

Saarland University's Participation in the GERman SenTiment AnaLysis shared Task (GESTALT)

Michael Wiegand, Christine Bocionek, Andreas Conrad, Julia Dembowski
Jörn Giesen, Gregor Linn, Lennart Schmeling
Spoken Language Systems
Saarland University
D-66123, Saarbrücken, Germany
michael.wiegand@lsv.uni-saarland.de

Abstract

We report on the two systems we built for Task 1 of the German Sentiment Analysis Shared Task, the task on *Source, Subjective Expression and Target Extraction from Political Speeches (STEPS)*. The first system is a rule-based system relying on a predicate lexicon specifying extraction rules for verbs, nouns and adjectives, while the second is a translation-based system that has been obtained with the help of the (English) MPQA corpus.

1 Introduction

In this paper, we describe our two systems for Task 1 of the German Sentiment Analysis Shared Task, the task on *Source, Subjective Expression and Target Extraction from Political Speeches (STEPS)* (Ruppenhofer et al., 2014). In that task, both *opinion sources*, i.e. the entities that utter an opinion, and *opinion targets*, i.e. the entities towards which an opinion is directed, are extracted from German sentences. The opinions themselves have also to be detected automatically. The sentences originate from debates of the Swiss Parliament (*Schweizer Bundesversammlung*).

The first system is a rule-based system relying on a predicate lexicon specifying extraction rules for verbs, nouns and adjectives, while the second is a translation-based system that has been obtained with the help of the (English) MPQA corpus (Wiebe et al., 2005).

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Page numbers and proceedings footer are added by the organizers. License details: <http://creativecommons.org/licenses/by/4.0/>

This shared task has been organized for the first time. No labeled training data have been available.

2 Rule-based System

The pipeline of the rule-based system is displayed in Figure 1. The major assumption that underlies this system is that the concrete realization of opinion sources and targets is largely determined by the opinion predicate¹ by which they are evoked. Therefore, the task of extracting opinion sources and targets is a lexical problem, and a lexicon for opinion predicates specifying the argument position of sources and targets is required. For instance, in Sentence (1), the sentiment is evoked by the predicate *liebt*, the source is realized by its subject *Peter* while the target is realized by its accusative object *Maria*.

(1) [Peter]_{subj}^{source} **liebt**_{sentiment} [Maria]_{obja}^{target} .
(Peter loves Maria.)

With this assumption, we can specify the demands of an opinion source/target extraction system. It should be a tool that given a lexicon with argument information about sources and targets for each opinion predicate

- checks each sentence for the presence of such opinion predicates,
- syntactically analyzes each sentence and
- determines whether constituents fulfilling the respective argument information about

¹We currently consider verbs, nouns and adjectives as potential opinion predicates.

sources and targets are present in the sentence.

In the following, we describe how we implemented these different steps. The rule-based system will be made publicly available allowing researchers to test different sentiment lexicons with different argument information about opinion sources and targets.²

2.1 Linguistic Processing

Even though the data for this task already come in a parsed format, we felt the need to add further linguistic information. In addition to the existing constituency parse provided by the Berkeley parser (Petrov et al., 2006), we also included dependency parse information. With that representation, relationships between opinion predicates and their sources and targets can be formulated more intuitively.³

As a dependency parser, we chose *ParZu* (Sennrich et al., 2009). We also carried out some normalization on the parse output in order to have a more compact representation. To a large extent, the type of normalization we carry out is in line with the output of dependency parsers for English, such as the Stanford parser (de Marneffe et al., 2006). It is included since it largely facilitates writing extraction rules. The normalization includes

- (a) active-passive normalization
- (b) conflating several multi-edge relationships to one-edge relationships
- (c) particle-verb reconstruction

Our extraction rules assume a sentence in active voice, therefore sentences in passive voice (we exclusively consider the frequent German *von-Passiv*) need to be converted to active voice (a).⁴

²The code will be made available via the website of the shared task <https://sites.google.com/site/iggsasharedtask/task-1>

³As a matter of fact, the most appropriate representation for that task is semantic-role labeling (Ruppenhofer et al., 2008; Kim and Hovy, 2006; Wiegand and Klakow, 2012), however, there currently do not exist any robust tools of that kind for German.

⁴From a semantic point of view, the content of a sentence in passive voice and that of a sentence in active voice are, more or less, identical. Therefore, normalizing passive voice sentences to active voice sentences is legitimate.

This conversion is illustrated in Figure 2.

