

SIGNOR



Annotating for Slovene Sign Language Corpus

Boštjan Jerko, Špela Vintar

Japina d.o.o.

University of Ljubljana, Faculty of Arts

bostjan@japina.eu, spela.vintar@ff.uni-lj.si

Project Signor

- 80 informants recorded
- 126 movies recorded
- 4198 glosses
- 120 compounds
- Glosses search engine

<http://www.lojze.si/signor/en>

Comparison between BSL annotation and Signor

Two-handed signs

We make no distinction between hands.

Buoys

We use classifiers with description. There is a distinction between motion and form classifiers.

Manual negative incorporation

Not used.

Directional verbs

Gramatical modification is encoded in separate ID-gloss UČITI [to teach], UČITI SE [to learn], UČITI ME [to teach me]

Sign names of persons

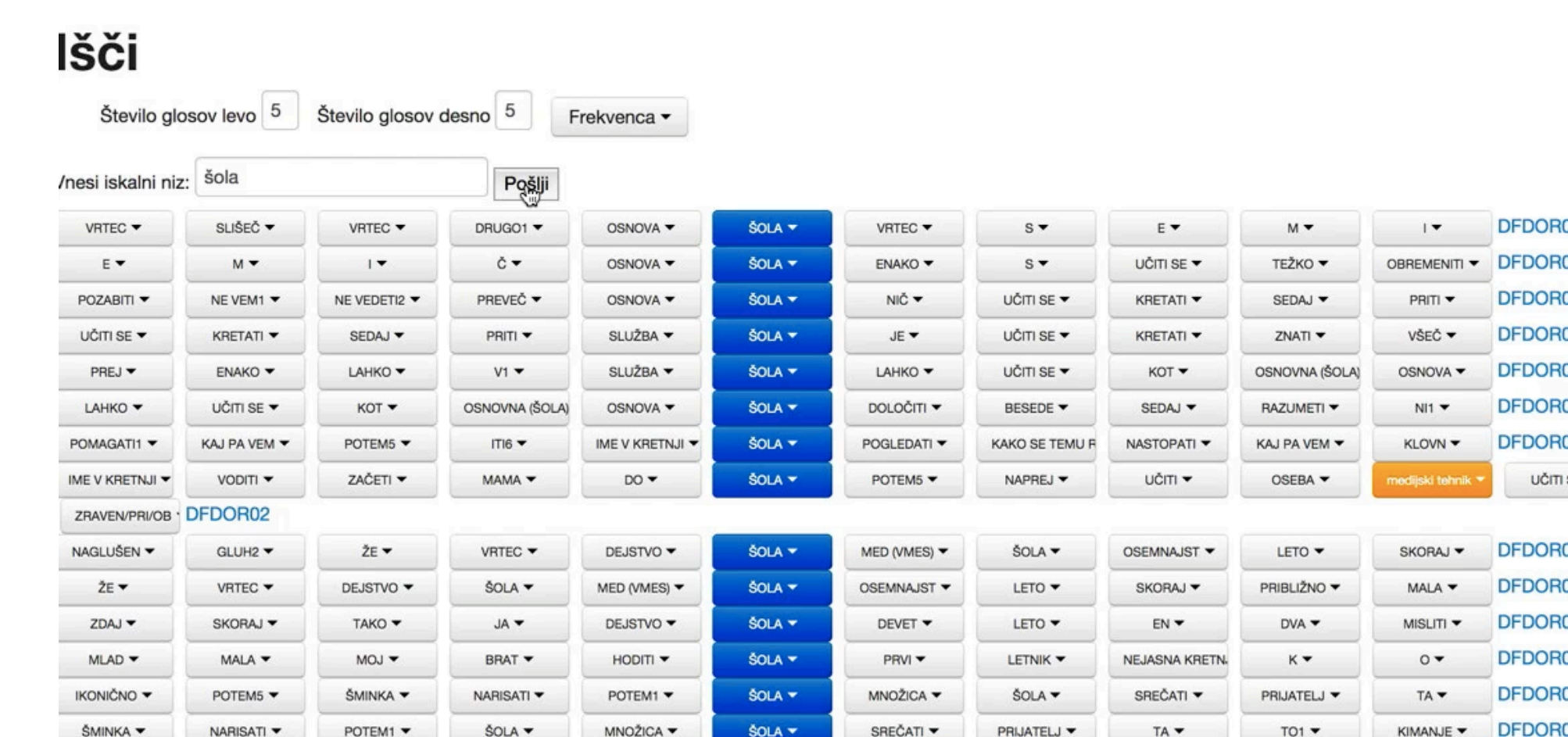
One gloss for all sign names due to data protection issues.

Uncertancies

The seed source of glosses was provided from previous sign language dictionaries. We marked indecipherable signs with one gloss (NEJASNA KRETNJA [UNCLEAR SIGN]). Since we had just two annotators they managed to discuss unknown or unrecognisable signs between themselves. No special glosses for other issues.

Top ten fillers with frequencies

TAKO [so] 218, KAJ PA VEM [I don't know] 82, TO JE VSE [that's all] 60, KAJ ŠE [what else?] 55, TAKO JE [that's right] 35, EH [um] 34, KAJ ČEŠ [what can you do?] 30, TA [that] 20, TO JE TO [that's that] 20, KAKO ŽE [what's it called?] 13



Compounds

Sort of the same. We use two glosses and join them in separate tier.

Ordinal numbers

Get a different ID-gloss.

Classifier/depicting signs

Shape constructions

Type-like classifier/depicting signs

ID-gloss is iconic movement (freq: 359) and iconic shape (freq: 323)

Variants

If a sign has at least one variant, all are numbered including the main lexical form. The advantage is that a number in ID-gloss indicates the existence of variants.

Segmentation into utterances

We annotated 3000 utterance boundaries. We plan to use these to explore SZJ syntax.

References

Crasborn, O., R. Bank and K. Cormier (2015). Digging into Signs: Towards a gloss annotation standard for sign language corpora, First draft
Vintar, Š. (2015). Lexical Properties of Slovene Sign Language: A Corpus Based Study, Sign Language Studies Vol. 15 No. 2 Winter
Vintar, Š., B. Jerko and M. Kulovec (2012). Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, Compiling the Slovene Sign Language Corpus, 159-162