Automatic annotation of tense, mood and voice
for English, French and German

Anita Ramm
Universität Stuttgart

Sharid Loáiciga
Uppsala University

Annemarie Friedrich
LMU München

ramm@ims.uni-stuttgart.de sharid.loaiciga@lingfil.uu.se anne@cis.uni-muenchen.de

Tag
Datum
Zeit
Raum

We present a rule-based system for automatic annotation of syntactic tense, mood and voice (TMV) for English, French and German in arbitrary texts (Ramm et al., 2017). The annotation rules are applied on verbal complexes (VCs) which are automatically extracted from Mate dependency parse trees[1]. We consider both finite (e.g. 'is saying'), as well as non-finite VCs (e.g. 'to say'). The annotation rules make use of different types of information, i.e., lexical, lemma, POS and morphological features of the verbs in a VC. For example, given the German VC 'seien beantwortet worden', the applicable rule is the following: 'VAFIN[pres/ind] VVPP VVPP[worden]' → *past/konjunktivI/passive*. The amount of the rules for each of the considered languages depends on the number of the different morpho-syntactic patterns of the VCs they include. For instance, German has about 170 different VC patterns. The rules in our tool capture all of them ensuring not only high annotation recall, but high annotation precision as well.

In a combination with the Mate trees, the overall annotation accuracy is between 75-76%. The accuracy increases to 78-88% when gold parse trees are used as the underlying representation of the sentences which are to be annotated. Most of the errors are due to inaccurate morphological analysis of the finite verbs and erroneous VCs. Additionally, there are ambiguous VC patterns, such as the stative passive which presents the same configuration like some active tenses, (e.g., (DE) 'ist/is gegangen/went' → past/indicative/**active** vs. 'ist/is geschrieben/written' → present/ indicative/**passive**). To distinguish between such constructions, additional resources such as verb lists were added to the system.

The annotation tool is open source[2], and has an online demo[3].

References: ● A. Ramm, S. Loáiciga, A. Friedrich and A. Fraser (2017): Annotating tense, mood and voice for English, French and German. *ACL - demo session.*

---

[1] https://code.google.com/archive/p/mate-tools/
[2] https://github.com/aniramm/tmv-annotator
[3] https://clarin09.ims.uni-stuttgart.de/tmv/index.html (thanks to André Blessing)