Extending effect annotation with lexical decomposition

Josef Ruppenhofer and Jasper Brandes

Department of Information Science and Natural Language Processing
Hildesheim University
D-31141 Hildesheim
{ruppenho|brandesj}@uni-hildesheim.de

Abstract

In this contribution, we report on an effort to annotate German data with information relevant to opinion inference. Such information has previously been referred to as *effect* or couched in terms of event-evaluation functors. We extend the theory and present an extensive scheme that combines both approaches and thus extends the set of inference-relevant predicates. Using these guidelines to annotate 726 German synsets, we achieve good inter-annotator agreement.

1 Introduction

In recent years, there has been increasing interest in inferring implicit opinions in addition to capturing explicit expressions of opinion. A series of papers by Reschke and Anand as well as Wiebe and her collaborators (Anand and Reschke, 2010; Reschke and Anand, 2011; Deng et al., 2013; Wiebe and Deng, 2014) has shown the great potential of opinion inference: speakers and authors leave many implicit opinions for hearers to infer. While these additional inferred opinions involve sentences or clauses that bear no explicit sentiment at all, they very often interact with sentences or clauses that do bear explicit sentiment, as in example (1).

(1)   She is disappointed that Peter is **happy** because the Colts LOST.

In (1), we can infer, for instance, that Peter, the (nested) source of the explicit sentiment (in boldface) towards the event as a whole, is also negative towards the Colts football team given that the event affected them negatively (in small caps). As laid out in great detail by Wiebe and Deng (2014), given partial knowledge about explicit subjective expressions, sources' attitudes about participants and knowledge about the effects of events on their

participants, people can generate contextually defeasible inferences about the missing pieces. And inferences can build on prior inferences: in the above example, we can further defeasibly infer that 'she' potentially holds the opposite opinion towards the Colts from Peter, given that she is disappointed at his attitude.

Although opinion inference is a pragmatic process, it relies on rich lexical knowledge of subjective expressions and of predicates, which entail some kind of effect.[1] While a great deal of effort has been devoted to the acquisition of explicit sentiment expressions, the acquisition of information that is relevant for opinion inference is in its infancy, by comparison.

In this paper, we report on an effort to manually annotate effect-relevant predicates in GermaNet (Hamp and Feldweg, 1997), a WordNet-like (Fellbaum, 1998) electronic lexical resource for German. The purpose of annotating the word senses of lemmas that have at least some effect-entailing senses is to construct a gold standard for evaluating automatic systems that provide a complete automatic annotation of the senses in the resource via label propagation along the lines of Choi and Wiebe (2014). Here we focus on the following contributions of our work:

- We extend the range of predicates covered, relative to Choi and Wiebe (2014).

- We provide a typology of effect-relevant predicates in terms of lexical decomposition.

- We explicitly take account of the syntax-semantics interface and mark the argument that is affected.

- We report inter-annotator agreement results for labeling lexical entries with argument-

---

[1] Work on connotation detection by Feng et al. (2011) can be seen as addressing the issue of determining sources' attitudes in those cases when they are aligned with their stereotypical attitudes within particular (discourse) communities or cultures.

specific effect information, whereas so far only agreement on annotations of corpus instances has been studied.

This paper proceeds as follows. In §2, we present related work. In §3, we describe the way we chose the lexical entries we annotated. §4 lays out our own annotation scheme and its differences to prior work. We report on the inter-annotator agreement that we were able to achieve applying our scheme in §5. We offer our conclusions in §6.

2 Related work

The relevance of predicates to the inference of attitudes towards the events they are used to refer to was explored by Reschke and Anand (2011) who treat predicates and their arguments as functors that map tuples of argument and verb properties to evaluations. An example is given in Table 1. The first line of the table applies to the situation where there is a possessor (x) who is valued positively by some nested source and a possession (y) that is also valued positively. If the relation between them is *have* (third column), that relation is valued positively from the external point-of-view. If the relation is *lack*, that relation is valued negatively. The table shows that the reasoning for *lack* also applies to events of withholding and depriving which result in lack.[2]

x	y	have	lack	withhold	deprive
+	+	+	-	-	-
+	-	-	+	+	undef.
-	+	-	+	+	+
-	-	+	-	-	undef

Table 1: Functors for verbs embedding a state of possession

It is important to keep in mind that the goal of the inference procedure is to assess the attitude of an external observer on the event. Thus, while an external viewer may feel positively towards a situation where a person he or she dislikes, *x*, lacks something desirable, *y*, the relevant possessor, *x*, will most likely feel negatively about this lack. Reschke and Anand's functor are intended only to capture the former attitude, of the observer, but not the latter, of the participant.[3]

