

# Deriving a PropBank Corpus from Parallel FrameNet and UD Corpora

International FrameNet Workshop 2020

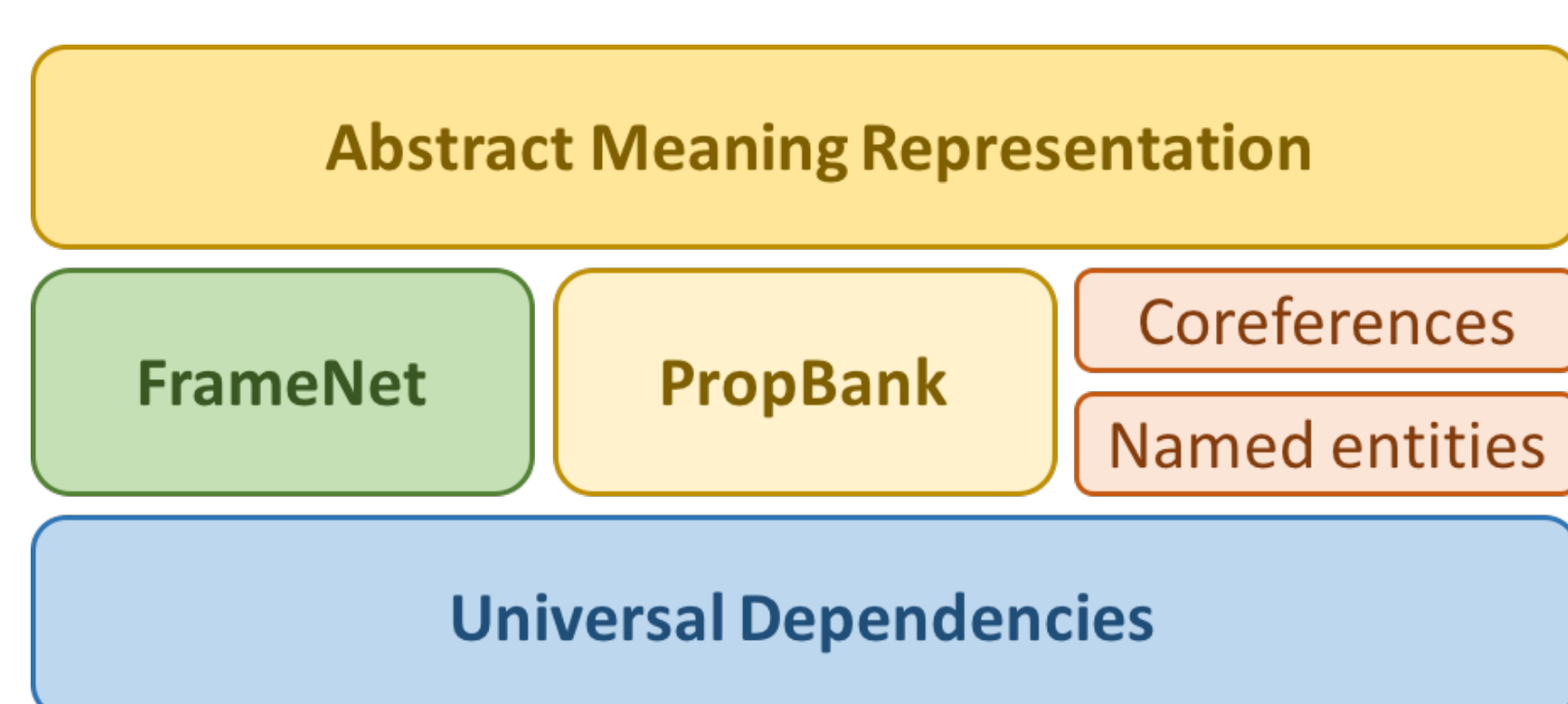
Normunds Grūzītis, Roberts Dargis, Laura Rituma, Gunta Nešpore-Bērzkalne, Baiba Saulīte

University of Latvia, Institute of Mathematics and Computer Science

<http://ailab.lv>

{normunds.gruzitis,roberts.dargis,laura.rituma,gunta.nespore,baiba.valkovska}@lumii.lv

**Abstract.** We propose an approach for generating accurate and consistent **PropBank** annotations, given a **FrameNet**-annotated corpus which has an underlying dependency annotation layer, namely, a parallel **UD** treebank. The PropBank annotation layer of such a multilayer corpus can be semi-automatically derived from the existing FrameNet and UD annotation layers, by providing a mapping configuration from LUs in a non-English FrameNet to English PropBank **predicates**, and a mapping configuration from FrameNet **FEs** 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 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.



**(1) Mapping from lexical units in Latvian FrameNet to predicates in English PropBank:**

LEMMA	UPOS	PRED <sub>FN</sub>	PRED <sub>PB</sub>
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

**(2) Mapping from FrameNet FEs to PropBank semantic roles, taking UD dependency relations into account:**

PRED <sub>FN</sub>	APRED <sub>FN</sub>	DEP	PRED <sub>PB</sub>	APRED <sub>PB</sub>
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

The screenshot shows the 'ķerties.VERB.Activity\_start' interface. It lists suggestions filtered by verb translation candidates, such as 'begin.01' (start, with agent), 'dawn.01' (to become known), and 'start.01' (begin). It also shows suggestions from SemLink and the Predicate Matrix, including 'initiate.01', 'launch.01', 'take\_up.31', and 'commence.01'.

The screenshot shows the 'PropBank predicate: begin.01' interface. It displays 'FrameNet elements => PropBank roles' with mappings for 'Activity - iobj' (A1-PPT: thing begun), 'Agent - nsubj' (A0-PAG: beginner, Agent), and 'Time - obl' (AM-TMP: temporal). Below are 'FrameNet corpus examples' with highlighted FrameNet elements and their corresponding PropBank roles.

The screenshot shows the 'Agent - nsubj' role selection interface. It displays a dropdown menu with 'A0-PAG: beginner, Agent' selected. Below are 'SUGGESTIONS FROM PREDICATE MATRIX' and 'CORE ROLES' (A1-PPT: Thing begun, A2-MNR: Instrument) and 'ARGM ROLES' (AM-ADJ: adjectival (nouns only), AM-ADV: adverbial modification, AM-CAU: cause).

**Statistics (PBLV vs. FNLV):**

- 1,322 verbs (97.3%)
- 2,377 LUs (92.2%)
- 521 BFN frames (96.5%)  
Note: 44.2% of 1,222 BFN frames
- 1,033 PB predicates  
Note: 9.7% of 10,687 PB predicates
- 20,054 annotation sets (96.0%)

