, typically containing only one suggestion. If this group is present, it always contains the accurate mapping

[2] https://hugo.lv

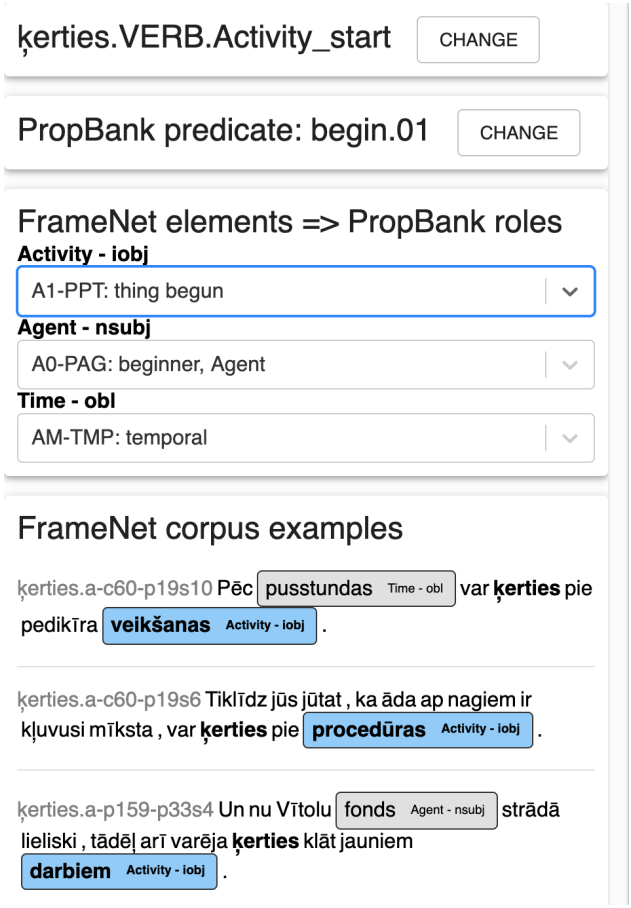


Figure 2: User interface for mapping FrameNet frame elements to PropBank semantic arguments for the given LU (*ķerties.VERB.Activity_start*) / predicate (*begin.01*). In general, the choice depends on the underlying UD relation.

candidate, and a linguist had to make a choice between two or three candidates only in 10% of the cases; the rest of the cases were unambiguous.

In 49% of all cases when no Predicate Matrix suggestions are available for a given FE, a linguist first considers the group of PropBank core arguments defined for the particular frameset. This group contains 3 candidates on average. Note that Latvian FrameNet mostly contains annotations of core FEs; non-core FEs are annotated rarely (for the time being). Therefore one of these suggestions is typically the correct one, and it is easy to make the choice.

The remaining group of non-core PropBank arguments (the *ArgM* roles) is always present, containing all the possible *ArgM* options regardless the particular frameset (except if already included in the priority group of Predicate Matrix suggestions). However, this group of suggestions is seldom consulted by a linguist because of the above mentioned nature of Latvian FrameNet.

## 2.3. Elimination of FrameNet and UD errors

A very important side result of the FrameNet to PropBank mapping process is that it has unveiled a number of annotation errors and inconsistencies both in Latvian FrameNet and in Latvian UD Treebank.

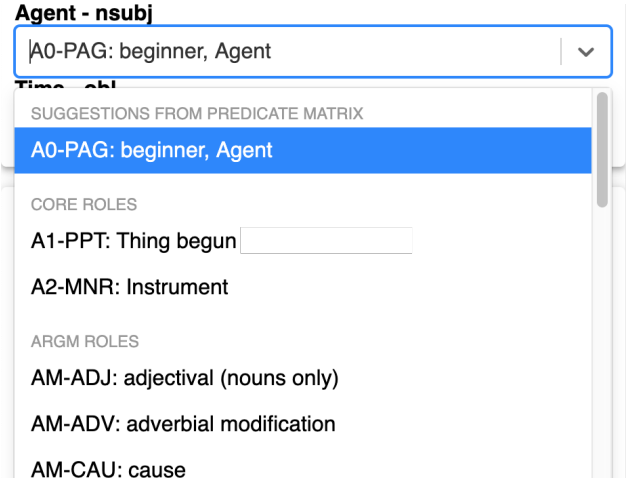


Figure 3: Candidate PropBank argument mappings for the selected pair of a FrameNet frame element (*Agent*) and a UD dependency relation (*nsubj*).

