
Cross-lingual parsing
with
Universal Dependencies

Team

Kostya Vinogorodskiy

Marina Kustova

Sasha Martynova

Pasha Stepachev

Mentors

Olga Lyashevskaya

Francis Tyers

What?

- Using annotated data from one language to parse another
- No annotation for target language
- The languages should be closely related

Work plan

- Research on existing studies on the subject
- Choose source and target languages
- Find parallel texts for each pair
- Annotate the source language using UD parser
- Transfer the annotation to the target language

State of the art

- Delexicalized parsing (without lexical info)
- Annotation projection (using aligned parallel corpora)
- Treebank translation (with a phrase-based statistical machine translation)

Source & Target languages

- Russian -> Belarussian
- Norwegian Bokmål -> Faroese
- German -> Yiddish
- some Romance language -> some other Romance language

Parallel texts ?

- … or MT (fao-nor)

And then ...

- Research on existing studies on the subject
- Choose source and target languages
- Find parallel texts for each pair
- Annotate the source language using UD parser
- Transfer the annotation to the target language

Further plans

- Will be discussed on the next meeting on the 29th of November

Bibliography

1. Jörg Tiedemann (2017) "Cross-Lingual Dependency Parsing for Closely Related Languages – Helsinki's Submission to VarDial 2017"

2. Jörg Tiedemann (2015) "Cross-Lingual Dependency Parsing with Universal Dependencies and Predicted PoS Labels". Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015), pages 340–349, Uppsala, Sweden, August 24–26 2015.

Bibliography

3. Jörg Tiedemann and Željko Agiç (2016) "Synthetic Treebanking for Cross-Lingual Dependency Parsing". Journal of Artificial Intelligence Research 55 (2016) 209-248

the current major approaches. We emphasize *synthetic treebanking*: the automatic creation of target language treebanks by means of annotation projection and machine translation.

In our setup, we always use the test sets provided by the Universal Dependency Treebank version 1 (UDT) (McDonald et al., 2013)...

Improved Annotation Projection; Phrase-Based,Syntax-Based Treebank Translation and some more...

That's it!