Stance Detection in Facebook Posts of a German Right-wing Party

Manfred Klenner and Don Tuggener and Simon Clematide
Computational Linguistics
University of Zurich, Switzerland
{klenner,tuggener,siclemat}@cl.uzh.ch

Abstract

We argue that in order to detect stance, not only the explicit attitudes of the stance holder towards the targets are crucial. It is the whole narrative the writer drafts that counts, including the way he hyposta- sizes the discourse referents: as benefac- tors or villains, as victims or beneficia- ries. We exemplify the ability of our system to identify targets and detect the writer's stance towards them on the basis of about 100 000 Facebook posts of a Ger- man right-wing party. A reader and writer model on top of our verb-based attitude extraction directly reveal stance conflicts.

1 Introduction

Recently, verb-based sentiment relation extraction has been used among others to derive positive and negative attitudes of Democrats and Republicans towards actors or topics. The system of Rashkin et al. (2016) accomplishes this task on the basis of crowd-sourced connotation frames of (transitive) verbs which indicate such relations. A connotation frame specifies, among others, the polar effects a verb role bears, if the verb is used affirmatively.

We are also interested in stance detection, but *stance*, in our model, is not only the attempt to identify the positive and negative attitudes of the writer of the text (the main opinion holder) to- wards *given* actors (e.g. political parties) or (con- troversial) topics (henceforth targets) (see e.g. Mo- hammad et al. (2016)). We also strive to *identify* targets in the first place. We claim that the way the writer conceptualizes actors, namely as polar- ized actors - benefactors, villains, victims, benefi- ciaries (and so on) - reveals who/what the targets are. This also unveils, as a by-product, the writer's stance. A writer might not directly call someone

a villain, but if he puts forward that a person has told a lie, then he obviously regards him as a vil- lain, which implies a negative attitude.

We propose the following model. The writer produces - under the assumption of truth com- mitment - some text. The reader, on the basis of shared (lexical) semantic knowledge, is able to identify what the text implies for the various tar- gets involved and described. The reader's per- sonal preferences (his stance, his moral values etc.) might be affected by a given exposition. He might agree with the (implications of the) pro- claimed facts or not. From what is being said, the reader is thus able to derive at least two things: How does the writer conceptualize the world (i.e. what is his stance, what are the targets) and how does this relate to the reader's stance. We focus on the interplay of these perspectives. Our model confronts the writer with the reader perspective. This way, conflicting conceptualizations of reality and incompatible stances become visible. This al- lows the reader to identify *charged* statements, i.e. main sources of disagreement.

We have implemented a system that predicts *advocate* and *adversary* attitudes and that further assigns sources and targets their polarized roles (benefactor, victim etc.) on the basis of a con- notation verb lexicon comprising 1500 manually specified connotation frames stemming from 1100 different verbs, also including about hundred nom- inalisations. In order to do so, event factuality in the sense of Saurí and Pustejovsky (2009) also needs to be coped with.

In this paper, we are interested in a qualita- tive validation of our approach. On the basis of 100 000 Facebook posts of a right-wing German party, the AfD (Alternative für Deutschland), and a virtual (kind of prototypical) reader, we exem- plify how conflicting perspectives can be identi- fied and how stance is detected.

31

2 Stance Detection: Ways to Go

Presumably, one finds directly expressed attitudes like *I hate Burger King* only in product reviews. In a political discourse, such aversions etc. are expressed more subtly. We are unlikely to find a sentence like *We, the AfD, are against refugees* in Facebook posts of AfD members. Then how can we get to know that this is apparently the case?[1] There are three ways how to identify the writer's stance.

1. Inference patterns (cf. exemplification E1). There are sentences, where *AfD* and *refugees* co-occur, but where the relation between them is given only indirectly.

E1: | If A0 is against an event that is good for A1, A0 is an adversary of A1 |

For example: *The AfD criticizes that refugees are tolerated by the German government.* Our model is able to derive an adversary relation between *AfD* and *refugee* from such complex sentences. The underlying inference pattern is: A negative attitude (*criticize*) of an opinion source (*AfD*) towards a situation (*tolerate*) that is positive for a target (*refugee*) means that an adversary relation holds between the two (*AfD* and *refugee*).

2. Inference chains (cf. E2).

Assume that our system derived that *refugee* is an adversary of *Germany*, and *AfD* is an advocate of *Germany* in the text. It then follows that *refugee* is an adversary of *AfD* and vice versa.

E2: | From: $A0_1$ is an adversary of A1
and: $A0_2$ is an advocate of A1
it follows that: $A0_1$ and $A0_2$ are adversaries |

We also tried to find cliques of accomplices of the main opinion source, here the AfD. Shared advocate or adversary relations help. For instance, from *AfD is an adversary of refugees* and *Pegida*[2] *is an adversary of refugees* it follows that the AfD is an advocate of Pegida.

