Tagger for Intensifiers – An Automatic Identification for Intensifiers of Adjectives

Imge Yüzüncüoglu[1]
[1]Ruhr-Universität Bochum
Imge.Yuezuencueoglu@rub.de

Intensifiers are rapidly changing words or wordparts which can be used to amplify the semantic meaning of a statement as in (1a) or to downtone it as in (1b) (Stratton, 2020). There is no finite amount of them and analyzing intensifiers turns out to be time-consuming as every token in a corpus has to be annotated individually.

(1)   a. *Der Film ist <u>sehr</u> sehenswert.* "The movie is <u>very</u> worth seeing."

        b. *Der Film ist <u>kaum</u> sehenswert.* "The movie is <u>hardly</u> worth seeing."

Until now there have been two ways for an automatic annotation of intensifiers. One is the usage of wordlists that contain intensifiers and the other is the usage of general POS-taggers that contain the PTKIFG-tag of the POS-tagset from Empirist (2015) like the POS-tagger SoMeWeTa by Proisl (2018). But both options need to be revised manually as the wordlists cannot distinguish if a word it finds is used as an intensifier or not. Additionaly, the PTKIFG-tag is a collective tag for intensifiers, degree- and focusparticles. Therefore, the goal is to implement a tagger that can identify intensifiers of adjectives in a reliable way.

The here presented tagger for intensifiers (TafIn) does exactly that. TafIn is based on the Conditional Random Field algorithm and trained with manually annotated data, consisting of 468 blogposts from the TwiBloCoP corpora (2021). Different token information based on the definition of intensifiers like POS-tags and tokens alongside of the current token were used to generate abstract features to identify intensifiers in a text contextfree. To get a clear impression of TafIn's results, it and two other baselines, the first one (B1) being the PTKIFG-tag within SoMeWeTa and the second one (B2) being a combination of a wordlist and heuristical methods, were evaluated. The resutls like the macroaverage F1-scores presented in the table below show that TafIn works nearly as twice as good as the previous methods, thus it can be used as a more reliable annotation tool for intensifiers of adjectives.

	B1: SoMeWeTa	B2: Wordlist	**TafIn**
Macroaverage F_1-Score	0.47	0.44	**0.88**

References: • EmpiriST. 2015. GSCL Shared Task: Automatic linguistic annotation of computer-mediated communication / social media. *EmpiriST* 2015. • Proisl, T. (2018). SoMeWeTa: A part-of-speech tagger for German social media and web texts. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC 2018), 665–670. • Stratton, J. M. (2020). Adjective Intensifiers in German. *Journal of Germanic Linguistics* 32(2), 183–215.