

LREC-COLING  2024

# SI-NLI: A Slovene Natural Language Inference Dataset and its Evaluation

Matej Klemen, Aleš Žagar, Jaka Čibej, Marko Robnik-Šikonja



**FRI**

UNIVERSITY OF LJUBLJANA

Faculty of Computer and Information Science

# Motivation

- What is natural language inference?

Premise: In autumn, the days are shorter and it gets dark quite early.

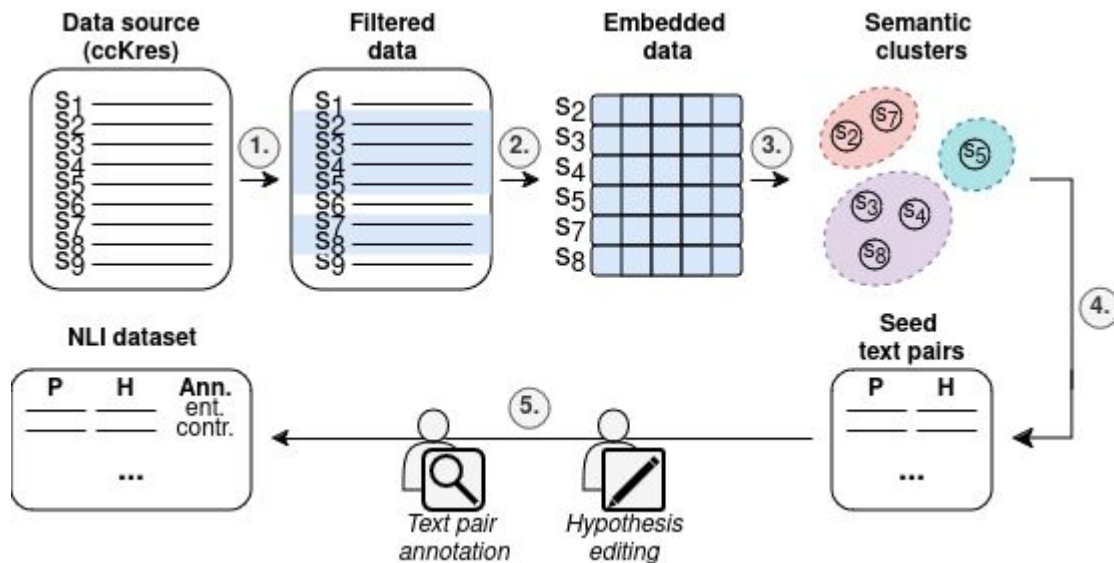
Hypothesis: Darkness comes relatively early in the shorter days of autumn.

Hypothesis follows premise => **Entailment**

- Problem: No resource for Slovene
- **How to create a new NLI resource?**

Translation, recasting, **creation from scratch**

# Our approach



# Pre-annotation

- **ccKres 1.0** reference corpus: free non-commercial use of data
- remove oddly structured sentences (partial sentences, signatures, dates, etc.)
- embed texts using **sentence embeddings** (LaBSE)
- cluster using **DBSCAN** algorithm
- sample pairs from clusters (according to our annotation budget)

# Annotation

## **Part 1: hypothesis editing**

- preliminary annotation and construction of annotation guidelines
- introductory session with the annotators
- small training batch for the annotators
- recurring weekly minimum creation quota: 240 pairs (est. 8 hrs of work)

## Part 2: cross-checking

- two rounds to “resolve” disagreements
- resolution of major disagreement by the authors

## (Part 3: translation into English)

- automatic translation using DeepL, followed by manual inspection/correction

**SI-NLI 1.0:** ~6000 examples

# Evaluation of models

Model	Accuracy
majority	0.344
gpt-sl-base	0.479 (0.019)
SloBERTa	0.744 (0.008)
CSEBERT	0.667 (0.005)
mBERTc-base	0.595 (0.011)
XLM-R-base	0.670 (0.011)
XLM-R-large	0.791 (0.014)
T5-sl-small	0.653 (0.004)
T5-sl-large	0.590 (0.007)
mT5-small	0.540 (0.012)
mT5-base	0.621 (0.007)
mT5-large	0.767 (0.005)
GPT3.5-turbo	<b>0.857</b>

**Slovene**

Model	Accuracy
majority	0.344
CSE-BERT	0.623 (0.011)
mBERTc-base	0.489 (0.031)
XLM-R-base	0.620 (0.038)
XLM-R-large	0.765 (0.017)
mT5-small	0.428 (0.011)
mT5-base	0.579 (0.029)
mT5-large	0.746 (0.017)
GPT3.5-turbo	<b>0.837</b>

**cross-lingual**  
**(sl -> en)**

# Room for improvement - misclassifications

## Example 1:

[P] V njej je prebujal občutke, ki jih ni poznala. (*He made her feel feelings she did not know before.*)

[H] Ob njem je začutila doslej neznana občutja. (*By his side she felt feelings she did not feel thus far.*)

**predicted:** neutral, **correct:** entailment

hidden pronoun "he"

## Example 2:

[P] »Ja?« sem rekel, ko sem zaslišal trkanje po vratih. (*"Yes?" I uttered once I heard the knocking on the door.*)

[H] Zaradi zanimanja sem se oglasil na trkanje po vratih. (*Due to my curiosity I answered the door knock.*)

**predicted:** neutral, **correct:** entailment

common sense



### Example 3:

[P] S prijateljem še ne bova vrgla puške v koruzo. (*Me and my friend will not give up yet.*)

[H] “S prijateljem sva se naveličala neuspehov, zato bova odnehala.” (*Me and my friend got tired of failure, so we will quit.*)

**predicted:** neutral, **correct:** contradiction      **idiom**

### Example 4:

[P] Vabljeni ste na konferenco, ki se bo zgodila v četrtek, 11. 11. 2010 v Ljubljani. (*You are invited to the conference happening on Thursday 11. 11. 2010 in Ljubljana.*)

[H] Vabimo vas na slavni dogodek, ki bo potekal 11. novembra 2010 v prestolnici. (*We invite you to the gala event happening on November 11 in the capital.*)

**predicted:** neutral, **correct:** entailment

**world knowledge**

# Conclusion

- new resource for Slovene and cross-lingual (EN -> SL and SL -> EN) NLI
- evaluation of baselines
- future directions: improvement of resource, alternative approach to modeling

**Contact:** *matej.klemen@fri.uni-lj.si*



**SI-NLI**  
(data)



**SI-NLI-en**  
(data)



**SloBENCH**  
leaderboard  
(evaluation)