For our extraction rules, we want to specify the relationship between opinion predicates and their sources/targets as direct (or first-order) dependency relationships. In current dependency parsers for German, however, those two types of entities are often not connected via a direct edge, i.e. they are multi-edge (or second-order) relationships. We, therefore, wrote a set of rules collapsing those multi-edge relationships. A simple example is illustrated in Figure 3 for the case of predicate adjectives and their subjects. In Figure 3(a) *schön* and *Auto* are connected via *pred+subj* which we collapse to just *subj* in Figure 3(b).⁵ In a similar fashion, we also collapse prepositional objects as illustrated in Figure 4.

Finally, a considerable fraction of German verbs are *particle verbs* which means that several inflectional forms are split into two tokens, i.e. verb stem and some particle. These two tokens may then be separated by other constituents in a sentence. This is illustrated for *aufgeben* in Sentence (2) which is split in *gab* and *auf*. The ParZu dependency parser connects stems and particles via a dedicated relation edge. Thus the full lemma (as listed in the lexicon specifying the extraction rules) can be reconstructed.

- (2) Er *gab* das Rauchen vor 10 Jahren *auf*.
(He *gave up* smoking 10 years ago.)

2.2 The Extraction Rules

As already indicated above, the heart of the rule-based system is a lexicon that specifies the (possible) argument positions of sources and targets. So far, there does not exist a lexicon with that specific information which is why we came up with a set of default rules for the different parts of speech. The set of opinion predicates are the sentiment expressions from the PolArt system (Klenner et al., 2009). (For some runs for the benchmark, we also add sentiment expressions from SentiWS (Remus et al., 2010).) Every mention of such expressions will be considered as a mention of an opinion predicate, that is, we do not carry out any subjectivity word-sense disambiguation (Akkaya et al., 2009).

⁵The copula *ist* needs to be inserted for syntactic reasons in that sentence. It does not carry any semantic content and, therefore, can be dropped for our purposes.

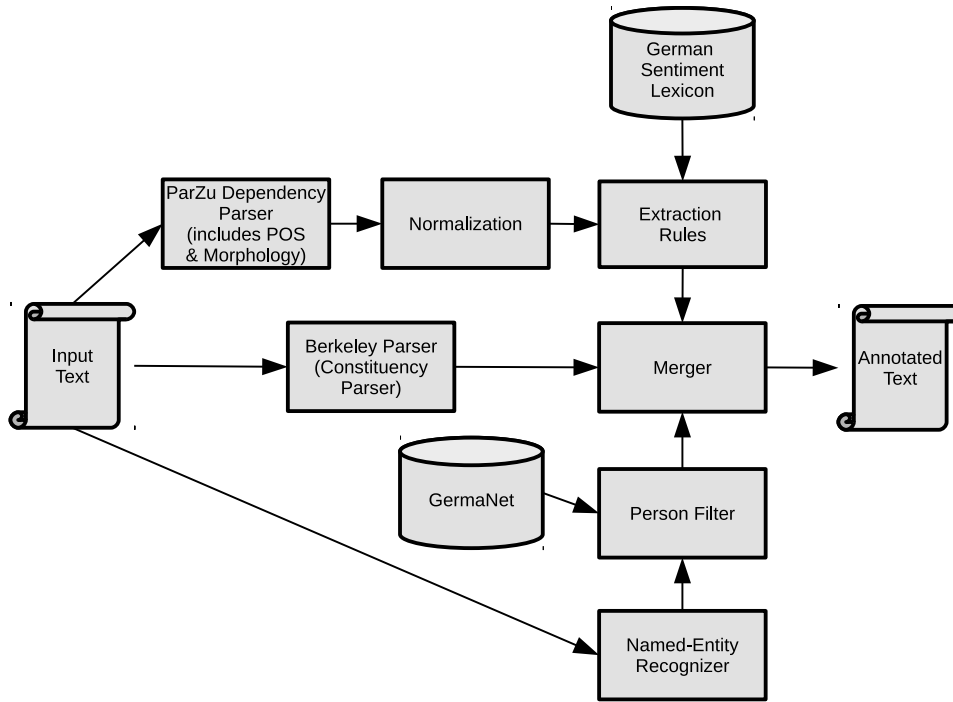
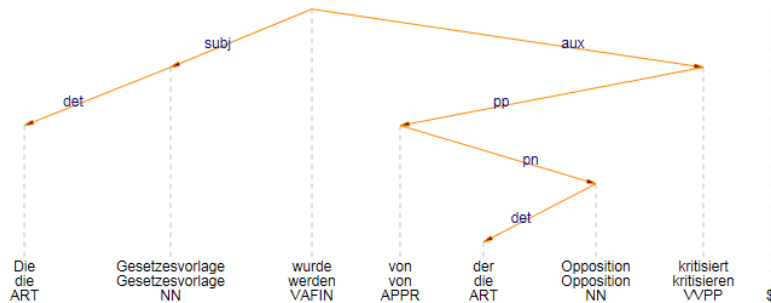
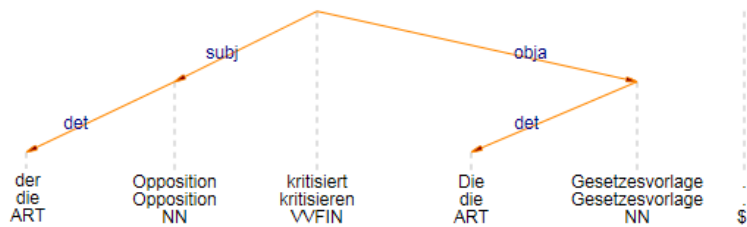


