Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019)

A Descriptive Analysis of a German Corpus Annotated with Opinion Sources and Targets

Michael Wiegand[*†], **Leonie Lapp**[†], **Josef Ruppenhofer**[‡]
[*]Leibniz ScienceCampus, Heidelberg/Mannheim, Germany
[†]Spoken Language Systems, Saarland Informatics Campus, Saarland University, Saarbrücken, Germany
[‡]Leibniz Institute for German Language, Mannheim, Germany
{wiegand|ruppenhofer}@ids-mannheim.de
leonie.lapp@lsv.uni-saarland.de

Abstract

We present a descriptive analysis on the two datasets from the shared task on Source, Subjective Expression and Target Extraction from Political Speeches (STEPS), the only existing German dataset for opinion role extraction of its size. Our analysis discusses the individual properties of the three components, subjective expressions, sources and targets and their relations towards each other. Our observations should help practitioners and researchers when building a system to extract opinion roles from German data.

1 Introduction

While there has been much research in sentiment analysis on typical text classification tasks, such as subjectivity detection, polarity classification and emotion classification, there has been notably less work on opinion role extraction. This particularly concerns research done on languages other than English. In opinion role extraction, we distinguish between *opinion source extraction*, where the entities expressing an opinion, i.e. the opinion sources, are to be extracted, and *opinion target extraction*, where the task is to extract the entities or propositions at which sentiment is directed, i.e. the opinion targets. For example, in (1) the subjective expression *criticizes* has as its source *Switzerland* and as its target *North Korea*.

(1) [Switzerland $_{\text{SOURCE}}$] **criticizes** [North Korea $_{\text{TARGET}}$].
(2) [The opposition $_{\text{SOURCE}}$] **claims** [that the health service is getting fewer resources $_{\text{TARGET}}$].

In this paper, we address opinion role extraction on German data. Research on this specific task and language has been kicked off by the *shared task on Source, Subjective Expression and Target Extraction from Political Speeches (STEPS)* with its two editions from 2014 (Ruppenhofer et al., 2014a) and 2016 (Ruppenhofer et al., 2016). We present a descriptive analysis of the two datasets from this shared task that serve as a gold standard for opinion role extraction on German. Our aim is not to produce a classifier to automatically extract opinion sources and targets. Instead, we look into the properties of this gold standard in order to guide researchers and practitioners who intend to build such a classifier. Our analysis should largely influence the choice of classifiers, particularly the underlying feature set that describes potential opinion roles.

The focus of our analysis is on the structure of the opinion frame (§3), i.e. the linguistic structure that relates opinion source and target to its subjective expression. For each of these three linguistic components (subjective expression, opinion source and opinion target), we look at their individual linguistic forms and also their (syntactic) relation towards each other. Since STEPS consists of two datasets, i.e. the editions from 2014 and 2016, we also compare in how far the observed properties between the two different datasets differ. Given that each dataset individually is very small (i.e. only about 600 annotated sentences) the combination of the two datasets is a desirable step when building a classifier for opinion role extraction. Only if the two datasets are compatible to a large degree, can they be used for building a single application.

For general accessibility, we will always provide English examples when German and English follow the same linguistic pattern. Since, to the best of our knowledge, this is the first descriptive corpus-based study for opinion role extraction in general, we believe that our insights may be relevant to research beyond the German language.

Syntactic information plays a significant role in opinion role extraction, particularly, dependency relations. In this work, we consider dependency parses produced by ParZu (Sennrich et al., 2009). We consider this parser since it is also the dependency parser which the organizers of STEPS employ in the release of their dataset.

2 Related Work

So far, work on opinion role extraction has mostly been carried out on English data, especially the MPQA corpus (Wiebe et al., 2005), the standard corpus for fine-grained sentiment analysis. There has also been a related shared task on the topic: the Sentiment Slot Filling track (SSF) that was part of the Shared Task for Knowledge Base Population of the Text Analysis Conference (TAC) (Mitchell, 2013). For Japanese and Chinese some comparable data have been created as part of the NTCIR Opinion Analysis Task (Seki et al., 2007; Seki et al., 2008; Seki et al., 2010).

To the best of our knowledge, the only descriptive analysis of opinion role extraction was presented by Ruppenhofer et al. (2008). The major difference to our work is that Ruppenhofer et al. (2008) enumerate linguistic phenomena involved in opinion role extraction without reference to some existing datasets. Since we examine a labeled corpus, our main contribution is that we quantify the linguistic phenomena involved.

For German sentiment analysis, there exist quite a few different corpora ranging from sentiment aspect classification (Säger et al., 2016; Wojatzki et al., 2017) to much more fine-grained tasks, such as attitude classification (Klenner et al., 2017). Apart from STEPS, however, there only exists the MLSA corpus (Clematide et al., 2012) with annotation of both opinion holders and targets on German text. The annotation scheme of STEPS was mainly inspired by *Layer 3: Expression-level Annotation* of MLSA. (The same researchers who annotated that layer of MLSA also created the two STEPS datasets.) The reason we conduct our study on the STEPS corpus rather than on the MLSA corpus is that the two STEPS corpora totaling about 1,200 sentences are significantly larger than the MLSA corpus with only 270 sentences.