Reschke and Anand (2011) focus on classes of verbs assuming that verb classes such as verbs of creation behave consistently due to lexical entailments about their arguments. They focus on three prominent kinds of entailments: ones related to possession, existence and affectedness. Reschke and Anand (2011) test the correctness of the predictions generated by their theory by annotating actual corpus instances. They do not evaluate the agreement on the presence of the lexical entailments for predicates themselves. Instead, they simply identified the verbs having particular lexical entailments by inspecting FrameNet's frames (Baker et al., 1998; Ruppenhofer et al., 2010) and its hierarchy.

While acquisition or validation of a large lexicon was not the aim of Reschke and Anand (2011), it is the focus of later work by Choi and Wiebe (2014) who seek to generate EffectWordNet, a large lexical resource based on WordNet, in which information relevant for opinion inferences is recorded. The notion to which the annotation of the WordNet word senses appeals is that of effect rather than the more specific entailments used by Reschke and Anand (2011). The idea is that in order to determine a nested source's view on an event, one first needs to look at the positive or negative effect that the event has on its object.[4] In combination with the attitude towards the object this yields the evaluation of the state that results from the event. That evaluation can then be transferred onto the agent or cause responsible for bringing about the effect, and onto the overall action brought about by the agent or cause. Consider this example taken from Choi and Wiebe (2014):

(2)   The bill would curb skyrocketing health care costs.

The reasoning applied to this example is that since *skyrocketing* conveys negative sentiment towards *curb*'s object *health care costs*, which is negativey affected by the event of curbing, we understand that the writer feels positively towards the curbing event and the *bill* that brings it about.

The idea of looking specifically at the positive or negative effect on a particular affected entity (=object) features only in Choi and Wiebe's

(2014) theory. In the work of Reschke and Anand (2011), it is left implicit. Consider again Table 1: the plus and minus signs encode attitudes of an external viewer towards participants and events but they do not capture the nature of how (some) participants are affected. Further, as indicated by the variable names for the participants (x,y), no affected entity is identified.

In the work of Choi and Wiebe (2014), all WordNet word senses of 1143 lemmas were manually annotated as +/-effect. This data was then used as seed and test data for a label propagation algorithm that spread effect information throughout WordNet. In a prior study by Deng et al. (2013), textual instances of effect-predicates were annotated as triples consisting of <agent, +/-effect event,object>.[5] In addition, the writer's attitude toward the agents and objects of those events was captured.

3   Data

As noted in the introduction, our ultimate goal is to create a WordNet-like resource for German with information relevant to opinion inference that is similar to the English EffectWordNet. Following the approach of Choi and Wiebe (2014), we want to annotate word senses in GermaNet V9.0 (Hamp and Feldweg, 1997), the sister resource to Word-Net for German, that can serve both as training / seed data for automatic methods for graph-based label propagation, and also as a gold standard for evaluating automatic methods.

In picking German synsets to annotate, we made use of the work done by the EffectWordNet group. We extracted all 745 synsets from the EffectWordNet gold standard that were annotated as either +effect or -effect. We omitted all synsets annotated as Null in the source synsets. We then retrieved 273 corresponding German synsets from BabelNet (Navigli and Ponzetto, 2010) on the basis of the WordNet synset IDs. Using the German lemma information and the POS information in BabelNet, we next extracted 998 unique synsets from GermaNet that contained any word senses for the lemmas found in the 273 BabelNet synsets. After expanding the set again based on lemmas found in GermaNet but not in BabelNet, we obtained 1492 GermaNet synsets.

As we will show in §4.2, our annotation scheme

does not only focus on the *effect* on an entity caused by an agent but also allows for the annotation of uncaused states an entity may find itself in. We therefore must include inchoative and resultative verbs as well as resultative adjectives in our data set. Thus, we manually culled inchoative verbs such as *aufblühen* 'blossom', resultative verbs like *verblühen* 'wither' and adjectives (e.g. *hoch*, 'high') from various German lexicons. We extracted all corresponding GermaNet synsets by their lemma and POS information, resulting in 163 verbal and 52 adjectival synsets. After removing duplicate synsets that were found a second time as part of the search for uncaused states, our final data set consists of 1667 GermaNet synsets.