Figure 1: Processing pipeline of the rule-based system.

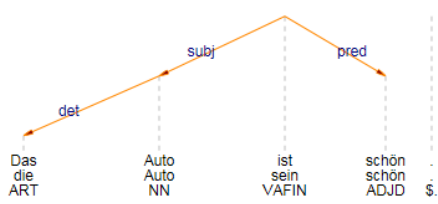


(a) original dependency parse

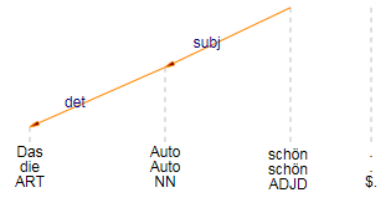


(b) normalized dependency parse

Figure 2: Illustration of normalizing dependency parses with passive voice constructions.

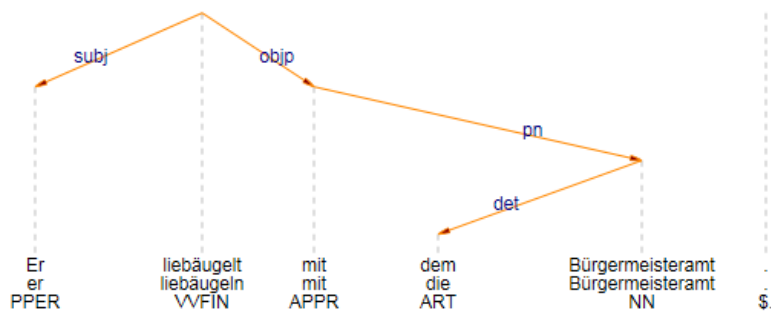


(a) original dependency parse

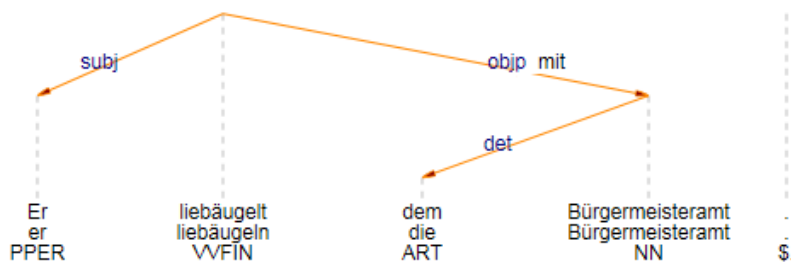


(b) normalized dependency parse

Figure 3: Illustration of normalizing dependency parses with predicative adjectives.



(a) original dependency parse



(b) normalized dependency parse

Figure 4: Illustration of normalizing dependency parses with prepositional complements.

These default extraction rules are designed in such a way that for a large fraction of opinion predicates with the pertaining part of speech they are correct. The rules are illustrated in Table 1. We currently have distinct rules for verbs, nouns and adjectives. All rules have in common that for every opinion predicate mention, at most one source and at most one target is assigned. The rules mostly adhere to the dependency relation labels of ParZu.⁶

The rule for verbs assumes sources in subject and targets in object position (1). Note that for targets, we specify a priority list. That is, the most preferred argument position is an dative object (*objd*), the second most preferred position is an accusative object (*obja*), etc. In computational terms, this means that the classifier checks the entire priority list (from left to right) until a relation has matched in the sentence to be classified. For prepositional complements, we also allow a wildcard symbol (*pobj*-*) that matches all prepositional complements irrespective of its particular head, e.g. *über das Freihandelsabkommen* (*pobj-ueber*) in (3).

- (3) [Deutschland und die USA]_{subj}^{source} **streiten** [über das Freihandelsabkommen]_{pobj-ueber}^{target}
(Germany and the USA quarrel over the free trade agreement.)

For nouns, we allow determiners (possessives) (4) and genitive modifiers (5) as opinion sources whereas targets are considered to occur as prepositional objects.