3 Opinion Frames

In STEPS, opinion roles are represented as *opinion frames*. An opinion frame is a triple <*subjective expression*, *opinion source*, *opinion target*>. The *subjective expression* is a word or phrase which conveys some opinion, its *source* is the entity that expresses that opinion, and its *target* is the entity or proposition towards which that opinion is directed. In this paper, subjective expressions will always be indicated by bold type font in examples and abbreviated by the acronym SE in the prose.

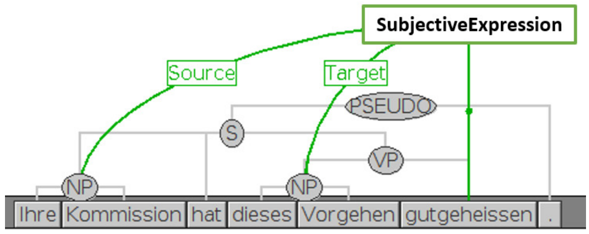


Figure 1: Illustration of an opinion frame.

Each opinion frame has exactly one SE and at most one source and target each. In other words, there can be opinion frames without a source (3), without a target (4) or without both (5).

(3) I don't understand this **interest** [in weapons $_{\text{TARGET}}$].
(4) [Peter $_{\text{SOURCE}}$] was so **unhappy** that he immediately left the party.
(5) **Altruism** is not a very common thing in our society.

4 IGGSA-STEPS: Shared Task on Source and Target Extraction from Political Speeches

For our experiments we employ the labeled datasets from the *Shared Task on Source and Target Extraction from Political Speeches*. In that shared task, German language speeches from the Swiss parliament were annotated with opinion frames. In German, there exists no comparable dataset of similar size for opinion role extraction. The data have been annotated in TIGER/SALSA XML (Erk and Padó, 2004), a format originally devised for representing frame-structures from FrameNet (Baker et al., 1998). This representation format combines syntactic constituency parses with some semantic annotation. Like FrameNet-frames, opinion frames represent semantic structures that build upon syntactic structures. Figure 1 illustrates the structure of a typical opinion frame.

There are two editions of the shared task (Ruppenhofer et al., 2014b; Ruppenhofer et al., 2016). For STEPS 2016, the STEPS 2014 dataset was revised in order to be compatible with the new annotation scheme introduced for STEPS 2016. We use this revision of the STEPS 2014 dataset. Another advantage of using the dataset from the revised annotation scheme is that it has been shown to produce a sufficiently high interannotation agreement (Ruppenhofer et al., 2016).

Table 1 displays some general statistics of the two datasets. The table already indicates that there

property	freq STEPS 2014	STEPS 2016
no. of sentences	605	581
avg. no. of tokens in sentence	22.58	24.08
no. of subjective exprs. (SEs)	2105	2166
avg. no. of SEs in sentence	3.58	3.94
no. of opinion frames	2228	2417
no. of sources	997	1064
no. of targets	1608	1770

Table 1: Statistics of the two STEPS datasets.

dataset	types	singletons	MWEs
STEPS 2014	1115	805	213
STEPS 2016	1110	769	214

Table 2: Statistics of subjective expressions (SEs).

dataset	verb	adj	noun	adv	other
STEPS 2014	270	206	418	18	227
STEPS 2016	280	207	405	27	224

Table 3: POS-Distribution of SEs (*types are counted*).

are no significant differences in the frequency of the different major constructions between STEPS 2014 and STEPS 2016.

5 Subjective Expressions (SEs)

The first step in opinion role extraction is to determine which words represent SEs. Table 2 provides some statistics about this linguistic entity. The most notable observation is that many SEs are singletons. (This ratio does not change much even if we merge the two datasets STEPS 2014 and STEPS 2016.) This is highly relevant for building a classifier to detect SEs. If most SEs only occur once in a gold standard, then they can hardly be learnt from this data directly. Instead, some form of sentiment lexicon listing SEs should be used. However, by computing the coverage of the SEs in the standard sentiment lexicon for German, the PolArt lexicon (Klenner et al., 2009), we found that only a small proportion (i.e. 25%) is actually covered.

With about 19% of the vocabulary, multiword expressions (MWEs) represent a considerable share in the set of SEs. About 75% are MWEs that consist of exactly 2 tokens, which, in most cases, are phrasal verbs, e.g. *tritt ab (stands down)*, *denkt nach (considers)*. Compared to idioms, e.g. *in den sauren Apfel beißen (to bite the bullet)*, which due to their free word order in German can have many different surface realizations (Wiegand et al., 2016a), phrasal verbs are relatively easy to detect.

