



Neural metaphor detection for Slovene

Matej Klemen
(supervisor: prof. Marko Robnik Šikonja)

The problem

- **what** is a metaphor, **why** detect them

His words cut deeper than a knife.

- the tools:

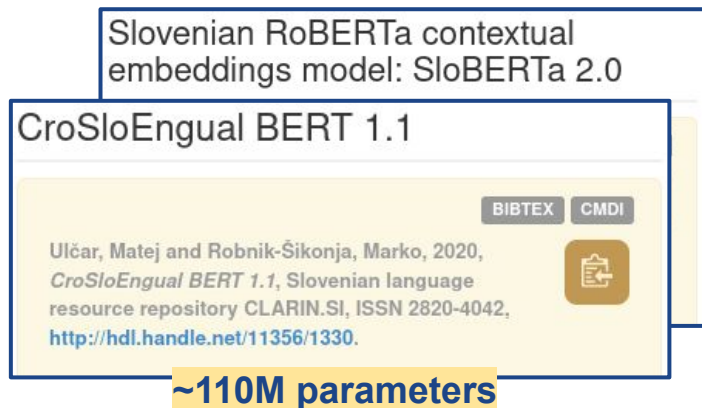
Slovenian RoBERTa contextual embeddings model: SloBERTa 2.0

CroSloEngual BERT 1.1

BIBTEX CMDI

Ulčar, Matej and Robnik-Šikonja, Marko, 2020, *CroSloEngual BERT 1.1*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1330>.

~110M parameters



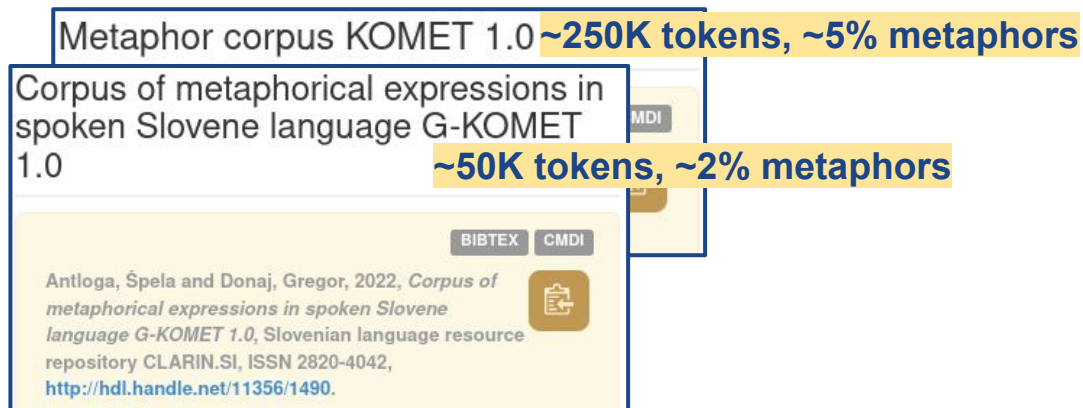
(models)

Metaphor corpus KOMET 1.0 **~250K tokens, ~5% metaphors**

Corpus of metaphorical expressions in spoken Slovene language G-KOMET 1.0 **~50K tokens, ~2% metaphors**

BIBTEX CMDI

Antloga, Špela and Donaj, Gregor, 2022, *Corpus of metaphorical expressions in spoken Slovene language G-KOMET 1.0*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1490>.



(data)

Results

- **F1 score**: correctness and coverage of the found metaphors

	Model	KOMET	G-KOMET
<i>full dataset</i>	(random)	0.114	0.016
	SloBERTa	0.597	0.374
	CSEBERT	0.552	0.310
<i>noun + verb dataset</i>	(random)	0.035	0.018
	SloBERTa	0.391	0.407
	CSEBERT	0.265	0.368

Future work

- considering metaphors at token level is limited, **consider spans**
- **cross-lingual transfer**: does learning metaphor detection on other languages improve detection on Slovene?
- **applications** of automatic metaphor detection