Figure 1 displays an instance of a GermaNet synset, along with its annotations. The key parts for our purposes are

a the initial pair of square brackets containing our effect annotations (bold-face);

b the orth forms set, which lists the lemmas of the synset (underlined);

c the paraphrases, which help us understand the intended senses of the lemmas (italics);

d and the example sentences (lines beginning with "# GermaNet").

Unlike in (Effect)WordNet, each example sentence is accompanied by a syntactic subcategorization frame which lists the set of arguments and adjuncts occurring with the lemma being exemplified. Thus, in the first sentence, *gefallen* is realized with a noun phrase in nominative case (NN) and another in dative case (DN). We refer to these morphosyntactic phrase labels in our effect annotations, as illustrated by the arrows in Figure 1. For instance, the initial block in the example says that for the verb *mögen* 'like' the participant coded as an accusative noun phrase (AN) is positively affected, whereas for the verb *gefallen* 'please' the participant coded as a nominative case noun phrase (NN) is positively affected.

As suggested by the synset in Figure 1, taking valence information into account is important: without it, we are unable to reliably identify which argument is affected in which way. Ideally, we could make use of either a German counterpart to PropBank or FrameNet. However, there is no German PropBank and the Salsa lexicon (Burchardt et al., 2006), which uses the FrameNet formalism, has too low a coverage. It thus makes most sense

---

[5]We follow Choi and Wiebe (2014) and apply the more recent term +/-*effect* instead of the earlier *goodFor/badFor*.

```
35:[+Effect:AN:mögen;+Effect:NN:gefallen] id: 52110, orth forms: [gefallen, mögen],
paraphrases: jemand findet etwas gut; jemandem angenehm sein; jemanden
oder etwas gern haben, lieben (etwas schwächere Form)
verben        Allgemein        [annotation-uncertain|meaning-uncertain]
# GermaNet:  Ihm gefällt Klassische Musik. NN.DN
# GermaNet:  Er mag diese Musik. NN.AN
```

Figure 1: Annotation of a synset

to use GermaNet which is large, structured on the sense-level and has example sentences with associated subcategorization patterns that allow us to access at least syntactic, if not semantic valence.[6] We supply our annotation labels with the understanding that they cover only the syntactic frames exemplified by GermaNet. While these may not include all possible frames, they seem to cover the major syntactic frames for the predicates.

4 Scheme

Our overall approach is a mixture of the functor approach to event evaluation of Reschke and Anand (2011) and of the object-focused approach of Choi and Wiebe (2014). Like Choi and Wiebe (2014) we annotate synsets in a WordNet-like resource but inspired by Reschke and Anand (2011) we annotate a wider variety of predicates and also cover cases where the focus on an affected object alone is not enough to allow a judgment about how the event as a whole should be evaluated.

For practical reasons, our annotations are done with reference to syntactic subcategorization frames that come with example sentences for the GermaNet synsets. Conceptually, we find it useful to reason about the annotation task in terms of lexical decompositions and semantic roles along the lines of e.g. Rappaport Hovav and Levin (1998) and Tenny (1994), *inter alia*.

We present our basic approach in §4.1 and discuss extensions to it and further considerations in the following subsections §4.2-§4.4.

4.1 Underlying linguistic model of effects

Let us consider the prototypical case of predicates relevant for opinion inference, namely ones that involve a causal event (*Cause*) that brings about a resulting event, the (*Effect*). A clear example of such a predicate is *produce* in (3).

(3)   [The explosion Cause] PRODUCED [a loud "bang" Effect].

Example (3) would be a simple +effect-verb in terms of Choi and Wiebe (2014). So it is for us: we mark its German counterpart *produzieren* as *+Effect:AN* to capture the positive effect on its accusative object.

However, in many cases, the effect does not appear directly as a single argument of the predicate but is expressed in two semantically interconnected phrases, one expressing an affected participant and another expressing the relevant situation that affects that participant. The *Patient* is typically realized as an object for verbal predicates. The situation-referring constituent, which we call *Eventuality*, can be of various forms: it can be a verb phrase (4), an adjectival phrase (5), a prepositional phrase, or some other type of unit that can function as a predicate. Semantically, the Patient is the 'subject' of the Eventuality predicate: e.g. (5) could be paraphrased as 'My mortal fear and faintness brings it about that I am deadly pale.'

(4)   [The explosion Cause] MADE [me Patient] [fear for my life Eventuality].

(5)   [My mortal fear and faintness Cause] must have MADE [me Patient] [deadly pale Eventuality].