3. Assignment of polar roles (cf. E3).

E3: | If a text has framed A0 as a villain then the writer is an adversary of A0 |

The majority of sentences in our corpus that contain *refugee* do not even mention the AfD explicitly. Nevertheless, some of them allow for the

inference of an adversary relation. All those sentences that imply that refugees are villains give rise to such an inference. If the writer conceptualizes someone as a villain, then he is an adversary of him. Clearly, this is defeasible, e.g. *He lied to me, but I still admire him.* This, however, must be done explicitly, otherwise we are entitled to assume an adversary relation. It is a distinctive component of our approach that we are able to determine such polar assignments. Their instantiation, however, must be licensed by the factuality or counterfactuality of events (cf. section 4).

Similarly, if someone conceptualizes someone as a victim, then he is - to a certain degree and maybe only situation-specific - an advocate of him, since normally victims do have our sympathy. Interestingly, the corresponding argumentation while true for benefactor, is not true for beneficiary. Someone who benefits from something could nevertheless be an adversary of us (we could take it as unjust that he benefits).

3 Attitudes and Polar Effects

According to Deng and Wiebe (2015) a verb might have a positive and a negative effect on the filler of the direct object, which they map to the patient or theme role. It is, however, not only the direct object that bears a polar connotation, but also the subject (e.g. *to whitewash*), the indirect object (di-transitive *to recommend*), the PP object (*to fight for*), and the complement clause (*to criticize that*). For German, we (Klenner and Amsler, 2016) have introduced a freely available verb lexicon which we called *sentiframes* (about 300 verb frames). For each verb, it specifies the positive and negative effects that the affirmative and factual use of the verb has on the filler objects on the basis of its grammatical functions. In case that a verb subcategorizes for a complement clause, the verb's implicature signature in the sense of Nairn et al. (2006) is specified as well. We have revised this resource[3] and substantially augmented it by adding 800 additional verbs, their frames (1200) and their verb signatures. We have also started to model verb derived nominalisations (e.g. *destruction*).

A sentiframe or connotation frame (we use the later term, henceforth) of a verb in our new model provides a mapping from argument positions (A0 etc.) to polar roles. We use two polar roles which

---

[1] We, the authors, use our knowledge as informed citizens in order to make such characterizations.

[2] Another xenophobic German movement.

[3] https://pub.cl.uzh.ch/projects/opinion/lrec_data.txt

we call *source* and *target*. Both roles are further qualified (verb specific) according to the polarity they bear. This is summarized in principle P1.

| P1: | *source* and *target* are polar roles of verbs, they bear a positive or negative effect (or none) |

Take for example the verb *whitewash*:

A0 (source) negative effect
A1 (target) positive effect

It is negatively connotated to whitewash (A0), while being whitewashed - at least given a naive point of view - is positive (A1).

A positive effect on the target indicates that the source either acts in a way that the target is positively affected (*cherish*), or it expresses a positive relationship directly (*admire*). Thus, principle P2 holds:

| P2: | The type of attitude a verb expresses depends on the target role effect: *advocate* if positive, *adversary* if negative |

The attitude relation is independent of the effect on the source (if any). We specify a function $\mathcal{R}$ (see Figure 1) that retrieves the attitude relation of a verb v given its affirmative status, i.e. *aff(v)* = 1 if v is affirmative and 0 if negated. The function *teff(v)* retrieves the effect of verb v on its target role.

$$
\mathcal{R}(v) =
\begin{cases}
adversary & \textit{if teff(v)=negative} \wedge \textit{aff}(v) = 1 \\
adversary & \textit{if teff(v)=positive} \wedge \textit{aff}(v) = 0 \\
advocate & \textit{if teff(v)=positive} \wedge \textit{aff}(v) = 1 \\
advocate & \textit{if teff(v)=negative} \wedge \textit{aff}(v) = 0
\end{cases}
$$

Figure 1: Attitude Determination for Verbs

Depending on the affirmative and factuality status of the verb (event), the polar roles (see P1) turn into what we call a *polar assignment*.

| P3: | *benefactor, villain, beneficiary, victim, pos_affected, neg_affected*, ... are polar assignments |

For instance, given *A0₁ regrets that A1 has been insulted*, the negative target role of the verb *insult* gives rise to the polar assignment *victim*. The reason why we distinguish polar roles from polar assignments is that negation might alter the realization of a polar role.

| P4: | The polar assignment (that a polar role gives rise to) depends on the affirmative status of the verb |

For instance in *A1 was not rewarded*, the positive target role of A1 (target of *reward*) is, given verb negation, either neutralized or could even be interpreted as negative (A1 receives a negative effect). Note that an advocate relation between A0 and A1 does not necessarily imply that any of them receive a polar assignment. Given *A0₁ fears that A0₂ has insulted A1*, we have an advocate relation between A0₁ and A1. But neither is A0₂ a villain nor A1 a victim. In the context of *fears*, the truth value (factuality status) of the event denoted by *insult* is unknown.

| P5: | Polar assignments occur if factuality or counterfactuality is given, but are blocked given non-factuality |

In the case of a factual, but non-affirmative use of the verb, the situation is a bit more complicated, since negation might pragmatically be used in various ways, e.g. as a reproach (*We complain that A0 has not helped A1*) or as a plain denial (*We confirm that A0 has not criticized A1*). Only in a reproach the attitudes and the effects can safely be (partially) inverted. Given *A0 has not whitewashed A1* (meant as a denial), we might infer that this is negative for A1. But we certainly would not say that A0 has a negative attitude towards A1 nor that A0 should receive a positive effect (inverting the negative effect of the affirmative use).