- (4) [Sein]_{det}^{source} **Hass** [auf die Regierung]_{pobj-auf}^{target} ...
(His hatred towards the government ...)
- (5) Die **Haltung** [der Kanzlerin]_{gmod}^{source} [zur Energiewende]_{pobj-zu}^{target} ...
(The chancellor’s attitude towards the energy revolution ...)

The rule for adjectives is different from the others since it assumes the source of the adjective to be the speaker of the utterance. Only the target

⁶The definition of those dependency labels is available at <https://github.com/rsennrich/ParZu/blob/master/LABELS.md>

Part of Speech	Source	Target
verb	subj	objd, obja, objc, obji, s, objp-*
noun	det, gmod	objp-*
adjective	author	attr-rev, subj

Table 1: Extraction rules for verb, noun and adjective opinion predicates.

has a surface realization. Either it is an attributive adjective (6) or it is the subject of a predicative adjective (7).

- (6) Das ist ein [guter]_{attr-rev}^{target} **Vorschlag**.
(This is a good proposal.)
- (7) [Der Vorschlag]_{subj}^{target} ist **gut**.
(The proposal is good.)

Our rule-based system is designed in such a way that, in principle, it would also allow more than one opinion frame to be evoked by the same opinion predicate. For example, in *Peter überzeugt Maria*/Peter convinces Maria, one frame sees Peter as source and Maria as target, and another frame where the roles are switched. Our default rules do not include such cases, since such property is specific to particular opinion predicates.

2.3 Filtering

Our extraction lexicon tends to overgenerate in several situations. This can be mainly ascribed to the fact that we do not carry out any word-sense disambiguation and we use simple default rules. The only means to rectify this shortcoming (to a certain extent) is by applying a heuristic filter. The filter that we apply concerns the plausibility of opinion sources. We only mark a phrase as an opinion source, if it denotes a person or a group of persons. We automatically detect this semantic information with the help of a named-entity recognizer (Faruqui and Padó, 2010) (in order to detect proper nouns) and GermaNet (Hamp and Feldweg, 1997), the German version of WordNet (Miller et al., 1990) (in order to cope with common nouns). In addition, we also formulate a set of rules for personal pronouns, e.g. the German pronoun *es*, similar to the English *it*, is fairly unlikely to denote a human being and therefore is not eligible to represent opinion sources.

2.4 Finding Phrases in the Constituency Parse

Having established a source or a target of an opinion predicate with the help of the extraction rules and (normalized) dependency parsing, we need to expand sources/targets to the corresponding phrases in a constituency parse. The dependency parser only specifies relations holding between words (i.e. *heads* of phrases). For this expansion, we use a simple heuristics which applies for both opinion sources and opinion targets. Figure 5 illustrates it for opinion sources. It identifies the lowest common ancestor for the opinion verb (i.e. *kritisiert*) and the head of its source (i.e. *Polizei*). Then, we choose as the phrase the node directly dominated by the lowest common ancestor and dominating the head of the source (i.e. the NP *die Polizei*).⁷ This heuristics is fairly reliable if both constituency and dependency parse provide a correct syntactic analysis of the pertaining sentence.

3 Translation-based System

Even though there currently do not exist any large datasets with sufficient labeled data for fine-grained sentiment analysis in German, there exist comparable resources for other languages, most notably for English. Therefore, we devised a translation-based system that tries to harness fine-grained labeled training data available in English. We chose the MPQA corpus (Wiebe et al., 2005). Due to the availability of annotation present in the MPQA corpus, the translation-based system only learns how to extract opinion sources from the MPQA corpus. In other words, that system will not detect any opinion targets. The pipeline of this system is illustrated in Figure 6.

The first step is to translate the MPQA corpus into German. This has been achieved by translating the raw text of this corpus by *Google Translate*⁸. Since the annotation of that corpus is not on the sentence level but on the phrase/word level, we need to align each word of a sentence with the

⁷Depending on the tree configuration, this node may, of course, also be a terminal node – in case the head of the source is immediately dominated by the lowest common ancestor. In such cases, the head of the source is already the constituent that we are looking for.