In our annotation scheme, rather than leave these cases out of consideration, we explicitly record that, in order to assess the effect on the Patient in object position, we need to consider the Eventuality expressed in the secondary predicate. We mark this dependence between the two phrases with the symbol ∼. For instance, for the German equivalent of *make* in (5), *machen*, we would mark *Effect:AN∼BM* not specifying + or -, where AN represents the accusative object and BM (adverbial of manner) covers the adjectival secondary predicate. This information can be used to, for instance, compose the polarity of the sentence in (4) as follows. First, we compute if 'me/I fear for my

---

life' is a positive or negative event, for which we can use information in our resource about the verb *fear*. Second, we input the result of that first computation to the basic cause-event reasoning used for the simple case in (3).

While verbs like *cause* and *make* are very generic and require the nature of the Eventuality to be specified explicitly, other verbs can be thought of as incorporating the Eventuality. For instance, we can think of the event structure of *carve* as "Agent causes Patient to exist by carving it" (cf. example (6)). In the syntax, the Eventuality "to exist" is not expressed because it is already contained in the meaning of *carve*.

(6)   As well as painting, [he `Agent`] CARVED [images `Patient`] from sandalwood.

For the verb *carve*, we record that its active-form object is its Patient, which we take to be positively affected by coming into existence, in line with Reschke and Anand (2011)'s Existence entailment or the guidance of Deng & Wiebe's annotation manual that "To Exist is Good". The annotation for German *schnitzen* would be *+Effect:AN*. In a parallel way, verbs of destruction can be thought of as including a negation of the existence-eventuality: *destroy* is "Agent causes Patient to no longer exist". Accordingly, we would mark the Patients of *destroy* and *kill* as negatively affected; the German counterparts *zerstören* and *töten* would receive the annotation *-Effect:AN* for their accusative objects.

If we allow that the Eventuality can be implicit in the verb's meaning, we can also analyze verbs related to transfer in a similar way. The meaning of *give* is "Agent causes Recipient to have Theme" but in the syntax there is no separate expression of the notion of having: it is incorporated by *give*. As in the case of *make* in (4–5) the positive or negative quality of the Effect cannot be assessed based on either the Recipient or the Theme alone.

(7)   [Bill `Agent`] GAVE [my mom `Recipient`] [a valuable painting `Theme`].

It is clear that we first need to figure out the Effect's polarity on the basis of "Recipient have Theme", for which we reason along the lines of Reschke and Anand's functor in Table 1. From there, we can proceed to the general level of "Cause causes Effect". Accordingly, the annotation for German *geben* would be *Effect:DN∼AN*, which captures the dependence between the dative Recipient phrase (DN) and the accusative Theme phrase (AN). Note that the order matters: the (animate) Recipient's state changes more saliently by coming into possession of the (inanimate) Theme than the other way around. For (7), we reason with Reschke and Anand (2011) that it's good if a person we like has something good. Assuming we like our mothers, the possession of the valuable painting is good. Since agents and causes get credit and blame for the good and bad things they bring about, Bill is evaluated positively.

The idea of decomposing verb meanings can be applied to yet more verb groups. For instance, verbs that refer to an item attaining a lower or higher position on a scale can be decomposed as "Cause causes Item to be higher/lower (on some scale)", with the Eventuality "to be higher/lower" being implicit.

(8)   [Water privatization `Cause`] RAISED [prices `Item`].

While Deng & Wiebe treat increases as a metaphorical case of existence, the evaluative logic behind these cases could also be couched as a functor in the style of Reschke and Anand (2011). Basically, "Increase is good, decrease is bad". For the German equivalent of *raise*, *erhöhen* we would annotate *+Effect:AN* to capture the positive effect on its accusative object. For the antonym *senken*, we would annotate *-Effect:AN*.

4.2   Evaluation of pure states / post-states

The last examples of the previous section suggest that there is no particular reason to focus solely on verbs that contain causation as part of their meaning. Just as we treat both *good* and *improve* (or: *bad* and *deteriorate*) as equally relevant explicit sentiment predicates, so we can treat the non-subjective adjective *high* (or: *low*), the intransitive verb *rise* (or: *decline, fall*) and the transitive verb *raise* (or: *decrease, lower*) as relevant to opinion inference.

As argued by Reschke and Anand (2011), "The evaluation of a change of state is equivalent to the evaluation of the end state.". This can be readily seen by taking into account sets of lexical decompositions such as those for the verbs *raise* and *rise* and the adjective *high* in (9–11), which alert us to series of related predicates and make explicit what the relevant (end) states are.