As Figure 2 illustrates, the linguist who verifies the mapping configuration also sees all corpus examples for the given LU. This not only helps to make decisions in both LU-to-predicate mapping and FE-to-argument mapping, but also helps to notice inconsistencies and errors in the underlying annotation layers. Such sentences can be marked with a *FixMe* tag, indicating the annotation layer and the type of the issue.

We have identified three types of typical issues so far:

- An incorrect UD relation associated with a FrameNet frame element, which means that most likely there is an error in the UD annotation layer. The FrameNet-to-PropBank mapping user interface allows to filter all corpus examples (along with their sentence identifiers) containing this error. The mapping configuration of LUs containing such issues is left unfinished until the issues are fixed and the mapping can be finalised.

- An incorrect root node of a subtree of the underlying UD tree is selected for a FrameNet frame element. It can also be the case that the whole FrameNet frame is chosen incorrectly for a particular sentence, and the PropBank perspective has helped to notice that. Again, the user interface allows to filter the problematic corpus examples, and the mapping configuration for the particular LUs is left unfinished until the issues are fixed.

- The mapping process also encourages to reconsider the whole LU – whether the selected FrameNet frame is best suited for the particular verb sense. For instance, we have observed that different verbs are annotated in Latvian FrameNet using the *Give_impression* frame, however, SemLink suggests the *Appearance* frame for the respective PropBank predicates. This helps to achieve a better consistency for both Latvian FrameNet and Latvian PropBank.

## 3. Results

The Latvian FrameNet to PropBank mapping process is nearly finished: we have so far specified mappings for 92% of the LUs in the latest version of Latvian FrameNet. A corresponding Latvian PropBank corpus is automatically derived, and all the annotation layers of the multi-layer corpus are released as open data.[3]

### 3.1. Data set

Current statistics of the parallel Latvian FrameNet and Latvian PropBank corpora is as follows:

- **Lexical units** For 2,377 (out of 2,577) LUs represented in Latvian FrameNet, a mapping configuration to PropBank has been specified (92.2%). These LUs represent word senses of 1,322 (out of 1,358) frequently used verbs represented in Latvian FrameNet (97.3%).
- **FrameNet frames** For 521 (out of 540) Berkeley FrameNet frames reused in Latvian FrameNet, at least one LU has been mapped to PropBank so far (96.5%). Latvian FrameNet, in turn, covers 44.2% of 1,222 frames defined in Berkeley FrameNet v1.7.
- **PropBank predicates** Current LU-to-predicate mappings cover 1,033 (out of 10,687) English PropBank v3.1 predicates (9.7%).
- **Corpus examples** The LU-to-predicate and FE-to-argument mappings specified so far cover 20,054 (out of 20,879) annotation sets in Latvian FrameNet (96.0%).

Latvian PropBank consists of two data sets: (i) a machine-readable mapping configuration for each LU in Latvian FrameNet, and (ii) a set of annotated corpus examples in an extended CoNLL-U format, compliant to Universal Propositions. Latvian FrameNet is a single data set of annotated corpus examples in an extended CoNLL-U format.

### 3.2. Inter-annotator agreement

To conduct an inter-annotator agreement (IAA) experiment, we selected 30 random LUs from Latvian FrameNet to be mapped to PropBank by three linguists experienced in treebanking as well as in frame semantics. The 30 LUs cover 205 corpus examples (annotation sets). First, we measured IAA w.r.t. LU-to-predicate mapping, then – w.r.t. FE-to-argument mapping.

**LU-to-predicate** Statistically, only in 13 cases out of 30 (43.3%) all 3 annotators have agreed on the corresponding PropBank predicate for a given LU. In 13 more cases, at least 2 annotators agreed on the same predicate, thus, at least 2 of 3 annotators could agree on a predicate in 86.6% of the cases. In the remaining 4 cases, no two annotators could agree on the same predicate. The qualitative analysis, however, shows that the cause of disagreement was mostly due to different preferences when deciding between close translation equivalents (having an equivalent argument structure). For instance, the LU *dauzīties.VERB.Impact* was mapped to three different predicates: *beat.02*, *bang.02* and *thud.01*. In general, all three predicates represent a situation when something hits something making a sound. However, each alternative has a slightly different meaning. Another example – the LU *noslaucīt.VERB.Emptying* – first annotator has selected a rather abstract predicate (*clear.01*), second annotator – a more specific predicate (*wipe.01*), while third – even a more specific one (*wipe-off.03*). These differences illustrate that the annotator's sense of the second language plays an important role.