⁸<https://translate.google.com>

clearpage

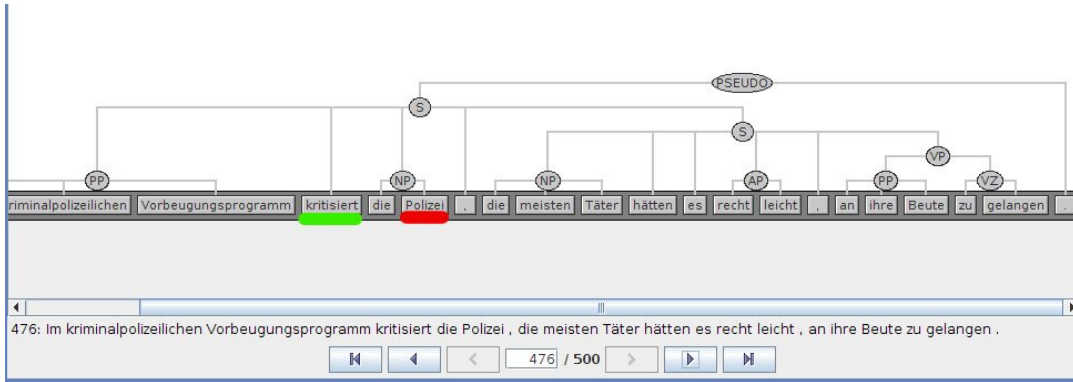
corresponding word in the German translation. With the translation from *Google Translate*, we just obtain a sentence alignment. In order to obtain a word alignment, we employ *GIZA++* (Och and Ney, 2003).

Once a German version of the MPQA corpus has been reconstructed, two supervised learning classifiers are trained. The first is to detect subjective expressions or phrases. For that, we employ a conditional random field (Lafferty et al., 2001). As an implementation, we chose *CRF++*⁹. As a motivation, we chose a sequence-labeling algorithm because the task of detecting sentiment expressions or even (continuous) sentiment phrases is similar to other tagging problems, such as part-of-speech tagging or named-entity recognition. The feature templates for our sentiment tagger are displayed in Table 2. We use *CRF++* in its standard configuration; as a labeling scheme, we used the simple IO-notation.

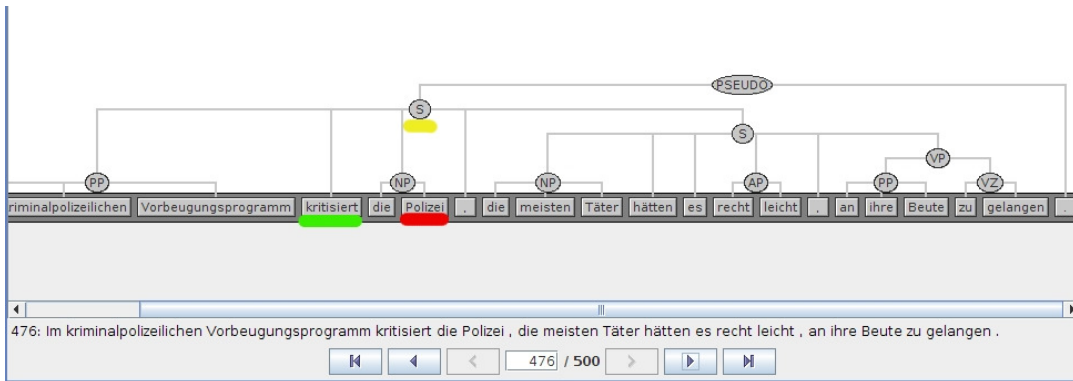
The second classifier extracts for a subjective phrase detected by the CRF the corresponding opinion source, if it exists. For this second task, a support vector machine (SVM) was chosen. As an implementation, we chose *SVM^{light}* (Joachims, 1999). The instance space is a set of tuples comprising candidate opinion sources (i.e. noun phrases of a sentence) and sentiment expressions/phrases (detected by the sentiment tagger). The setting is a binary classification deciding for each tuple whether the noun phrase is a genuine opinion holder of the sentiment expression/phrase, or not. Opinion sources are typically persons or groups of persons. Such entities can only be expressed by noun phrases which is why we reduce our instances to those types of constituents. SVM was chosen as a learning method since this task deals with a more complex instance space, and SVM, unlike sequence labelers, allow a fairly straightforward encoding of that instance space. The feature templates of the SVM are illustrated in Table 3.

Figure 6 indicates that a different parser (Stanford parser (Rafferty and Manning, 2008)) was used for the translation-based system compared to the rule-based system (Berkeley parser & ParZu parser). The reason for this is that those two