(9)   causative *raise.v*: [x CAUSE [BECOME [y <high>]]]

(10)   inchoative *rise.v*: [BECOME [y <high>]]

(11)   stative *high.a*: [y <high>]

Regarding the scalar predicates *high* and *low*, we reason "more is good, less is bad". Analogously to the object of the transitive *raise* or the subject of the intransitive *rise*, we thus mark the head noun that is modified by the adjective *high* as being in a positive state. Similarly, we annotate the head of *low* as being in a negative state. Recall that "in a positive state" for *high* is meant to metaphorically take the perspective of the entity referred to by the head, for example *rents* in (12). The overall evaluation of the situation described by (12) also depends on the evaluation of the external viewer's attitude towards high rents: if he or she is a landlord, it will tend to be positive; if he or she is a renter, it is most likely negative.

(12)   She said [rents `Item`] are HIGH in Paris.

The states that are relevant to opinion inference are not necessarily ones with a single participant. There are also cases of relational states that we need to take into account. The verbs of possession covered by the functor in Table 1 are one important subclass. But we can also consider predicates like *similar*, *like*, and *resemble*, which talk about states of similarity ([y <similar> z]).

(13)   Charles Krauthammer said ... "[Putin `Item1`] is LIKE [Hitler `Item2`] but he's more subtle and he's also weaker, ..."

Basically, if there is sentiment towards Item2, then it is imputed to Item1 as well by dint of the comparison (cf. Table 2). By paying attention to lexical decompositions, we know that we can reason in a parallel way about inchoative verbs and causative verbs that denote events with an end state of (dis)similarity. Accordingly, for the causative predicate like *angleichen* 'make sth equal to sth else', we would mark *Effect:AN∼DN* to capture the fact that the Effect on the changed accusative object depends on the nature of the dative object. Note that in this case, unlike with *geben* in (7) the accusative phrase refers to the affected participant. The ordering of the arguments around the "∼"-symbol thus captures information that is lost in Anand and Reschke's notation, where affected participants are not identified.

A second prominent class of relational states are, interestingly enough, inherent sentiment predicates.

(14)   My best friend Martha just told me that [she `Experiencer`] LOVES [Sarah Palin `Stimulus`].

Item1	Item2	*similar*	*differ*
+	+	+	-
+	-	-	+
-	+	+	-
-	-	-	+

Table 2: Functor for predicates of similarity

While the Experiencer's inherent positive sentiment specified by *love* toward the Stimulus-object is clear, the state itself is also open to support opinion inferences. The basic reasoning could be couched as "positive/negative sentiment is useful/harmful for the Stimulus". Accordingly, for the verb *lieben* 'love', we mark the object as being in a positive effect state (*+Effect:AN*).

Example (14) is underspecified. Assuming that the speaker approves of Sarah Palin herself, she will feel positive that her friend shares the sentiment. Assuming that the speaker has so far not liked Sarah Palin, he or she may now have a conflicted attitude towards her friend Martha. Martha's positive sentiment benefits Sarah Palin, and since the speaker does not like Palin, he or she should therefore also disapprove of the source of that (emotional) support for Palin.

A third large class of relational states concerns locative prepositions and causative predicates such as *put, throw, remove* etc. If the (post-)state is valued in some way, then so is the event (and potentially its author). For concrete physical locations, it is, however, often not obvious what value to attach to the Figure and Ground arguments a priori. Even Grounds that come heavily connotated with one polarity, can carry a different polarity, when we take into account the specifics of Figure and Ground (cf. (15) vs. (16)). For metaphorical locations, it seems that the 'Ground' expressions that metaphorically evoke a state are often readily interpretable as to polarity, as in (17). The operative logic for these cases is that "good/bad entities should be in good/bad states" (cf. Table 3).

(15)   [The laundry `Figure`] LAY [in the mud `Ground`].

(16)   [The rhinoceros `Figure`] LAY [in the mud `Ground`].

(17)   ...they DROVE [the company `Figure`] [into the ditch `Ground`].

The German verb *liegen* 'lie' would be marked as *Effect:NN∼BL* to capture the fact that the effect on the Figure, coded as a nominative noun

Figure	Ground	*in*	*out of*
+	+	+	-
+	-	-	+
-	+	-	+
-	-	+	-

Table 3: Functor for predicates of location

phrase (NN), depends on the nature of the Ground, coded as a kind of locative adverbial phrase (BL).

