

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

Anita Ramm^{1,4}, Sharid Loáiciga^{2,3}, 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** TüBa-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, genre, etc.

CONTRIBUTIONS

Tool specifics:

- First publicly available tool for TMV annotation for English, German and French
- Output information: TMV values, negation, progressiveness (EN), coordination, clause boundaries (DE)
- Tool output in the TSV format
- Open-source implementation
- Online demo provided

METHOD

Verbal complexes (VCs):

- Extracted from dependency trees
- Finite as well as non-finite VCs

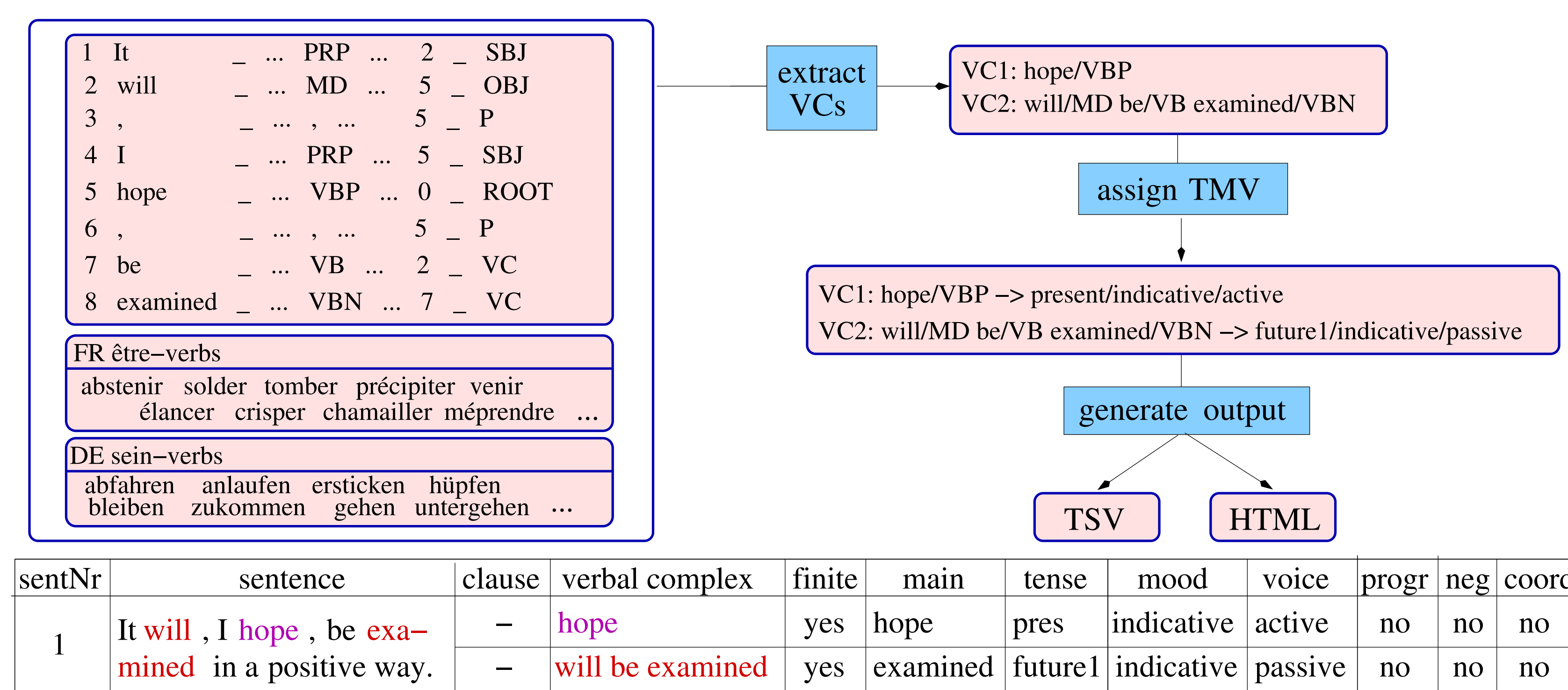
TMV assignment:

- 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*

PROCESSING PIPELINE



PROBLEM COMPLEXITY ⇔ VERBAL COMPLEX COMPLEXITY

GERMAN PAST TENSE 'PERFEKT'

HABEN/SEIN _{pr,ind} + PPART	
hat gelesen	perf ind act
has read	
ist gefahren	perf ind act
was driving	
ist geschrieben	pres ind pass
is written	
SEIN _{pr,ind} + PPART + WERDEN _{ppart}	
ist gelesen worden	perf ind pass
has been read	
MODAL _{pr,ind} + PPART + HABEN/SEIN _{ppart}	
kann gekommen sein	perf ind act
could have come	
kann gelesen haben	perf ind act
could have read	
HABEN _{pr,ind} + INF + MODAL _{inf}	
hat sehen können	perf ind act
could have seen	
...	

Observations:

- **Two TMV combinations** - perf|ind|act & perf|ind|pass - but **15 different VCs** (different POS sequence and/or finite verb morphology)!
- Different mood values (e.g. *hätte gesehen* (would have seen), *könnte gesehen haben* (could have seen)) 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

ANNOTATION ACCURACY

Evaluation of annotations on Mate trees:

Language	tense	mood	voice	all
EN	81.5	88.1	86.1	76.8
DE	80.8	84.0	81.5	76.4
FR	86.1	93.4	82.5	75.2

⇒ Many errors due to erroneous parses and/or morphological analysis

Evaluation of annotations on gold trees:

Language	tense	mood	voice	all
EN	90.9	90.9	90.0	88.2
DE	88.2	90.9	88.2	88.2
FR	89.2	95.0	85.2	78.4

Observations:

- Confusion between active tenses and certain stative passive constructions
- FR rule set requires further improvements
- VC extraction: falsely attached gerunds: *Elizabeth hates being called 'Liz'*.

FUTURE WORK

Annotations:

- Development of 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!