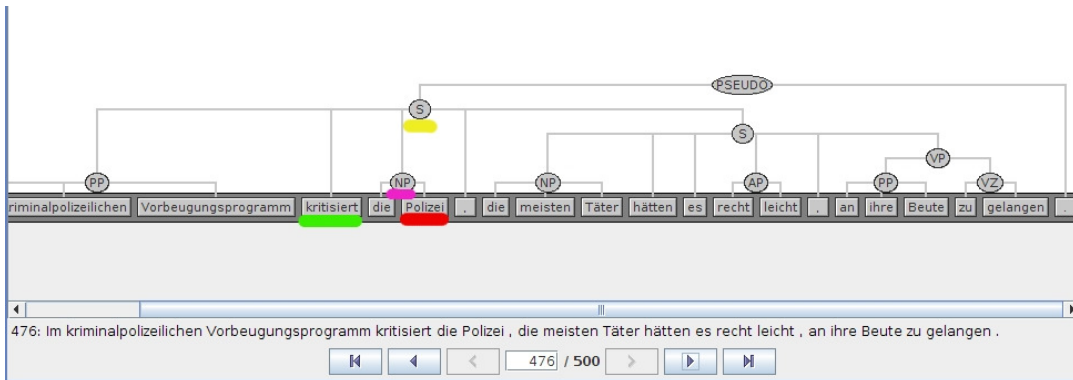
⁹<https://code.google.com/p/crfpp/>



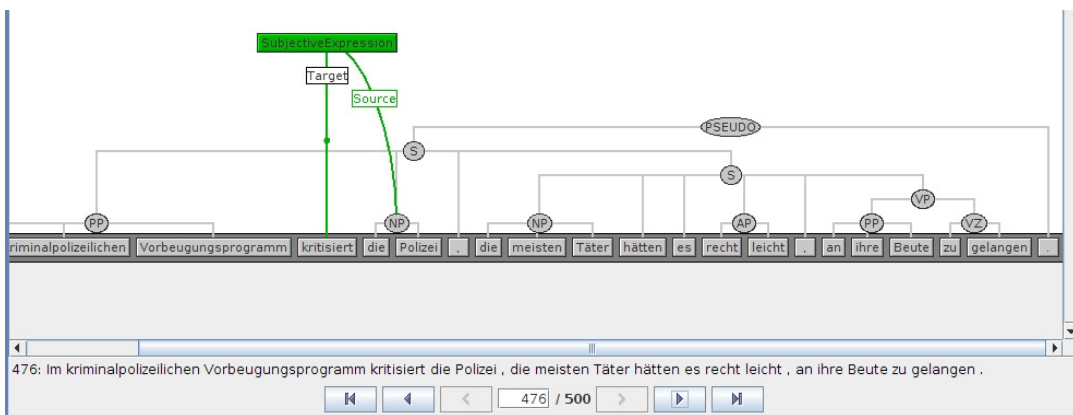
(a) start: given an opinion predicate (*kritisiert*) and the head of its source (*Polizei*)



(b) find lowest common ancestor node (*node underlined in yellow*)



(c) find direct descendant of lowest common ancestor also dominating head of source (*node underlined in violet*)



(d) final frame structure for opinion predicate and its source phrase

Figure 5: Illustration of how phrases are found for heads.

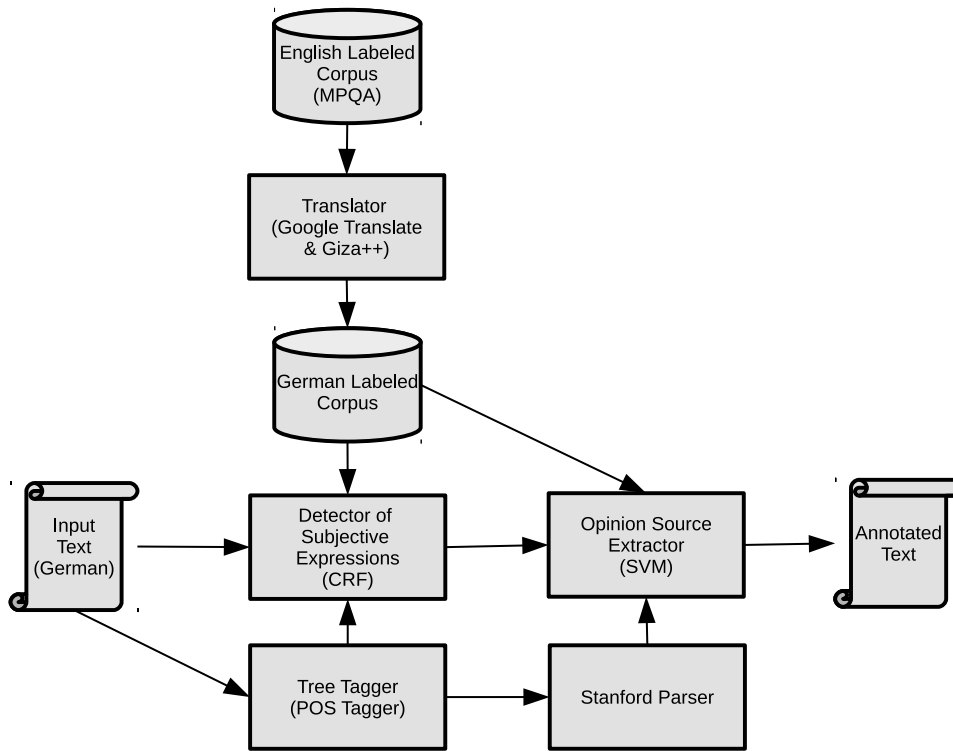


Figure 6: Processing pipeline of the translation-based system.