With some locative verbs, we face a certain difficulty because they incorporate their Theme argument. E.g. *asphaltieren* 'to asphalt, tarmac' refers to a transfer of some material onto a surface. The evaluation of tarmacking depends on whether we think the 'Theme' (=asphalt) is appropriately put on the location in question. We can mark the location as affected but the polarity depends on the theme. Since it is not realized explicitly as a phrase, we hint at its relevance by annotating *Effect:AN~Context*.

4.3 Logical operators

While we can often identify states (and specifically, post-states) as the most relevant concept for opinion-inference, it is clear that we need to deal with certain logical operators that can occur inside complex semantic decompositions. Consider *enable* and *possible*.

(18)   This workshop ENABLED delegates to learn more about practical intervention strategies.

(19)   This workshop made it POSSIBLE for delegates to learn more about practical intervention strategies.

(20)   It is POSSIBLE for delegates to learn more about practical intervention strategies.

The causative verb *enable* allows for paraphrases as in (19). If we focus solely on the effect of (19), we get sentence (20). Although possibility is certainly different from existence, the logic is the same: to be possible is good, to be impossible is bad. Accordingly, if we approve of the delegates' learning about practical intervention strategies, we will approve of the workshop. Since permission and requirement can be couched in terms of possibility, the aforementioned logic also applies to predicates such as *obligatory*, *permit* etc.

(21)   It was OBLIGATORY to eat fish on certain fast days laid down by the church.

In (21), we can paraphrase *obligatory* as "not possible not to do X". Assuming eating fish is viewed negatively, not eating it is positive. The possibility of not eating it is positive, too, but once that possibility is negated, we are left with a negatively evaluated situation. Similarly, if in (18) we replaced *enable* by *prevent* (and adjusted other syntax accordingly), we would come to the opposite conclusion because *prevent* can be decomposed as "cause it not to be possible for X to do Y".

4.4 Pragmatic inference vs. lexical sentiment

It is desirable to keep the positive/negative state characterization of a predicate separate from any additional sentiment that the predicate may carry. To modify an example of Reschke and Anand (2011), consider the following pair of sentences:

(22)   Tom injured the burglar.

(23)   Tom abused the burglar.

Both sentences imply a negative effect on the burglar. Given that the burglar is likely evaluated negatively, we could infer that the injuring event and its agent are evaluated positively or at least neutrally in (22). In (23), this is not possible since *abuse* lexically specifies negative evaluation on the abusing event and its author, however deserving of injury the abusee may be. In our annotation, the negative effect on the victim of abuse is preserved. We prefer to let the lexical sentiment information conflict with, and override, the effect-based inference. This makes sense as the negative evaluation of the victim may (at least for some people) constitute an attenuating circumstance. If we ignored the effect altogether, this nuance would be lost.

For some words, the choice between a treatment in terms of lexically inherent sentiment and a treatment as pragmatic inference is difficult to make. As an example, consider the verb *fawn*. On the lexical analysis, the negative characterization of another person's speech as fawning is inherent in the meaning of *fawn*. On the inference account, it just happens that speakers often describe other people's behavior as fawning when they themselves dislike the person that the other people have positive sentiment for. The inference that the speaker disapproves of the fawning and of the fawners would then simply follow from the logic applied to sentiment states (cf. §4.2).

(24)   The critics at Cannes FAWNED all over these like they'd never seen kung fu before ...

5 Inter-annotator agreement

We assess the inter-annotator agreement for the annotation task by two measures: percent agreement and κ (Cohen, 1960). Percent agreement is the ratio of all identically annotated synsets against all annotated synsets. For each annotated synset, we compare the annotations from both annotators. If they are identical, the synset receives a credit of 1, while the synset receives no credit in case they are different. Finally, for all annotated synsets we add up the credit and divide the sum by the number of annotated synsets. Note that for synsets with multiple annotations, we do not consider the order of the individual annotations.

For κ, we proceed as follows. For each GermaNet synset, we extract the phrase labels from the valence frames that come with the example sentences. In order not to artificially inflate or deflate the inter-annotator agreement, we discard any duplicate valence frames that may arise from syntactically identical examples. From the extracted phrase labels, we construct three types of units: (i) phrase labels-only (**PHL**), (ii) phrase label relations (**REL**) and (iii) context-dependent phrase labels (**CON**). **PHL** relates to annotations like +*Effect:AN* where there is positive/negative effect on one phrase label. Units of type **REL** correspond to all pairwise combinations of phrase labels of a given valence frame. An annotation example is *Effect:NN∼AN* where the effect on NN is dependent on the evaluation of AN. Context-dependent effect on an entity, annotated as e.g. *Effect:AN∼Context*, corresponds to **CON**. For this unit, we construct a combination between a given phrase label and the term "Context", connected by the ∼ symbol. Finally, for each annotation of a synset, we project the annotation onto these units. For cases, where there is no match, we ascribe "Default" for no annotation. We then compute Cohen's κ separately for each of the three units.