The situation changes if we know that a statement is meant as a reproach, e.g. if it is embedded into a verb with a negative effect on its subclause, as in *A0₁ criticizes that A0₂ has not whitewashed A1*. We interpret this as a negative attitude of A0₁ towards A0₂ and a positive attitude of A0₁ towards A1. Here is the (partial) frame for *criticize* (A3 denotes a proposition):

A0 no effect
A3 negative effect

Note that in this example we have to combine two attitudes stemming from different verbs, namely *criticize* and *whitewash*.

| P6: | An attitude towards an event might lead to a polar assignment for some roles of that event |

An adversary relation on *whitewash* stemming from *criticize* combines with the adversary relation of *not whitewash* (=adversary) to

give an advocate relation between $A0_1$ and $A1$. Figure 2 shows the definition of the function $\mathcal{C}$, which realizes relation composition. In the current example, the call would be: $\mathcal{C}(\mathcal{R}(criticize), \mathcal{R}(whitewash))$.

$$\mathcal{C}(r, s) = \begin{cases} adversary \text{ if } r = advocate \wedge s = adversary \\ advocate \text{ if } r = adversary \wedge s = adversary \\ adversary \text{ if } r = adversary \wedge s = advocate \\ advocate \text{ if } r = advocate \wedge s = advocate \end{cases}$$

Figure 2: Attitude Composition

P7:	Attitudes combine with attitudes to form a derived attitude

In P5, we saw that polar assignments depend on (counter-)factuality. We have not yet discussed how to determine (counter-)factuality. In the context of verbs that have clausal complements, we need a further notion, namely that of an implicature signature (Nairn et al., 2006). It relates to the truth or falsehood commitment that a verb casts on its clausal complement. The factuality of an event denoted by a clausal complement can be determined from the implicature commitment of the matrix verb, the affirmative status of the matrix verb and the affirmative status of the clausal complement. We discuss this in the next section. For the moment we postulate P8.

P8:	The polar assignment (that a polar role gives rise to) not only depends on the affirmative status of the verb, but also on the affirmative status and the implicature signature of the matrix verb

4 Truth Commitment, Negation, Factuality Status

We distinguish factual (true), counterfactual (not true) and non-factual (truth value is unknown). We call this the *factuality status* of an event denoted by a verb. In order to determine the factuality status of a clausal complement of a verb, its implicature signature and the affirmative status of the matrix and the subclause verb have to be taken into account. In order to specify the implicature signature, we use T, F, N for truth committed, falsehood committed, and no commitment, respectively, which is along the lines of Nairn et al. (2006), though not totally identical (e.g. they use *polarity* to denote what we call *affirmative status*).

For instance, *to regret* as a factive verb is truth committing (T), both in its affirmative and negated usage. Thus, the clausal complement of an instance of *to regret* is factual if affirmative, and counterfactual if negated. While *to refuse* is falsehood committing (F) if affirmatively used, there is no commitment (N) if negated. The clausal complement of negated *to refuse* thus is, in any case (i.e., affirmative or negated) non-factual.

In order to determine factuality, we first define a function $\mathcal{T}(v)$ which assigns a signature to a verb (especially with clausal complements) given the affirmative status of the verb. Figure 3 gives a (partial) definition of such a verb (class) specific mapping. Here, *aff(v)* = 0 (again) means that the verb v is negated, while *aff(v)* = 1 indicates an affirmative use of the verb.

$$\mathcal{T}(v) = \begin{cases} T & \text{if } v \in \{force, \ldots\} \wedge \textit{aff}(v) = 1 \\ N & \text{if } v \in \{force, \ldots\} \wedge \textit{aff}(v) = 0 \\ F & \text{if } v \in \{forget, \ldots\} \wedge \textit{aff}(v) = 0 \\ T & \text{if } v \in \{forget, \ldots\} \wedge \textit{aff}(v) = 1 \end{cases}$$

Figure 3: Implicature Signature

In order to determine the factuality of an event denoted by a subclause of a matrix verb, we apply the function defined in Figure 4. We use 'fact', 'cfact' and 'unk' for factual, counterfactual and unknown. Note that we interpret factuality as event factuality in the sense of Saurí and Pustejovsky (2009). The function m applied to the verb v delivers the embedding matrix verb.