Table 3 lists the distribution of the different parts of speech among the SEs. To our surprise, nouns are the most frequent type of SEs. One typically associates sentiment with adjectives (e.g. *bad*, *nice*) or verbs (e.g. *adore*, *hate*) and, therefore, one would expect a higher proportion of these parts of speech. The high frequency of subjective nouns can be explained by the fact that many of these nouns are nominalized adjectives (e.g. *badness*) and nominalized verbs (e.g. *hatred*). Additionally, many subjective nouns are some form of compound, e.g. *Bombenattentat (bombing attack)* or *Expertenmeinung (expert advice)*. Wiegand et al. (2016b) state that every other sentence in STEPS 2014 contains a noun compound.[1] A noun compound (*Bombenattentat*) typically consists of two constituents, a modifier (*Bombe*) and a head (*Attentat*). Noun compounds are very productive. In principle, noun heads may combine with a large number of different noun modifiers (e.g. *Bombenattentat*, *Selbstmordattentat*, *Flugzeugattentat*, *Sprengstoffattentat*, *Säureattentat* etc.) This results in a large number of different compounds in STEPS (please keep in mind that in Table 3, we count types and not tokens). Each of these compounds only occurs once or twice on average which explains the high number of noun types, particularly singleton nouns in STEPS.

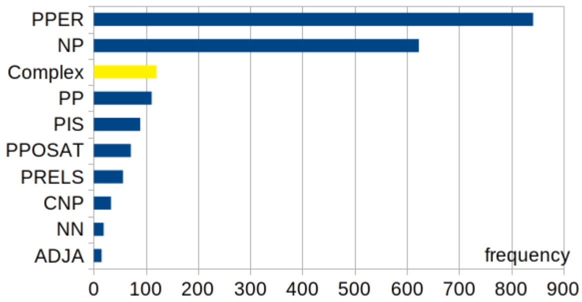
In order to detect SEs automatically and given the large number of sparse noun compounds on both datasets, some form of noun normalization would be advisable. Only the head of a noun compound is relevant for detecting SEs, i.e. *Attentat (attack)* in *Bombenattentat (bombing attack)*. We, therefore, anticipate a higher coverage of matching SEs in a sentiment lexicon by reducing noun compounds to their heads.

If one pursues a lexicon-based approach to detect SEs, not only is a lexicon sought that has a good coverage, one should also keep the reliability of the entries in mind. Table 4 compares the precision of an oracle lexicon, i.e. a lexicon comprising all words being labeled at least once as an SE expression in either of the two editions of STEPS, with the precision of the words in the PolArt sentiment lexicon. (We evaluate here on the concatenation of STEPS 2014 and STEPS 2016. In the following sections, we always merge the distributions of STEPS 2014 and STEPS 2016 whenever there was not sufficient space and we did not observe any significant difference between the two datasets.)

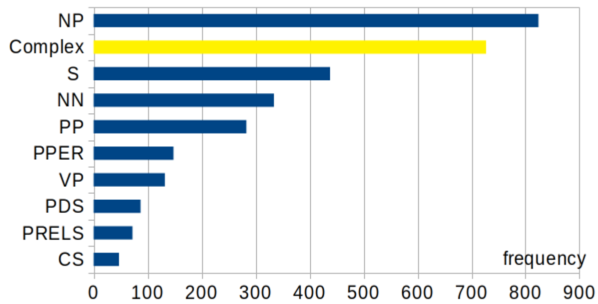
[1] We also confirm a similar proportion in STEPS 2016.

lexicon	Prec
union of words being labeled as SE at least once	72.1
PolArt lexicon	89.9

Table 4: Precision of different lexicons.



(a) sources



(b) targets

Figure 2: Distribution of phrase labels.

Although the PolArt sentiment lexicon has a low coverage of SEs in STEPS, the entries that match that lexicon are fairly reliable. A lexicon with a full coverage of SEs (as our oracle lexicon) would not solve the problem of detecting SEs, as a large proportion of SEs are ambiguous words. One additionally would have to devise a subjectivity word-sense disambiguation (Akkaya et al., 2009) once a word within a sentence has been matched with that lexicon. The task would be to decide whether an ambiguous word, such as *alarm* is used in a subjective sense, as in (6), or in a non-subjective one, as in (7). If no reliable disambiguation is possible, a sentiment lexicon may still be a good solution because of its high precision.

(6) When he heard that particular news, his **alarm** grew even more.

(7) Our new smoke detector is malfunctioning. The alarm went off twice yesterday although there was no smoke.

6 Inherent Properties of Opinion Roles

We now examine properties of opinion sources and targets. We start by looking at inherent properties. Figure 2 compares the phrase label distribution of opinion sources and targets. Typically, both source and target exactly match one phrase node in the constituency parse tree representing the sentence (Figure 1). We introduce a phrase label *Complex* by which we subsume all cases in which an opinion role does not match exactly one constituent. A large fraction of those instances will be parse errors.[2]

Figure 2 shows that sources and targets have notably different phrase labels. Opinion sources are mostly noun phrases (NP) or personal pronouns. This result is quite intuitive. Opinion sources can only be persons or groups of persons as other types of entities typically do not have any specific sentiment. The fact that prepositional phrases are also frequent can be explained by passive constructions (9) in which the opinion source is realized as a prepositional phrase rather than an NP which is the case in the more canonical active constructions (8).