In total, two annotators, both authors of this paper, independently annotated 726 GermaNet synsets over two annotation phases. In the exploratory phase I, we annotated 226 GermaNet synsets following the annotation guidelines for EffectWordNet by Choi and Wiebe (2014). In phase II, we annotated 500 synsets using the scheme presented in §4. In both phases, for each annotation round, we randomly selected the synsets to be annotated and discussed differences after the annotation and accordingly adjusted the guidelines.

The agreement results are presented in Table 4. Due to different annotation formats, we only report κ values for the annotations from phase II. For phase II, the agreement after the first 100 instances is very good with percent agreement values around 0.8 and κ values between 0.75 and 0.94. The first round (1-100) had low results because it was the first attempt to apply our own more extensive scheme. After each annotation round, we adjudicated the annotations, resolving almost all differences between the annotators. The bottom row in Table 4 shows the agreement results for all annotations *after* adjudication. Note that we manually reannotated the adjudicated synsets from phase I according to the final guidelines, in order to be able to include them in the overall computation of the inter-annotator agreement and for use as additional seed data in future work on label propagation.

It is not possible to directly compare our results to the annotation for EffectWordNet. Though that annotation effort was bigger in that all 1030 WordNet senses of 1143 lemmas were covered, the annotation was only done by a single annotator (Choi and Wiebe, 2014). Thus, no agreement information is available for those annotations. However, the relevant annotator had taken part in a prior annotation study (Choi et al., 2014), where two annotators achieved a κ of 0.75 on 151 senses that were to be labeled as either +effect; -effect; or neutral.

By contrast, we performed double annotation on 726 GermaNet synsets. However, our annotation scheme is more extensive in several respects. It allows for the annotation of more opinion inference-relevant predicates and it takes into account syntactic valence information about the entity affected. Despite this added complexity, we have achieved good agreement results for all of the GermaNet senses that we have annotated so far.

Among the synset members of the 726 annotated synsets, there are 148 unique lemmas with more than one GermaNet synset whose different senses have all been annotated. Following Choi and Wiebe (2014), we conduct an analysis of the effect ambiguity for these lemmas across their different senses. We find 110 of the 148 lemmas (74.3%) to have an inconsistent effect on an affected entity (polarity / affected entity, or both) across their different senses. 26 of these 110 lemmas show effects with different polarities on an affected entity. Consider e.g. *ausstoßen* which entails a positive effect on its object in its 'to emit

	Synsets	Percent agreement	Cohen's Kappa PHL	Cohen's Kappa REL	Cohen's Kappa CON
Phase I	1-59	0.75	n/a	n/a	n/a
Phase I	60-133	0.68	n/a	n/a	n/a
Phase I	134-226	0.76	n/a	n/a	n/a
Phase II	1-100	0.46	0.16	0.17	0.18
Phase II	101-150	0.76	0.87	0.92	0.93
Phase II	151-200	0.68	0.78	0.85	0.78
Phase II	201-250	0.76	0.81	0.86	0.91
Phase II	251-300	0.80	0.80	0.75	0.84
Phase II	301-401	0.80	0.84	0.81	0.85
Phase II	402-500	0.82	0.91	0.90	0.95
Adjud.	1-726	0.97	0.98	0.98	0.99

Table 4: Inter-annotator agreement for phase I (EffectWordNet scheme) and phase II (our scheme). Bottom row: agreement after adjudication.

sth' sense[7] but negative effect on its object in the sense of 'to expel so'. This indicates that effect ambiguity is also prevalent in German and confirms the need for a sense-level lexical resource.

6 Conclusion and Future Work

We have presented an annotation scheme for effect-related information in a lexical resource that significantly extends prior work. We have broadened the coverage, and made more systematic the understanding of effect-related predicates by framing them in terms of lexical decomposition. First, in addition to effects resulting from causing events, we also take into account resulting states of events that need not involve a specific cause (e.g. *fall*) and even simple states and properties (e.g. *high*). Second, the states or relations that occur in effect-related predicates are not limited to ones referring to existence, possession or affectedness. Verbs of location, similarity, and sentiment are relevant, too. Third, our annotation scheme deals explicitly with predicates where the evaluation of an event requires considering a relation between two semantic roles (e.g. *give [me] [a cookie]*, *make [Kim] [happy]*).

