ANOTATING TENSE, MOOD AND VOICE FOR ENGLISH, FRENCH AND GERMAN

Anita Ramm1,4, Sharid Loáiciga2,3, Annemarie Friedrich4 and Alexander Fraser4
1University of Stuttgart, 2Université de Genève, 3Uppsala University, 4LMU Munich

Motivation
Lack of tools for automatic annotation of syntactic tense, mood and voice (TMV):
- English PropBank: annotations for tense, mood and aspect, but no annotations for subjunctive constructions
- German TuBa-D/Z: only morphological features of the verbs
- French FTB: only morphological features of the verbs

TMV annotations are interesting for different fields of linguistics and NLP:
- Theoretical research and automatic modeling of mono/cross-lingual use of tense, mood and voice
- Useful features for classification tasks such as authorship, epoch, domain, etc.

Method
Verbal complexes (VCs):
- Extracted from dependency trees
- Finite as well as non-finite VCs

TMV assignment:
- Based on hand-crafted rules
- Syntactic TMV values of a whole VC

TMV rules:
- Rely on POS sequences, morphological as well as lexical information
- Use external verb lists for handling ambiguous active/passive constructions, e.g.: ist gegangen vs. ist geschrieben

Problem complexity ⇔ Verbal complex complexity

Observations:
- Two TMV combinations
  - perfect & pass - but 15 different VCs (different POS sequence and/or finite verb morphology)
  - Different mood values
    (e.g. hätte gesehen (had/see), könnte gesehen haben (could see))
  - Additionally enlarge both the TMV set, as well as the number of the differing VC
  - Total number of the DE VCs: 170!

Rules need to:
- Consider many different VCs to ensure both high precision, as well as high recall
- Distinguish between ambiguous VCs (e.g. will drive vs. would drive)
- Have access to the relevant language-specific information: POS tags, morphological analysis, lemmas

Future work
Annotating rules that handle ambiguous constructions
- Addition of missing FR rules
- Improvement of the VC extraction procedures

Tool adaptation/extension:
- Adaptation to the universal dependency trees
- Join us to develop TMV rules for other languages!

Download, test and use the tool!
Download: https://github.com/aniramm/tmv-annotator
Online-demo: https://clarin09.ims.uni-stuttgart.de/tmv/
Feedback: ramm@ims.uni-stuttgart.de

Acknowledgement
This work has received funding from the DFG grant Models of Morphosyntax for Statistical Machine Translation (Phase 2), the EU’s Horizon 2020 research and innovation programme under grant agreement No. 644402 (HimL), and from the ERC under grant agreement No. 640550.

We thank André Blessing for developing the demo version of the tool!