**FE-to-argument** Given that annotators have agreed on a predicate, the mapping of FrameNet frame elements to PropBank semantic arguments is straightforward. Our IAA experiment shows that annotators can agree in 95.2% of the cases. The remaining 4.8% are cases where at least one annotator has tagged the given FrameNet frame element or UD dependency relation as an annotation error to be fixed in the FrameNet or UD layer respectively.

## 4. Discussion

This section summarises discussion of linguistic issues regarding LU-to-predicate and FE-to-argument mapping.

### 4.1. LU-to-predicate mapping

If there are several predicates with similar meaning in PropBank, it is not always clear which of them should be chosen. If we consider, for instance, the LU *parādīties.VERB.Circumscribed_existence*, its meaning roughly corresponds to PropBank predicates *appear.01*, *show_up.02* and *emerge.02*. In such cases, we choose the predicate with an argument structure that best matches the argument structure of the Latvian verb, i.e., the predicate that covers as many core FEs of the corresponding FrameNet frame as possible.

In some cases, more than one PropBank predicate corresponds to the meaning of a LU in Latvian FrameNet – the meaning of the FrameNet frame is more general than the meaning of the candidate PropBank predicates. For instance, the LU *izvirzīt.VERB.Choosing* covers corpus examples *izvirzīt mērķi* ('to set a goal') and *izvirzīt kandidātu* ('to nominate a candidate'), but PropBank does not provide a predicate that covers both meanings. In such cases, we consider the possibility of making the sense split at the FrameNet layer, if possible, by applying different FrameNet frames to represent these differences. Another example: the LU *slēgt.VERB.Closure*. The FrameNet frame *Closure* covers LUs of both meanings: opening and closing something. In PropBank, there are different predicates for each of the two meanings. In Latvian, however, these both meanings can be expressed by the same verb, using different adverbial modifiers: *slēgt ciet* ('to close') and *slēgt vaļā* ('to open'). We do not have a good solution for this issue yet, although such cases are quite rare.

There are some cases when a LU does not have an appropriate PropBank predicate for mapping, and would require a constructicon kind of a sense inventory. For instance, the LU *klusēt.VERB.Volubility* is expressed in English as the predicate adjective construction 'to be/keep silent'.

[3] https://github.com/LUMII-AILab/FullStack

A related issue are several Latvian verbs with modal meaning, which are not considered as modal verbs. In English PropBank, modal verbs like *must* and *can* are not annotated as predicates, therefore we cannot select a verbal predicate for a Latvian verb with such a meaning. However, we can choose an adjectival predicate, for instance, *able.01* or *unable.01*.

For around 25% of LUs in Latvian FrameNet, it was challenging to select the corresponding semantic predicate from English PropBank. In such cases, it took up to 1 hour for a linguist to decide the best fitting mapping, sometimes resulting in no mapping at all (see Section 3.1). In the remaining 75% of cases, it took up to 5 minutes for a linguist to decide the mapping. Overall, it took around 1 person month (PM) to map the easy cases (around 1,933) and 4 PMs to map the difficult cases (around 644).

### 4.2. FE-to-argument mapping

There are cases when it is impossible to assign a PropBank argument to a core FE of a FrameNet frame:

- In a syntax tree, the potential argument of a PropBank predicate is not a syntactic argument of this predicate. For example, consider the sentence *ļauj man paskatīties* 'let me look': the argument *ARG0* of the predicate *look.01* semantically is *man* ('me'), but syntactically this is an argument of the verb *ļaut* ('to let').

- Similarly, there are cases when a syntax subtree with a verb as its root node depends on another part of the sentence which represents a semantic argument of the verb but is not its syntactic argument. Consider, for instance, the sentence *kā pastāstīja organizācija, nebija iespējams lietot elektrības ģeneratoru* ('as it was told by the organization, it was impossible to use the power generator'). The verb *pastāstīt* ('to tell') corresponds to the PropBank predicate *tell.01* that has the argument *ARG1: utterance*, but the utterance itself is represented by the root node of the whole syntax tree on which the instance of *tell.01* depends.

- A core FE of a FrameNet frame is not defined as a core argument of the corresponding PropBank predicate. A typical example is the frame *Change_position_on_a_scale*: in FrameNet, there are two core FEs – *Item* (the entity that has a position on the scale) and *Attribute* (a scalar property that the *Item* possesses) – that both correspond to one argument of a corresponding PropBank predicate. Consequently, the FE *Item* is not mapped to an argument, if both *Item* and *Attribute* are present in the sentence.