$$\mathcal{S}(v) = \begin{cases} fact & \mathcal{T}(m(v)) = T \wedge \textit{aff}(v) = 1 \\ cfact & \mathcal{T}(m(v)) = T \wedge \textit{aff}(v) = 0 \\ fact & \mathcal{T}(m(v)) = F \wedge \textit{aff}(v) = 1 \\ cfact & \mathcal{T}(m(v)) = F \wedge \textit{aff}(v) = 0 \\ unk & \mathcal{T}(m(v)) = N \end{cases}$$

Figure 4: Factuality Determination

The main clause is non-factual if modals are present, otherwise it is factual or counterfactual. Negation turns the event denoted by the main verb into counterfactuality. Under counterfactuality as well as under factuality, polar assignments are licensed. Only if a main clause is non-factual, polar assignments are blocked.

Another distinctive feature of our approach is that not only clausal complements receive an implicature signature, but any verb role that could

take a nominalisation as a filler receives one. Moreover, nominalisations themselves have signatures. We are not aware of any model that also considers these cases. Take: *He criticized the destruction of the monuments of Palmyra through Isis*. Here *destruction* is the direct object and there is a truth commitment stemming from *criticize*. The destruction, thus, is factual (affirmative use of *criticize*). Since *destruction* has 'T N' (= 'affirmative negated') as signature and since it is affirmative (T holds) Isis is recognized as a villain. If we take *to fear* instead of *to criticize*, this no longer holds. Also, if we add *supposed* or *postponed* to *destruction* (*supposed destruction*) we have 'N' as commitment and polar assignments are blocked.

The factuality status is determined outside-in. Slightly simplifying, we can say that in order to infer an attitude between actors, the verb (event) of A0 or the verb (event) of A1 must be factual or counterfactual. If both, A0 and A1 are arguments of the same verb, then it needs to be factual (A0 cheats A1). Counterfactuality (e.g. A0 no longer admires A1) might - depending on the verb or other indicators like *no longer* - also license an attitude derivation.

5 Reader Perspective

The *reader perspective* distinguishes opponents from proponents. These classes need to be specified in advance by the reader. For instance, he could select particular political parties or politicians as proponents. In our experiments described below, we created a virtual reader along the following lines. Our reader is a proponent of Europe, Merkel, Germany, refugees and so on and against the AfD. His values and aversions, hopes and fears are those of a typical member of the Western society, which we fixed in a reader profile that assigns polarities along the lines of the Appraisal theory (Martin and White, 2005), i.e. judgment, affect and appreciation. For instance, *honesty* is judgment positive and represents a moral value of the reader, while *terrorist* is a contemner of the reader's values. This lexicon[4] serves two purposes. It forms the basis of the reader's ability to understand what a text implies. But it also represents (or better approximates) his moral values, his aesthetic preferences, his emotional dispositions. He is able to discern that in *A0 ap-*

---

[4]The lexicon is an adaptation of the lexicon described in Clematide and Klenner (2010).

proves terrorism someone is an advocate of something he finds immoral or inhuman. Altogether, we have six roles for the reader perspective: *myValues, myAversions, mvValueConfirmer, myValueContemner, myProponent, myOpponent*.

Note that the instantiation of these roles (except myOpponent and myProponent) sometimes raise the need for sentiment composition (not only lexicon access). *terrorist* is a myValueContemner since the word - according to the lexicon - denotes a judgment negative animate entity. In order to classify *cheating colleague* as a myValueContemner of the reader, composition is needed. The judgment negative adjective *cheating* combined with the neutral noun *colleague*, which denotes an actor, gives rise to a judgment negative phrase denoting a myValueContemner of the reader. The phrase *sick minister*, on the other hand, although *minister* is an actor and *sick* is a negative word (but appreciation negative, not judgment), does not denote a myValueContemner, but a neutral entity.

6 Writer Perspective

The *writer perspective* tells the reader what the writer wants him to believe (to be true) and explicates what this implies for the status of the targets involved, i.e. whether they are benefactors etc. It is the way the writer conceptualizes the world through his text.

The roles stemming from the polar assignment, e.g. *victim, villain, benefactor and beneficiary* are actor roles related to the moral dimension (verb-specific), while the additional roles *pos_actor, neg_actor, pos_affected, neg_affected* are used for the remaining cases (roles of not morally loaded verbs).

Given a sentence, we combine the attitudes, the reader and the writer perspective into a single view. Formally, we instantiate the relation 5-tuple $\langle L_r, L_w, rel, L_r, L_w \rangle$ where L_R, L_w is the reader and writer perspective, respectively and rel represents the attitude of the source towards the target. The reader and writer view L_r and L_w are applied twice, to the source (left hand part of the 5-tuple) and the target (right hand part of the 5-tuple) connected by the (directed) attitude relation rel.

The writer perspective, L_w is determined by calling the function $\mathcal{A}(a, v)$ (see. Figure 5) with the verb v and the polar role a of the entity in question (target or source).