We have achieved good levels of agreement given the complexity of the task. In successfully working on German data, we have provided further evidence that opinion inference and the relevance of lexical knowledge to support it are independent of specific languages.

A significant benefit of relating our annotations to example sentences and their syntactic valence descriptions is that we thereby generated sense-disambiguated data that can be used in evaluat-

ing an opinion-inference system for German. The GermaNet data that we have annotated so far will be made available to the research community.[8] In ongoing work, we are finishing the double annotation of the second half of our data set, which we will then also publish. In addition, we are beginning to experiment with ways to propagate the effect information on our seed data throughout GermaNet's verbal and adjectival synsets.

With regard to the annotation scheme, one issue that we have not dealt with so far is that for a given predicate multiple end states could be relevant, depending on the context. As an example consider the synset containing verbs of resource extraction such as *gewinnen* and *fördern*, which can co-occur with arguments realizing the agent, the theme and the source location. On the one hand, the agent's possession that results from oil/gas/mineral extraction may be relevant in some contexts such as thinking about the wealth of nations. On the other hand, the theme's (dis)location can be relevant, for instance, when arguing about whether to extract fossil fuels or leave them in the ground to mitigate climate change. Predicates with multiple relevant post-states may account for some of our annotation differences. Studying this issue more calls for performing annotation of actual corpus instances along the lines of Deng et al. (2013).

Another issue that we are pursuing is the inventory of different functors that are needed to reason about the post-states. Compare the functors for possession in Table 1, for similarity in Table 2, and for location in Table 3. All of them involve two arguments. Ignoring the role names, we see that the functors for possession and location are isomorphic, while that for similarity is different. Given a small number of arguments for a (post-)state and the possible assignments of +/- values to these arguments and to the state, only a relatively small number of functor types is at all possible. The question is which of the possible functors actually occur, and with what frequency.

Acknowledgments

---

[7]For instance, emitting smoke causes the smoke to exist.

References

Pranav Anand and Kevin Reschke. 2010. Verb classes as evaluativity functor classes. *Proceedings of Verb 2010*, pages 98–103.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98, pages 86–90, Stroudsburg, PA, USA. Association for Computational Linguistics.

Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal. 2006. The salsa corpus: a german corpus resource for lexical semantics. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-2006)*, pages 969–974.

Yoonjung Choi and Janyce Wiebe. 2014. +/-effectwordnet: Sense-level lexicon acquisition for opinion inference. In Alessandro Moschitti, Bo Pang, and Walter Daelemans, editors, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1181–1191. ACL.

Yoonjung Choi, Lingjia Deng, and Janyce Wiebe. 2014. Lexical acquisition for opinion inference: A sense-level lexicon of benefactive and malefactive events. In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 107–112, Baltimore, Maryland, June. Association for Computational Linguistics.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1):37–46.

Lingjia Deng, Yoonjung Choi, and Janyce Wiebe. 2013. Benefactive/malefactive event and writer attitude annotation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 120–125, Sofia, Bulgaria, August. Association for Computational Linguistics.

Christiane Fellbaum. 1998. *WordNet*. Wiley Online Library.

Song Feng, Ritwik Bose, and Yejin Choi. 2011. Learning general connotation of words using graph-based algorithms. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1092–1103. Association for Computational Linguistics.

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.

Verena Henrich, Erhard Hinrichs, and Tatiana Vodolazova. 2012. Webcage: a web-harvested corpus annotated with germanet senses. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 387–396. Association for Computational Linguistics.

Roberto Navigli and Simone Paolo Ponzetto. 2010. Babelnet: Building a very large multilingual semantic network. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 216–225. Association for Computational Linguistics.

Malka Rappaport Hovav and Beth Levin. 1998. Building verb meanings. *The projection of arguments: Lexical and compositional factors*, pages 97–134.

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

Josef Ruppenhofer, Michael Ellsworth, Miriam R.L. Petruck, Christopher R. Johnson, and Jan Scheffczyk. 2010. *FrameNet II: Extended Theory and Practice*. International Computer Science Institute, Berkeley, California. Distributed with the FrameNet data.

Carol L Tenny. 1994. *Aspectual roles and the syntax-semantics interface*. Springer.

Janyce Wiebe and Lingjia Deng. 2014. An account of opinion implicatures. *CoRR*, abs/1404.6491.