(8) [Peter SOURCE] **loves** [Mary TARGET]^{NP}.

(9) [Mary TARGET] is **loved** [by Peter SOURCE]^{PP}.

Opinion targets, on the other hand, are much more heterogeneous. Targets do not have to be entities. They can also represent entire propositions. This explains why other constituents, such as (complement) sentences (10) or verbal phrases (11), are also frequently labeled as targets.

(10) [Peter SOURCE] **thinks** [that Mary should work harder TARGET]^{S}.

(11) [Peter SOURCE] **wants** [to go shopping TARGET]^{VP}.

It is also surprising that the second most frequent phrase label is *Complex*. By manually inspecting these cases, we found that in most of them there was an error in the parse. Targets representing entire propositions are typically much longer phrases (i.e. they comprise more tokens) than source-phrases representing simple entities. Figure 3 illustrates the different token lengths of sources and targets. It confirms that sources tend to be shorter than targets. The long phrases that represent targets, such as sentences or verbal phrases,

[2] The constituency-parse trees in STEPS have been automatically generated by the Berkeley parser (Petrov and Klein, 2007). In case of parsing errors, the annotators were instructed to label those spans of text that they thought represent the correct span. This often meant that one opinion role was assigned more than one phrase node in the constituency-parse tree.

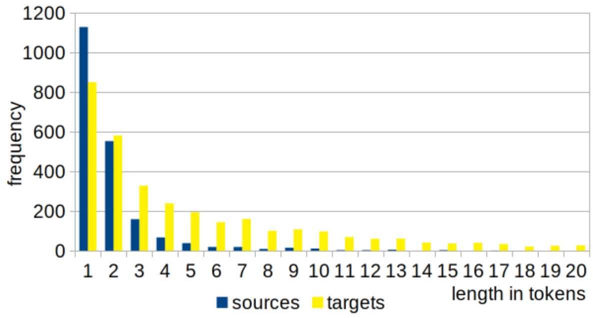


Figure 3: Distribution of phrase length.

pos	relation path	ratio [%]		freq
		sources	targets	
verb	↑subj	81.2	18.8	479
	↓aux_↑subj	67.9	32.1	324
	↑obja	7.7	92.3	221
	↑pp	12.6	87.4	87
	↑objd	21.2	78.8	52
adj	↓attr	1.3	98.7	235
	↓pred_↑subj	30.9	69.1	55
	↓aux_↑subj	69.2	30.8	26
	↓adv_↑subj	33.3	66.7	21
	↑pp	66.7	33.3	12
noun	↑gmod	38.4	61.6	199
	↑pp	15.2	84.8	79
	↓obja_↑subj	68.9	31.1	74
	↑det	76.9	23.1	65
	↑attr	48.0	52.0	25

Table 5: Ambiguity of dependency-relation paths between sources and targets.

are also much more likely to be affected by parsing errors.

7 Opinion Roles and Their Relation towards SEs

We continue our examination of opinion roles by looking at the relation between opinion roles to the SEs they evoke. We start by looking into syntactic relationships.

Table 5 lists the 5 most frequent dependency-relation paths observed between the individual opinion roles and the SEs they evoke. The table lists the paths for subjective verbs, adjectives and nouns each. It suggests that some dependency-relation paths are predictive for either sources or targets. For example, for subjective verbs, sources are often realized as subjects (↑*subj*) while targets are realized as accusative objects (↑*obja*), as illustrated by (12). For subjective attributive adjectives, one can very reliably predict targets by looking for the noun that modifies them (↓*attr*) as illustrated by (13).

(12) [Mary $_{\text{SOURCE}}$]^{subj} **likes**$_{verb}$ [Peter's new flat $_{\text{TARGET}}$]^{obj}.
(13) I just saw a **beautiful**$^{attr}_{adj}$ [rainbow $_{\text{TARGET}}$].

However, there are also relation paths that are ambiguous. The most notable example is the genitive modifier of subjective nouns (↑*gmod*) which is the most frequent dependency relationship connecting both opinion sources (14) and targets (15). This analysis proves that opinion roles cannot be extracted exclusively on the basis of dependency-relation paths.

(14) Das entsprach auch der **Sichtweise**$_{noun}$ [der meisten Bürger $_{\text{SOURCE}}$]^{gmod}.
(*This was also the view of most citizens.*)
(15) Er hob die **Einfachheit**$_{noun}$ [des Ansatzes $_{\text{TARGET}}$]^{gmod} hervor.
(*He emphasized the simplicity of that approach.*)

One major obstacle in processing German text is the high degree of errors in automatic syntactic parse analyses. The longer a sentence is the more likely errors in syntactic parsing occur. This certainly is an issue with STEPS 2014 and STEPS 2016 since both datasets contain long sentences (between 22 and 24 tokens on average, see also Table 1). The ParZu parser may fail to produce a fully connected dependency tree for long sentences. Instead only a set of partial trees are produced. In that case, there is often no dependency-relation path available that connects opinion roles with the SEs they evoke.