Given that *terrorist* is a value con-

$$\mathcal{A}(a,v) =$$

$$
\begin{cases}
benefactor & if & v \in \{help..\} \wedge a = source \\
& & \wedge\, \mathcal{S}(v) = fact \\
beneficiary & if & v \in \{help..\} \wedge a = target \\
& & \wedge\, \mathcal{S}(v) = fact \\
villain & if & v \in \{cheat..\} \wedge a = source \\
& & \wedge\, \mathcal{S}(v) = fact \\
victim & if & v \in \{cheat..\} \wedge a = target \\
& & \wedge\, \mathcal{S}(v) = fact \\
victim & if & v \in \{help..\} \wedge a = target \\
& & \wedge\, \mathcal{S}(v) = cfact \\
\dots
\end{cases}
$$

Figure 5: Polar Assignment

temner of the reader then *The politician helps the terrorist* would lead to the tuple $\langle$some,benefactor,advocate,myValueContemner,beneficiary$\rangle$ which reads: some benefactor is an advocate of a value contemner as a beneficiary. This immediately reveals the *charge* of the statement: a value contemner as a beneficiary. We could think of even more charged cases, e.g. a proponent of us as a villain. Or a proponent of us as an advocate of an opponent. Our tuple notation makes this transparent, it enables the search for such cases and we have defined secondary relations on top of it. We have identified 16 pattern that instantiate 4 new relations: new proponent, new opponent, no longer proponent and no longer opponent. A proponent of the reader who is an advocate of an opponent might no longer be a proponent etc. We give a couple examples in section 9.

7 Attitude Prediction

Given a sentence, we consider all pairs $\langle x,y \rangle$ such that x and y denote a noun position that acts as a polar role of one or more verbs. Given a pair of actors or entities, $\langle x,y \rangle$, both might occupy polar roles of the same or of different verbs. If x and y are arguments of the same verb, then, if factuality (or counterfactuality) holds, the attitude between them comes from the underlying verb. That is $\mathcal{R}(v)$ is applied if $\mathcal{S}(v) = $ 'fact' (or 'cfact').

If x and y have different verbal heads, i.e. the verbal head of x either directly or recursively embeds a verb with y as a polar role, then the relations stemming from the intermediary v_i are composed into a single relation rel (see Figure 2). If $A0_1$ *approves that $A0_2$ criticizes $A1$*, the relation between $A0_1$ and $A1$ is that of an adversary.

This depends on the advocate relation of *approve* and the adversary relation of *criticize*. Technically, we call $\mathcal{C}(\mathcal{R}(v), \mathcal{R}(v))$, depending on $\mathcal{S}(v)$. In general, we use a recursive function where attitude composition is performed outside-in along the lines just discussed. For instance, given *$A0_1$ criticizes that $A0_2$ does not help $A1_1$ to free $A1_2$*, we get: $\mathcal{C}(\mathcal{C}(\mathcal{R}(criticize), \mathcal{R}(help)), \mathcal{R}(free))$ which gives *adversary* (*criticize*) combined with *adversary* (negated *help*) which gives *advocate* which in turn is combined with *advocate* (*free*) which gives *advocate*: $A0_1$ advocate $A1_2$.

8 Sentiframes: Additional Details

The main components of a sentiframe or connotation frame are the effects that the source and target roles receive. We have shown that the actual assignments and relations also depend on the affirmative and factuality status. Of course, ambiguity is a problem. We found that shallow selectional restrictions distinguishing roles that require their filler to be an actor (persons, organizations, nations etc.) from roles where the filler must not be an actor actually help to reduce verb ambiguity. Other restrictions that are very useful are constraints that check the polarity of a filler object bottom-up. There are a couple of verbs that only (should) trigger if a bottom restriction is met. Take *prevent*: the one who prevents the solution of an urgent problem is a negative actor while if he prevents an assault, he is a hero. Other examples are verbs like *to call, to take*, e.g. *to call it a good/bad idea that* produces a positive or negative effect on the clausal complement. 95 connotation frames do have such bottom-up restrictions. We have implemented a straight-forward phrase-level sentiment composition (Moilanen and Pulman, 2007) in order to check bottom-up restrictions.

9 Example

Take the sentence (relevant positions are indexed): *The left-wing politician$_3$ criticized$_4$ that Merkel$_6$ helps$_7$ the refugees$_9$.* We get three pairs: $v_4{:}\langle x_3, y_6 \rangle$, $v_4{:}\langle x_3, y_9 \rangle$ and $v_7{:}\langle x_6, y_9 \rangle$. Let's say the reader has no prior attitudes towards left-wing politicians but that refugees has his sympathy (are *myProponents* of his). We discuss the case of $v_4{:}\langle x_3, y_9 \rangle$, i.e. the directed relation of the *left-wing politician* towards the *refugees*. The source of *criticize* has a negative attitude towards the *help* event. Since affirmative *help* represents an advo-