Type	Feature Templates
words	unigram features: target word and its two predecessors/successors bigrams features: bigrams of neighbouring words from unigram features
part of speech	unigram features: part-of-speech tag of target word and its two predecessors/successors bigram features: bigrams of neighbouring part-of-speech tags from unigram features bigram features: trigrams of neighbouring part-of-speech tags from unigram features
sentiment lexicon	is either of the words (window is that of the unigram features) a sentiment expression acc. to sentiment lexicon

Table 2: Feature templates employed for the CRF classifier to detect subjective expressions.

Type	Feature Templates
noun phrase	phrase label of noun phrase (e.g. <i>NP</i> , <i>MPN</i> , <i>PPER</i> etc.) words in phrase grammatical function if present (e.g. <i>SUBJ</i> , <i>OBJA</i> etc.)
sentiment expression	words in phrase part-of-speech tag of head of phrase
relational	distance between noun phrase and sentiment information

Table 3: Feature templates employed for the SVM classifier to detect opinion sources.

Run	Properties
Run 1	rule-based system combined sentiment lexicon dependency-parse normalization person filtering
Run 2	rule-based system combined sentiment lexicon
Run 3	rule-based system single sentiment lexicon dependency-parse normalization person filtering
Run 4	rule-based system single sentiment lexicon
Run 5	translation-based system only extracts sources

Table 4: The different properties of the different runs.

systems have been built in parallel. In particular, the superior dependency-parse normalization from the rule-based system was not implemented when that information was required for the translation-based system.

4 Experiments

In this section, we evaluate the five runs officially submitted to the shared task. Table 4 displays the different properties of the different runs. Runs 1-4 are rule-based systems, while Run 5 is a translation-based system. Runs 1 and 2 employ a large sentiment lexicon, being the concatenation of the sentiment lexicon of the PolArt system (Klenner et al., 2009) and SentiWS (Remus et al., 2010). Runs 3 and 4 are identical to Runs 1 and 2, respectively, with the exception that they only employ the sentiment lexicon of the PolArt system. Runs 1 and 3 employ normalization of the dependency parse output (Section 2.1) and person filtering for opinion sources (Section 2.3). Runs 2 and 4 neither contain normalization of the dependency parse output nor person filtering.

Table 5 displays the performance of the different configurations. **SE** evaluates the detection of subjective expressions. **Source** evaluates the detection of opinion sources, while **Source.SE** evaluates the detection of opinion sources given a correct match of subjective expression between system output and gold standard. Similarly, **Target** evaluates the detection of opinion targets, while **Target.SE** evaluates the detection of opinion targets given a correct match of subjective expression between system output and gold standard. As

there is no adjudicated gold standard but 3 individual annotations provided by the different annotators for each sentence, all numbers displayed in Table 5, i.e. *precision*, *recall* and *f-score*, are the average between the system output and each of the 3 annotators’ gold standards.

Table 5 shows that, on the detection of subjective expressions (SE), the combined sentiment lexicon (Runs 1 and 2) outperforms the single lexicon (Runs 3 and 4), however, the latter produces a better precision. Surprisingly, the best precision is achieved by the translation-based system (Run 5). This is most likely due to the fact that this system may be able to disambiguate subjective expressions. All rule-based systems consider each occurrence of a subjective expression in their respective sentiment lexicon as a case of a genuine sentiment.

On both the extraction of opinion sources and targets (Source and Target), the rule-based systems carrying out normalization and person filtering (Runs 1 and 3) outperform the systems without this type of processing (Runs 2 and 4). The rule-based system with the small lexicon (Run 3) outperforms its counterpart with the large lexicon on the tasks Source.SE and Target.SE since in that task, the detection of subjective expressions as such is not evaluated.

5 Conclusion

We reported on the two systems we devised for the German Shared Task on Task 1 of the German Sentiment Analysis Shared Task, the task on *Source, Subjective Expression and Target Extraction from Political Speeches (STEPS)*. The first system is a rule-based system relying on a predicate lexicon specifying extraction rules for verbs, nouns and adjectives, while the second is a translation-based system that has been obtained with the help of the MPQA corpus.