For all pairs of <*opinion role*, *SE*> where the members of the pair are separated by a specified token distance, Figure 4 shows the proportion of pairs for which there is no connecting dependency path available in the output of ParZu. The figure shows that the longer the token distance is the higher the proportion of pairs are that have no dependency-relation path. Even for pairs with a short token distance, there is still a considerable number of pairs for which there is no dependency-relation path. All in all, this analysis underlines that errors in the syntactic parse output will have an impact on classification performance.

As an alternative to syntactic dependency-relation paths, we also examine whether the order of pairs <*opinion role*, *SE*> is predictive for this task. Unlike syntactic information, information about the sequential order of two constituents is not dependent on the output of a syntactic parser. Table 6 displays the ratio of different orders. The table shows that there may be certain correlations between certain orders. For example, the source mostly precedes subjective verbs or adjec-

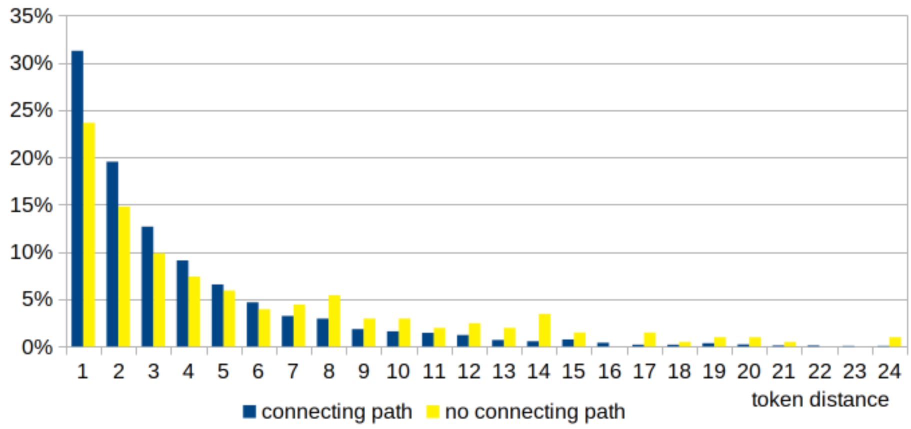


Figure 4: Connecting paths between opinion role and SE for different token lengths.

	STEPS 2014			STEPS 2016		
order	verb	adj	noun	verb	adj	noun
<source, SE>	81.7	91.1	62.0	83.4	97.6	55.9
<SE, source>	18.3	8.9	38.0	16.6	2.4	44.1
<target, SE>	67.5	24.2	32.1	66.8	36.5	30.9
<SE, target>	32.5	75.8	67.9	33.2	63.5	69.1

Table 6: Order of opinion role (i.e. source or target) and SE (in percentage).

tives. However, in the case of targets, these correlations are less pronounced. The general lack of a predictive sequential order can be explained by the fact that depending on tense or sentence type, the order between different constituents may vary. For instance, the canonical order for subjective verbs <*SE*, *opinion target*> as can be observed in a present tense main clause (16) changes if that sentence is shifted into present perfect (17) or a subordinate clause (18).

(16) Peter **hasst**$_{verb}$ [Julia $_{\text{TARGET}}$]^{obja}.
(*Peter hates Julia.*)

(17) Er hat schon immer [Julia $_{\text{TARGET}}$]^{obja} **gehasst**$_{verb}$.
(*Peter has always hated Julia.*)

(18) ... weil Peter [Julia $_{\text{TARGET}}$]^{obja} **hasst**$_{verb}$.
(... *because Peter hates Julia.*)

However, in all of these cases, the dependency relation between subjective expression and opinion target remains the same ($\uparrow$*obja*). This example illustrates that, in principle, syntactic dependency relations are more expressive than sequential order.

8 Frame Structure Configurations

According to the definition of opinion frames (§3), the presence of both source and target is not obligatory. We want to examine how often the opinion frame structure deviates from the canonical form in which both source and target are present. Table 7 lists the frequency of different frame structure configurations. The table clearly shows that both in STEPS 2014 and STEPS 2016 there is a significant number of frames that do not include both source and target. This observation is quite important since it suggests that joint modelling of opinion source and target using simple constraints of the type *an opinion frame has to comprise exactly one opinion source and one opinion target* would not work.