#	Relation Tuple	Illustration
1	⟨myProp,entity,adversary,myProp,neg_affected⟩	US refuses Germany something
2	⟨some,entity,is,adversary,of,myAversions,neg_affected⟩	someone condemns terror
3	⟨some,entity,is,advocate,of,myAversions,pos_affected⟩	someone insists on vengeance
4	⟨myProp,benefactor,advocate,myValContemner,beneficiary⟩	US supports dictator
5	⟨some,villain,adversary,myValues,neg_affected ⟩	someone ridicules human behavior

Table 1: Charged Relation Tuples

cate relation, we get adversary ∧ advocate = adversary (see Figure 2). The *left-wing politician* is just an entity, but the reader is a proponent of *refugees*. Since *help* is factual, *refugees* are beneficiaries. This yields:

⟨ some,entity,adversary,myProponent,beneficiary⟩.

We could paraphrase this as *some entity is an adversary of my proponent being a beneficiary*. Note that beneficiary as a role comes from factual *help*. This is the writer or text perspective. It tells us that the *refugees*, the reader proponents, are beneficiaries of some event that happened in reality. The relation also tells us that some entity is an adversary of this. That is, he does not approve the status of the reader's proponents, the refugees, as beneficiaries. This immediately makes him a candidate for the list of actors that are opponents of the reader.

Our tuple notation directly confronts the writer and the reader perspective and thus allows one to search for interesting cases. We have used a corpus comprising 3.5 million sentences taken from German periodicals (ZEIT and Spiegel) to explore this idea. Examples are given in table 1. The third column illustrates the underlying cases; US and Germany are set to be proponents of the reader (for short: *myProp*). In 1, two proponents are (surprisingly) adversaries. In 2, someone disapproves what the reader disapproves (a new proponent?). In 3, someone approves what the reader disapproves (a new opponent?). In 4, a proponent acts in a way the reader finds morally questionable (no longer a proponent?), and in 5, someone might turn out to be an opponent, since he violates the reader's values.

10 Empirical Evaluation

We have evaluated our approach quantitatively on the basis of 160 sentences. The data consists of 80 (rather complex) made-up sentences (one or more subclause embeddings) and 80 real sentences. Our goal was to verify the generative ca-

pacity of our model, thus the made-up sentences. It is much more convenient to invent complex sentences, where e.g. negation is permuted exhaustively over all subclauses, than to try to sample such rare constellations. Two annotators specified advocate and adversary relations and harmonized their annotations in order to get a gold standard.

Our goal was to see how our lexicon, including the principles of factuality determination, determines the performance. The precision was 83.5%, recall was 75.2%, which gives an F measure of 79.1%[5]. We then dropped the verb signatures from the lexicon, that is, we replaced the individual signatures by a default setting. There are three possible settings. We set the signature for the affirmative use of the verbs to 'T' (truth commitment), the signature for negated cases was set to 'F', 'N' and 'T' in turn. We got a precision of 69.06%, 75.36% and 74.88% and a recall of 69.36%, 71.62% and 75.2%. The F measure for the best default setting ('T T') is 75.06% which is about 4% points worse than the system's result, 79.1%. We also see that precision droped by 8% points which is a substantial loss. This demonstrates that verb-specific information is crucial.

Encouraged by these results, we decided to carry out a qualitative study in stance detection. We took 360 000 sentences from 100 000 Facebook posts of AfD members. Our system produced 44 000 polar facts from them: attitudes and polar assignments. Since these posts are (mostly) from AfD members, they implicitly represent their stance. The key messages, the self-conception of the party and the proclaimed friends and enemies should be accessible through these posts.

We aggregated polar facts by counting how often an actor was conceptualized as a villain etc., but also by counting the number of advocate and adversary relations between actors. We evaluated these aggregated polar facts through introspection.

---

[5]We use a dependency parser (Sennrich et al., 2009) and a rule-based predicate argument extractor, see Klenner and Amsler (2016) for the details

That is we relied on our knowledge about the AfD, its goals, methods, ideological stance etc. as portrayed by the mainstream German media.

The most important (since most frequent) polar fact derived by our systems already was in heart of the AfD's stance, namely that Angela Merkel, the German chancellor, is an adversary of Germany. That is exactly what the AfD claims. Actually, we get a very strong statement, in our tuple notation (recall that Merkel and Germany are reader proponents):

⟨myProponent,villain,adversary,myProponent,victim⟩

That is: myProponent (Merkel) as villain is an adversary of myProponent (Germany) as a victim. Conversely, these texts imply that the AfD is an advocate of Germany, that the refugees are adversaries of Germany, while the German government is an advocate of the refugees. Curiously enough, for the relation of the AfD towards refugees, we got inconsistent evidence (three times adversary, three times advocate). However, if we look at the polar assignment of refugees, which is villain, the picture is clear (see below). There are a couple of polar facts related to an event on New Year's Eve in 2015, where groups of men including migrants sexually assaulted women (that is the official statement). Our system came up with the polar fact that refugees are adversaries of (these) women.