The time spent to provide mapping at the semantic role level is included in the estimated time spent to provide mapping from lexical units in Latvian FrameNet to English PropBank predicates (see Section 4.1).

## 5. Conclusion

We have demonstrated in practice that a quality PropBank-compliant lexical database and annotated text corpus can be consistently and rapidly derived from an existing multilayer corpus that contains both FrameNet and UD annotation layers (or equivalent annotation layers). While mapping lexical units from a non-English FrameNet to English PropBank predicates is often (around 25% cases) a linguistically challenging task, the mapping at the semantic role level is straightforward, although it depends on the syntactic roles in general. Note that neither SemLink nor Predicate Matrix mappings contain information about the corresponding syntactic roles. This kind of information is created in our approach, and it could be added to these resources.

Although it is often the case that a PropBank corpus is created before a FrameNet corpus, as a layer on top of a treebank, since PropBank closely follows the syntactic verb-argument structure, it has paid us off to start with the manual creation of the more abstract FrameNet annotation layer from which the PropBank layer can be derived semi-automatically. It would not be possible the other way around.

It is also often the case that language-specific framesets are defined in advance to create language-specific FrameNet or PropBank annotations. Our design decision to reuse the existing framesets of English FrameNet and English PropBank, although introduce some cross-lingual issues, allow for cross-lingual linguistic studies and for the development of cross lingual semantic parsers.

## 6. Acknowledgements

This work has received financial support from the European Regional Development Fund under the grant agreements 1.1.1.1/16/A/219 and 1.1.1.2/VIAA/1/16/188. Later development was supported by the Latvian Council of Science under the grant agreement lzp-2018/2-0216.

## 7. Bibliographical References

Akbik, A., Chiticariu, L., Danilevsky, M., Li, Y., Vaithyanathan, S., and Zhu, H. (2015). Generating high quality proposition banks for multilingual semantic role labeling. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 397–407, Beijing, China.

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

Bhatt, R., Narasimhan, B., Palmer, M., Rambow, O., Sharma, D., and Xia, F. (2009). A Multi-Representational and Multi-Layered Treebank for Hindi/Urdu. In *Proceedings of the 3rd Linguistic Annotation Workshop*, pages 186–189, Suntec, Singapore.

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2006). The SALSA Corpus: a German Corpus Resource for Lexical Semantics. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*, Genoa, Italy.

Cai, R. and Lapata, M. (2019). Semi-Supervised Semantic Role Labeling with Cross-View Training. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1018–1027.

Duran, M. S. and Aluisio, S. M. (2012). Propbank-Br: a Brazilian Treebank annotated with semantic role labels. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, pages 1862–1867, Istanbul, Turkey.

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

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

Gruzitis, N., Pretkalnina, L., Saulite, B., Rituma, L., Nespore-Berzkalne, G., Znotins, A., and 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)*, pages 4506–4513, Miyazaki, Japan.

Hajič, J., Ciaramita, M., Johansson, R., Kawahara, D., Martí, M. A., Màrquez, L., Meyers, A., Nivre, J., Padó, S., Štěpánek, J., Straňák, P., Surdeanu, M., Xue, N., and Zhang, Y. (2009). The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages. In *Proceedings of the 13th Conference on Computational Natural Language Learning (CoNLL): Shared Task*, pages 1–18.

Haverinen, K., Kanerva, J., Kohonen, S., Missilä, A., Ojala, S., Viljanen, T., Laippala, V., and Ginter, F. (2015). The Finnish Proposition Bank. *Language Resources and Evaluation*, 49(4):907–926.

Lopez de Lacalle, M., Laparra, E., Aldabe, I., and Rigau, G. (2016). A Multilingual Predicate Matrix. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pages 2662–2668, Portoroz, Slovenia.

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., and Zeman, D. (2016). Universal Dependencies v1: A Multilingual Treebank Collection. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*, pages 1659–1666.

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

Palmer, M. (2009). SemLink: Linking PropBank, VerbNet and FrameNet. In *Proceedings of the Generative Lexicon Conference*, pages 9–15, Pisa, Italy.

Schuler, K. K., Dang, H. T., and Palmer, M. (2000). Class-Based Construction of a Verb Lexicon. In *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI)*.

Xue, N. (2008). Labeling Chinese Predicates with Semantic Roles. *Computational Linguistics*, 34(2):225–255.