The rule-based system benefits from some linguistic processing and a large sentiment lexicon. Currently, the translation-based system is outperformed by the rule-based approach, however, there needs to be a more thorough evaluation in order to make qualified statements as to which approach is more effective for the given task. In addition, there is still plenty of space of improving either of the two approaches.

Run	Measure	SE	Source	Source_SE	Target	Target_SE
Run 1	Prec	56.83	44.35	73.16	50.40	79.57
	Rec	36.21	13.73	37.23	19.68	53.31
	F	44.24	20.97	49.35	28.31	63.85
Run 2	Prec	56.89	35.88	62.15	51.77	80.62
	Rec	35.97	13.06	35.64	14.87	40.58
	F	44.07	19.15	45.30	23.11	53.98
Run 3	Prec	63.42	48.55	74.89	56.25	79.71
	Rec	26.10	11.32	42.46	15.60	58.00
	F	36.98	18.36	54.19	24.43	67.14
Run 4	Prec	63.62	41.86	66.12	55.59	79.28
	Rec	25.80	10.98	41.68	11.74	44.19
	F	36.71	17.39	51.13	19.38	56.75
Run 5	Prec	80.56	47.98	58.55	N/A	N/A
	Rec	29.97	10.44	32.65	N/A	N/A
	F	43.69	17.14	41.92	N/A	N/A

Table 5: Evaluation of the different runs

Acknowledgements

This work was partially supported by the German Research Foundation (DFG) under grant WI 4204/2-1.

References

- Cem Akkaya, Janyce Wiebe, and Rada Mihalcea. 2009. Subjectivity Word Sense Disambiguation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 190–199, Singapore.
- Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 449–454, Genoa, Italy.
- Manaal Faruqui and Sebastian Padó. 2010. Training and Evaluating a German Named Entity Recognizer with Semantic Generalization. In *Proceedings of KONVENS*, Saarbrücken, Germany.
- Birgit Hamp and Helmut Feldweg. 1997. GermaNet - a Lexical-Semantic Net for German. In *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 9–15, Madrid, Spain.
- Thorsten Joachims. 1999. Making Large-Scale SVM Learning Practical. In B. Schölkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*, pages 169–184. MIT Press.
- Soo-Min Kim and Eduard Hovy. 2006. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In *Proceedings of the ACL Workshop on Sentiment and Subjectivity in Text*, pages 1–8, Sydney, Australia.
- Manfred Klenner, Angela Fahrni, and Stefanos Petrakis. 2009. PolArt: A Robust Tool for Sentiment Analysis. In *Proceedings of the Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 235–238, Odense, Denmark.
- John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the International Conference on Machine Learning (ICML)*, Williamstown, MA, USA.
- George Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1990. Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3:235–244.
- Franz Josef Och and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19–51.
- Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. 2006. Learning Accurate, Compact, and Interpretable Tree Annotation. In *Proceedings of the International Conference on Computational Linguistics and Annual Meeting of the Association for Computational Linguistics (COLING/ACL)*, pages 433–440, Sydney, Australia.
- Anna Rafferty and Christopher D. Manning. 2008. Parsing Three German Treebanks: Lexicalized and Unlexicalized Baselines. In *Proceedings of the ACL Workshop on Parsing German (PaGe)*, pages 40–46, Columbus, OH, USA.
- Robert Remus, Uwe Quasthoff, and Gerhard Heyer. 2010. SentiWS – a Publicly Available German-language Resource for Sentiment Analysis. In *Proceedings of the Conference on Language Resources*

- and Evaluation (LREC)*, pages 1168–1171, Valletta, Malta.
- Josef Ruppenhofer, Swapna Somasundaran, and Janyce Wiebe. 2008. Finding the Source and Targets of Subjective Expressions. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 2781–2788, Marrakesh, Morocco.
- Josef Ruppenhofer, Roman Klinger, Julia Maria Struß, Jonathan Sonntag, and Michael Wiegand. 2014. IGGSA Shared Tasks on German Sentiment Analysis (GESTALT). In G. Faaß and J. Ruppenhofer, editor, *Workshop Proceedings of the 12th Edition of the KONVENS Conference*, Hildesheim, Germany, October. Universität Hildesheim.
- Rico Sennrich, Gerold Schneider, Martin Volk, and Martin Warin. 2009. A New Hybrid Dependency Parser for German. In *Proceedings of the German Society for Computational Linguistics and Language Technology (GSCL)*, pages 115–124, Potsdam, German.
- Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation*, 39(2/3):164–210.
- Michael Wiegand and Dietrich Klakow. 2012. Generalization Methods for In-Domain and Cross-Domain Opinion Holder Extraction. In *Proceedings of the Conference on European Chapter of the Association for Computational Linguistics (EACL)*, pages 325–335, Avignon, France.