Another question is, of course, who is to blame for the situation (in Germany). The mere fact that, in the perception of the AfD, Merkel is an adversary of Germany does not tell us whether this is positive or negative (in the eyes of the AfD). An adversary relation might be positive (e.g. A0 adversary terrorism), or negative (e.g. A0 adversary truth), i.e. the holder of the adversary attitude might be someone who shares or contemns our values, depending on the event underlying the adversary relation. If Merkel is said to cheat Germany, then the writer wants the reader to believe that Merkel is a villain and Germany her victim. Only then we know that the writer is (must be) an adversary of Merkel and an advocate of Germany.

In order to see who are villains and victims according to the AfD posts, we determined the most frequent actors that are classified as villains etc. To give a couple of examples: Among the villains are the refugees (ranked highest), immediately followed by Merkel and men (representing male refugees), the German word for villain itself (*Täter*) and government. We believe that these are

perfect hits. Victims are Germany, women (New Year's eve event), the AfD (presumably since misunderstood), and the citizen of Germany (AfD seems to believe: the government cheats the citizen). Among the beneficiaries are men (male refugees who are free to molest women without consequences), but also refugees (there is a welcome culture), Europe, criminals (since the government is weak) and the government. From these lists we can also see that the AfD conceptualizes itself as a victim, a positive actor and even a benefactor.

If someone is an adversary of the values of the reader, he might be a new opponent: we defined this and similar relations (no-longer-opponent) on top of our tuple notation. We found 80 different new opponent candidates, including various politicians, countries (their governments), parties, institutions (e.g. Nato) and concepts like *Flüchtlingswelle* (flood of refugees) or *politische Elite* (political elite). The list of entities we should no longer consider a proponent of the reader is perfect, it comprises Asylbewerber (refugee), Bundesregierung (government), Bundestag (parliament), EU, and Merkel. This exactly reflects the stance of the AfD.

11 Related Work

In this paper, stance detection is accomplished on the basis of opinion inference. A basic form of opinion inference is event evaluativity in the sense of Reschke and Anand (2011). They determine the polarity of an event as a function of the polarity of the arguments of the verb denoting the event. Work in the spirit of Reschke and Anand (2011) for the German language is described in Ruppenhofer and Brandes (2016a) and Ruppenhofer and Brandes (2016b). The goal of their approach is to create a verb-specific mapping from the prior attitude a so-called external viewer of an event has towards the verb arguments onto his overall evaluation of the event. For instance, if an immoral person lacks a good job, this is positive in the eye of the external viewer. Their approach focuses on a lexical resource, not on a system carrying out opinion inference. Thus, the authors do not take truth commitment, negation, and factuality determination into account. Nevertheless, their findings might be useful for what we call the reader perspective (where the prior polarity are needed).

A rule-based approach to sentiment implica-

tures (their term) is described in Deng and Wiebe (2015). This is the most recent and most elaborated version of a number of models of these authors. The goal is to detect entities that are in a positive or negative relation to each other. PosPair and NegPair are used as relation names, respectively. The model of Deng and Wiebe (2015) also copes with event-level sentiment inference, however factuality is not taken into account at all. Also, the reader is not modeled explicitly. Moreover, only attitude relations are derived, no polar assignments (beneficiary etc.) are modeled.

Recently, Rashkin et al. (2016) have presented an elaborate model that is meant to explicate the relations between all involved entities: the reader, the writer and the entities referred to by a sentence. Also, the internal states of the referents and their values are part of the model. The underlying resources, called connotation frames, were created in a crowd sourcing experiment, and the model parameter (e.g. values for positive and negative scores) are average values. Our resource, in contrast to such a layman's guess, was specified by an expert. The authors use belief propagation to induce the connotation frames of unseen verbs; they also use the connotation frames to predict entity polarities. This was applied to analyze the preferences and dispreferences of Democrats and Republicans. Choi et al. (2016) presented another application of that resource. Rashkin et al. (2016) claim to have a reader and a writer model, however, they do not seem to use it. This is in sharp contrast to our approach. Like Deng and Wiebe (2015), Rashkin et al. (2016) do not incorporate polar assignments (and factuality) in their model, which we deem crucial for stance detection.

Our previous model (Klenner, 2016; Klenner and Clematide, 2016) was realized with Description Logic OWL and the rule language SWRL (Horrocks and Patel-Schneider, 2004). The goal was to extract pro and contra relations from text. 42 SWRL rules were needed in order to establish such a functionality. In this paper, we have introduced a new model based on functions carrying out (a lean) attitude composition. We have also revised our approach for factuality determination. We now have a tripartite distinction while previously, our factuality labels were binary. The most important new feature of our current approach is the specification of our relation tuple which integrates the writer and the reader view.