In Table 7, we also observe that partial frames with only a target are much more frequent than the frames with only a source. We ascribe this large amount to the so-called *implicit* sources. Implicit sources are sources without a concrete surface realization (19). They typically represent the speaker of the utterance in which the opinion frame is evoked. Strictly speaking, therefore, frames with such a source are not partial frames. These frames just lack an *explicit* source, that is, a constituent in the sentence in which the SE occurs which has been annotated as an opinion source. Whether an SE comes with an explicit or implicit source largely depends on the SE itself. In other words, it is a lexical property of SEs. For English, Wiegand et al. (2016c) developed methods to distinguish whether an SE is more likely to have explicit or implicit sources. While SEs with a tendency for implicit sources are called *speaker-views* SEs, SEs with a tendency for explicit sources are referred to as *actor-views* SEs. Most of these methods should be largely reproducible on German language data.

Apart from opinion frames with implicit sources, there may, of course, also be opinion frames lack-

	STEPS dataset	
frame structure	**2014**	**2016**
frames with source and target	845	850
frames with only source	152	214
frames with only target	763	920
frames with neither source nor target	468	433

Table 7: Distribution of different frame structures.

ing both an explicit and implicit source. An example of the latter type is (20). It does not contain an explicit source and from the context it is clear that it is not the speaker of the utterance either since the speaker (represented by *I*) explicitly distances themselves from the interest in weapons. Therefore, the exact source remains unspecified.

(19) [The reasons for voting to leave the EU TARGET] are **obvious**.

(20) I don't understand this **interest** [in weapons TARGET].

Figure 5 compares the distribution of frames without source and frames without target across SEs with different parts of speech. While for verbs, we observe fewer frames that exclude either source or target, we observe that for nouns and adjectives partial frame structures are much more frequent. (This also matches our previous examples (3)-(5).) Particularly, most frames without a target are evoked by subjective nouns. The fact that mostly adjectives and nouns are likely to form partial opinion frames might be explained with the help of subcategorization. Although the subcategorization frames of verbs and nouns can be similarly complex (for instance, both the subjective verb in (21) and the subjective noun in (22) have two arguments), for verbs the realization of its arguments is usually obligatory in order to make a sentence grammatical (cp. (21) with (23)). For nouns (and adjectives follow similarly), however, it is quite often the case that they come with fewer arguments than their valency suggests (24). The fewer arguments a subjective expression has, the more likely partial frames are to be evoked. (24) contains a partial frame lacking a target.

(21) [Mary SOURCE]^{subj} **loves**$_{verb}$ [Peter$_{\text{TARGET}}$]^{obj}.

(22) Everyone knew about [Mary's SOURCE]^{gmod} **love**$_{noun}$ [to Peter$_{\text{TARGET}}$]^{pobj}.

(23) ?[Mary SOURCE]^{subj} **loves**$_{verb}$.

(24) In public, only few people talk about [Mary's SOURCE]^{gmod} **love**$_{noun}$.

9 Inferred Sources

Most opinion roles are syntactic dependents of the SE by which they are evoked. For instance, the

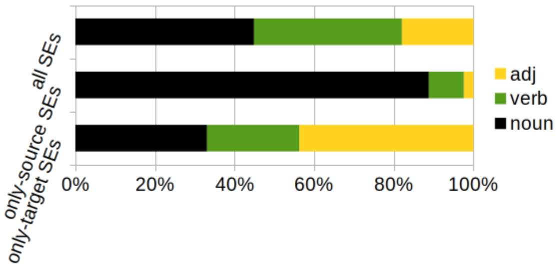


Figure 5: POS-Distribution of partial frames.

source of *like* in (25) is its subject. In STEPS, there is a special subset of sources, referred to as *inferred sources*. By that we understand sources that are not associated with any of the syntactic dependents of its SE (26). (In (26), the SE *impressive* has only one syntactic dependent which is its subject.) These sources are called *inferred* since from the subcategorization frame of the SE, we cannot conclude their presence. This makes them more difficult to detect than normal sources.

(25) [Mary SOURCE]^{subj} **likes**$_{verb}$ [Peter's new flat TARGET]^{obj}.

(26) [Mary INFERRED_SOURCE] said [Peter's new flat TARGET]^{subj} was **impressive**$_{adj}$.

26% of the opinion sources in STEPS 2014 and STEPS 2016 have been flagged as inferred sources by the annotators. Since this is a substantial amount, we want to investigate whether we can further characterize this subset of sources. If we look at the distribution of parts of speech of the SEs evoking inferred sources (Figure 6), we find that there is a notable difference to the general part-of-speech distribution of SEs. While the proportion of nouns remains fairly constant, there is a disproportionately high amount of inferred sources with subjective adjectives. For SEs being verbs, the proportion of inferred sources, on the other hand, is fairly low.

We assume that the valency of the individual SEs is responsible for that distribution. The prototypical (subjective) adjective has one syntactic argument, for example, a subject (27) which is its target. There is no argument position for the opinion source and therefore, the source is the implicit speaker of the utterance.[3] However, if this SE is

[3] The source in this sentence is not unspecified, since *impressive* in (27) comes with all its obligatory syntactic arguments, i.e. its subject. According to Wiegand et al. (2016c) more than 90% of all subjective adjectives are *speaker-view* words, i.e. these are subjective expressions that tend to have implicit sources.