A crucial difference of our model to existing approaches from the field of stance detection is that we do not only strive to classify the stance of a writer towards known controversial topics (e.g. abortion, climate change) like in e.g. Somasundaran and Wiebe (2010), Hasan and Ng (2014) or Anand et al. (2011). We also seek to identify the targets of the writer's stance in the first place. Among others, it is the way the writer frames the entities in his discourse (as villains etc.) that indicates his likes and dislikes.

12 Conclusions

We claim that the writer's conceptualization of reality as a narrative reveals his stance. In our case, the members of a political party together write that narrative which reflects how the AfD, a German right-wing party, divides the world into proponents and opponents, benefactors and villains and so. In contrast to previous approaches, we stress the point that an attitude between an opinion source and an opinion holder alone does not necessarily tell anything about how the writer perceives it. Only if we know the roles the source and the target play (e.g. villain, victim) in the whole discourse, we can identify the writer's stance towards them.

On a more technical level, the contributions of our approach are: 1200 new connotation frames for German, and a framework that integrates inferences both in verbal and nominal contexts. Our relation tuples jointly encode the reader and the writer perspective as well as the attitude among the source and target expressed by the underlying verb. Such a relation directly shows what the writer wants the reader to believe and how the reader - given his personal stances - might perceive this. This enables the reader to search for interesting constellations, where e.g. a proponent of his acts in an unexpected way.

Obvious future work stems from the need to define a more elaborated evaluation scenario. A small quantitative and an introspective qualitative evaluation was just a first (though successful) step.

References

Pranav Anand, Marilyn Walker, Rob Abbott, Jean E. Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, pages 1–9, Portland, Oregon, USA.

Eunsol Choi, Hannah Rashkin, Luke Zettlemoyer, and Yejin Choi. 2016. Document-level sentiment inference with social, faction, and discourse context. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 333–343, Berlin, Germany, August.

Simon Clematide and Manfred Klenner. 2010. Evaluation and extension of a polarity lexicon for German. In *Proceedings of the First Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, pages 7–13, Lisbon, Portugal.

Lingjia Deng and Janyce Wiebe. 2015. Joint prediction for entity/event-level sentiment analysis using probabilistic soft logic models. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 179–189, Lisbon, Portugal.

Kazi Saidul Hasan and Vincent Ng. 2014. Why are you taking this stance? Identifying and classifying reasons in ideological debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 751–762, Doha, Qatar.

Ian Horrocks and Peter F. Patel-Schneider. 2004. A proposal for an OWL rules language. In *Proceedings of the Thirteenth International World Wide Web Conference (WWW)*, pages 723–731, New York, NY, USA.

Manfred Klenner and Michael Amsler. 2016. Sentiframes: A resource for verb-centered German sentiment inference. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC)*, pages 2888–2891, Portoro, Slovenia.

Manfred Klenner and Simon Clematide. 2016. How factuality determines sentiment inferences. In Ivan Titov Claire Gardent, Raffaella Bernardi, editor, *Proceedings of *SEM 2016: The Fith Joint Conference on Lexical and Computational Semantics*, pages 75–84, Berlin, Germany, August.

Manfred Klenner. 2016. A model for multi-perspective opinion inferences. In Carlo Strapparava Larry Birnbaum, Octavian Popescu, editor, *Proceedings of IJCAI Workshop Natural Language Meets Journalism*, pages 6–11, New York, USA.

James R. Martin and Peter R. R. White. 2005. *Appraisal in English*. Palgrave Macmillan, London, England.

Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiao-Dan Zhu, and Colin Cherry. 2016. SemEval task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval)*, pages 31–41, San Diego, CA, USA.

Karo Moilanen and Stephen Pulman. 2007. Sentiment composition. In *Recent Advances in Natural Language Processing (RANLP)*, pages 378–382, Borovets, Bulgaria.

Rowan Nairn, Cleo Condoravdi, and Lauri Karttunen. 2006. Computing relative polarity for textual inference. In *Proceedings of Inference in Computational Semantics (ICoS 5)*, pages 67–75, Buxton, England.

Hannah Rashkin, Sameer Singh, and Yejin Choi. 2016. Connotation frames: A data-driven investigation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 311–321, Berlin, Germany, Angust.

Kevin Reschke and Pranav Anand. 2011. Extracting contextual evaluativity. In *Proceedings of the Ninth International Conference on Computational Semantics (IWCS)*, pages 370–374, Oxford, England.

Josef Ruppenhofer and Jasper Brandes. 2016a. Effect functors for opinion inference. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC)*, pages 2879–2887, Portoro, Slovenia, May.

Josef Ruppenhofer and Jasper Brandes. 2016b. Verifying the robustness of opinion inference. In Stefanie Dipper, Friedrich Neubarth, and Heike Zinsmeister, editors, *Proceedings of the 13th Conference on Natural Language Processing (KONVENS)*, pages 226–235. Bochum, Germany.

Roser Saurí and James Pustejovsky. 2009. FactBank: a corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3):227–268.

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.

Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124, Los Angeles, California, USA.