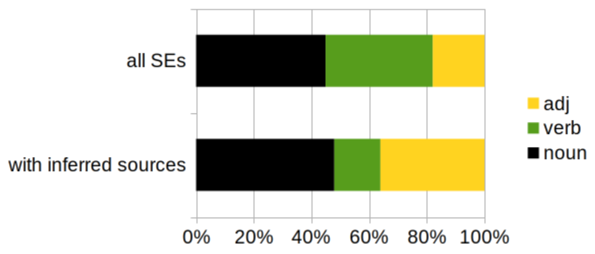


Figure 6: POS-Distribution of SEs with inferred sources.

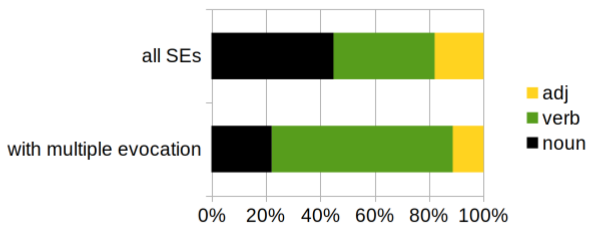


Figure 7: POS-Distribution of SEs with multiple frame evocation.

further syntactically embedded, as in (26), there may be some explicit source but it is inferred. For subjective verbs, unlike subjective adjectives, there is a syntactic argument in their subcategorization frame, typically their subject, as in (25), that is associated with their opinion source. Therefore, fewer inferred sources occur with subjective verbs.

(27) [Peter's new flat $_{\text{TARGET}}$]^{subj} is **impressive**$_{adj}$.

10 Multiple Frame Evocation

There are SEs that evoke more than one opinion source and target. In STEPS this is modeled by allowing the same SE to evoke more than one single opinion frame. For example, the verb *force* can evoke three different opinion frames at the same time as illustrated by (28)-(30). (28) describes the view that James has some request to someone. (29) describes the view of James towards walking the dog. Finally, (30) represents Alice's negative sentiment towards walking the dog (if she did not have that sentiment, James would not need to force her to do so).

(28) [James $_{\text{SOURCE}}$] **forced** [Alice $_{\text{TARGET}}$] to walk the dog.
(29) [James $_{\text{SOURCE}}$] **forced** Alice [to walk the dog $_{\text{TARGET}}$].
(30) James **forced** [Alice $_{\text{SOURCE}}$] [to walk the dog $_{\text{TARGET}}$].

12% of the SEs in STEPS evoke more than one opinion frame. Figure 7 shows the distribution of multiple frame evocation across SEs with different parts of speech. The statistic shows that by far most SEs evoking multiple frames are verbs. This can be explained by the fact that verbs have the most complex subcategorization frames (e.g. in (28)-(30) *force* has three different syntactic arguments). We assume that the more syntactic arguments a SE has in a sentence, the more likely there is some multiple frame evocation.

11 Conclusion

We presented a descriptive analysis of the STEPS 2014 and 2016 datasets, a resource for building and evaluating opinion role extraction systems in German. We found that the linguistic properties of the two datasets are very similar which means that they can be usefully merged into one resource. A large proportion of subjective expressions are nouns including noun compounds. The majority of subjective expressions are singletons. We assume that in order to increase the coverage of subjective expressions in lexical resources, such as as sentiment lexicons, more effectively, some noun normalization that reduces compounds to their heads may be helpful. Opinion sources and targets differ very much from each other. Opinion sources tend to be realized as (short) noun phrases, while opinion targets are long phrases of various types. For both opinion sources and targets there is a small set of characteristic dependency relationships towards the subjective expression they evoke. Conceptually speaking, dependency relationships are more predictive than sequential order. However, reliable syntactic information is difficult to produce since parsers for German are fairly error prone. STEPS includes a substantial number of inferred sources. Those subjective expressions that come with inferred sources have more often few syntactic arguments, such as adjectives. Subcategorization frames also play a role when it comes to partial opinion frames. Subjective expressions with very complex subcategorization frames, such as verbs, typically come with complete opinion frames unlike adjectives and nouns, which more often evoke partial opinion frames. There is also a significant number of subjective expressions that evoke multiple frames, however, this phenomenon is largely restricted to subjective verbs.

Acknowledgements

This research has been partially supported by the Leibniz ScienceCampus "Empirical Linguistics and Computational Modeling", funded by the Leibniz Association under grant no. SAS-2015-IDS-LWC and by the Ministry of Science, Research, and Art (MWK) of the state of Baden-Württemberg. The authors were also partially supported by the German Research Foundation (DFG) under grants RU 1873/2-1 and 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.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the International Conference on Computational Linguistics and Annual Meeting of the Association for Computational Linguistics (COLING/ACL)*, pages 86–90, Montréal, Quebec, Canada.

Simon Clematide, Stefan Gindl, Manfred Klenner, Stefanos Petrakis, Robert Remus, Josef Ruppenhofer, Ulli Waltinger, and Michael Wiegand. 2012. MLSA – A Multi-layered Reference Corpus for German Sentiment Analysis. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 3551–3556, Istanbul, Turkey.

Katrin Erk and Sebastian Padó. 2004. A powerful and versatile xml format for representing role-semantic annotation. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 799–802, Lisbon, Portugal.

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.

Manfred Klenner, Simon Clematide, and Don Tuggener. 2017. Verb-Mediated Composition of Attitude Relations Comprising Reader and Writer Perspective. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics (CICLing)*, Lecture Notes in Computer Science, pages 141–155, Budapest, Hungary. Springer.

Margaret Mitchell. 2013. Overview of the TAC2013 Knowledge Base Population Evaluation: English Sentiment Slot Filling. In *Proceedings of the Text Analysis Conference (TAC)*, Gaithersburg, MD, USA.

Slav Petrov and Dan Klein. 2007. Improved inference for unlexicalized parsing. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT/NAACL)*, pages 404–411, Rochester, NY, USA.

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, Marrakech, Morocco.

Josef Ruppenhofer, Roman Klinger, Julia Maria Struß, Jonathan Sonntag, and Michael Wiegand. 2014a. IGGSA Shared Tasks on German Sentiment Analysis (GESTALT). In G. Faaß and J. Ruppenhofer, editors, *Workshop Proceedings of the KONVENS Conference*, pages 164–173, Hildesheim, Germany. Universität Hildesheim.

Josef Ruppenhofer, Julia Maria Struß, Jonathan Sonntag, and Stefan Gindl. 2014b. IGGSA-STEPS: Shared Task on Source and Target Extraction from Political Speeches. *Journal for Language Technology and Computational Linguistics*, 29(1):33–46.

Josef Ruppenhofer, Julia Maria Struß, and Michael Wiegand. 2016. Overview of the IGGSA 2016 Shared Task on Source and Target Extraction from Political Speeches. In *Proceedings of IGGSA Shared Task Workshop*, Bochumer Linguistische Arbeitsberichte, pages 1–9, Bochum, Germany.

Mario Säger, Ulf Leser, Steffen Kemmerer, Peter Adolphs, and Roman Klinger. 2016. SCARE - The Sentiment Corpus of App Reviews with Fine-grained Annotations in German. In *Proceedings of the Conference on Language Resources and Evaluation (LREC)*, pages 1114–1121, Portorož, Slovenia.

Yohei Seki, David Kirk Evans, Lun-Wei Ku, Hsin-Hsi Chen, Noriko Kando, and Chin-Yew Lin. 2007. Overview of Opinion Analysis Pilot Task at NTCIR-6. In *Proceedings of the NTCIR-6 Workshop Meeting*, Tokyo, Japan.

Yohei Seki, David Kirk Evans, Lun-Wei Ku, Hsin-Hsi Chen, and Noriko Kando. 2008. Overview of Multilingual Opinion Analysis Task at NTCIR-7. In *Proceedings of the NTCIR-7 Workshop Meeting*, Tokyo, Japan.

Yohei Seki, Lun-Wei Ku, Le Sun, Hsin-Hsi Chen, , and Noriko Kando. 2010. Overview of Multilingual Opinion Analysis Task at NTCIR-8 - A Step Toward Cross Lingual Opinion Analysis. In *Proceedings of NTCIR-8 Workshop Meeting*, pages 209–220.

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, Germany.

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, Nadisha-Marie Aliman, Tatjana Anikina, Patrick Carroll, Margarita Chikobava, Erik Hahn, Marina Haid, Katja König, Leonie Lapp, Artuur Leeuwenberg, Martin Wolf, and Maximilian Wolf. 2016a. Saarland University's Participation in the Second Shared Task on Source, Subjective Expression and Target Extraction from Political Speeches (STEPS-2016). In *Proceedings of IGGSA Shared Task Workshop*, Bochumer Linguistische Arbeitsberichte, pages 14–23, Bochum, Germany.

Michael Wiegand, Christine Bocionek, and Josef Ruppenhofer. 2016b. Opinion Holder and Target Extraction on Opinion Compounds – A Linguistic Approach. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT/NAACL)*, pages 800–810, San Diego, CA, USA.

Michael Wiegand, Marc Schulder, and Josef Ruppenhofer. 2016c. Separating Actor-View from Speaker-View Opinion Expressions using Linguistic Features. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT/NAACL)*, pages 778–788, San Diego, CA, USA.

Michael Wojatzki, Eugen Ruppert, Sarah Holschneider, Torsten Zesch, and Chris Biemann. 2017. GermEval 2017: Shared Task on Aspect-based Sentiment in Social Media Customer Feedback. In *Proceedings of the GermEval 2017 – Shared Task on Aspect-based Sentiment in Social Media Customer Feedback*, pages 1–12, Berlin, Germany.