Published in: Calzolari, Nicoletta et al. (eds.): Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016). Paris: European Language Resources Association (ELRA), 2016. p. 2879-2887.

Effect Functors for Opinion Inference

Josef Ruppenhofer, Jasper Brandes
Universität Hildesheim
Universitätsplatz 1, 31141 Hildesheim
{ruppenho, brandesj}@uni-hildesheim.de

Abstract

Sentiment analysis has so far focused on the detection of explicit opinions. However, of late implicit opinions have received broader attention, the key idea being that the evaluation of an event type by a speaker depends on how the participants in the event are valued and how the event itself affects the participants. We present an annotation scheme for adding relevant information, couched in terms of so-called effect functors, to German lexical items. Our scheme synthesizes and extends previous proposals. We report on an inter-annotator agreement study. We also present results of a crowdsourcing experiment to test the utility of some known and some new functors for opinion inference where, unlike in previous work, subjects are asked to reason from event evaluation to participant evaluation.

Keywords: Opinion Inference, Effects, Subjectivity

1. Introduction

So far, sentiment analysis has mainly focused on the detection of *explicit* opinions. However, recently the relevance of *implicit* opinions has received broader attention within the field. Here, the work of Klenner and colleagues on verb polarity frames (Klenner et al., 2014; Klenner, 2015), the functor-based framework of event evaluation by Anand and Reschke (Anand and Reschke, 2010; Reschke and Anand, 2011) and the work on opinion implicature by Wiebe and her colleagues (Deng et al., 2013; Deng and Wiebe, 2014; Wiebe and Deng, 2014), based on the effects that events have on object entities, have been the most important contributions to the theory of implicit opinions.

The key observation of this line of work is that, while some predicates do not *denote* sentiment, in combination with expressions of explicit sentiment and knowledge about attitudes towards entities, one can infer implicit opinions being conveyed that hinge on these inference-relevant predicates. Consider (1). From the explicit positive sentiment (in **boldface**) by Peter towards the fact that the Colts lost, an event which affects them negatively (in SMALL CAPS), one can infer that he has a negative sentiment towards the Colts. Furthermore, depending on the context, one may infer that the speaker/writer producing (1) may have a positive or negative attitude towards Peter, depending on whether the speaker disapproves or approves of the Colts.

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

While we expect a sentiment lexicon to provide us with information that e.g. *happy* expresses positive sentiment, the information that the effect-bearing word *lose* entails negative affect on the subject can so far not be retrieved from any lexical-semantic resource.

In this contribution, we push the effort begun by Ruppenhofer and Brandes (2015) toward a synthesis between the effect-based approach and the functor approach further in several respects. Firstly, we identify and propose additional functors for verbs embedding states of *possibility*, predicates expressing *location*, and predicates expressing *similarity*. These functors allow us to cover predicates whose entailments have so far not been allowed for by Anand and Reschke's functors (Anand and Reschke, 2010; Reschke and Anand, 2011). We added these new functors because they seem to us to be relevant to opinion inference, which we empirically confirmed for some of them (cf. Section 7.). Unlike Ruppenhofer and Brandes (2015), we add explicit annotations of functor types to the annotations. In doing so, we also allow for multiple functors applying to the same effect predicate. Moreover, we do not only index the arguments relevant to the end state (=effect), but also those encoding the cause or agent.

The remainder of this paper is structured as follows. Section 2 discusses related work. In Section 3, we present the data that we annotated following the annotation scheme outlined in Section 4. We develop a systematic account of the inventory of possible functor types and their interrelationships in Section 5. We assess the reliability and validity of this extended annotation scheme via an inter-annotator agreement experiment in Section 6. We next report on a crowdsourcing experiment that we performed to assess the utility of some of the well-known and some of our new functors for opinion inference (cf. Section 7). We conclude the paper and discuss directions for future work in Section 8.

2. Related Work

Resources that contain information relevant to opinion inference are rather sparse, compared to the abundance of sentiment lexica for English (Wilson et al., 2005; Taboada et al., 2011) and German (Klenner et al., 2009; Remus et al., 2010; Waltinger, 2010). In fact, even for English there seems to exist only one resource for this task, namely +/-EffectWordNet (Choi et al., 2014). This resource is structured on the sense level, building on WordNet (Miller et al., 1990) as its underlying structure. The effect information in +/-EffectWordNet is intended to feed into a rule-based inference machinery along the lines of the ideas laid out in Wiebe and Deng (2014), which computes implicit opinions from the presence of effect information and explicit sentiment within a sentence, similar to the reasoning used in our discussion of (1) above. Note that

+/-EffectWordNet in its current form does not include (morphosyntactic or semantic) information with regard to the entity affected. Inspired by +/-EffectWordNet, Ruppenhofer and Brandes (2015) began work on a German resource, annotating GermaNet (Hamp and Feldweg, 1997) synsets with information relevant to opinion inference.

3. Data

We use the same GermaNet V9.0 (Hamp and Feldweg, 1997) synsets as Ruppenhofer and Brandes (2015). These authors extracted 1667 synsets for annotation in order to have gold standard data for automatic methods.[1] Note that the relevant synsets include verbs and adjectives, but no nouns. Figure 1 displays an instance of a GermaNet synset, along with its annotations. We discuss the meaning of the annotations in the next section.

4. Annotation Scheme

Figure 1 shows an example of our annotations. As in the earlier work of Ruppenhofer and Brandes (2015), the key parts of the scheme, which is exemplified in Figure 1, 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 comes with a syntactic subcategorization frame which lists the set of arguments and adjuncts occurring with the lemma being exemplified. Thus, in the example sentence in Figure 1, *übergeben* 'turn sth over to sb' is realized with a noun phrase in nominative case (NN), another in dative case (DN), and a third in accusative case (AN). 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 the possession functor applies to the verb. Further, the effect polarity on the participant coded as a dative noun phrase (DN) representing the Possessor is dependent on the evaluation of the Possession coded as an accusative case noun phrase (AN). We mark this dependence with the "∼" symbol. The participant coded as a nominative phrase (NN) is causally responsible for the effect on the Possessor.

Note that the annotations of Ruppenhofer and Brandes (2015) do not name the specific functor type (e.g. possession) and neither do they include the marking of a causally responsible participant, if there is one.

[1]For information on how Ruppenhofer and Brandes (2015) selected the relevant synsets, we refer the interested reader to their paper.

4.1. Functor Inventory

Following Reschke and Anand (2011), we treat predicates and their arguments as functors that map tuples of argument and verb properties to state evaluations. For instance, the functor embedding a state of possession in Table 1 is to be read as follows. The first line of the table applies to the situation when there is a possessor who is positively evaluated by an external viewer (e.g. 'her best friend') as well as a positively evaluated possession (e.g. 'a good job'). If the relationship between the two is *have* (third column), as in (2), the state of possession is evaluated positively by the external viewer. If the relationship is *lack* (fourth column), the state is evaluated negatively. Note that here and with other functors, labels such as *have* and *lack* are intended only as mnemonics for the intended abstract relations. They do not stand for specific lexical items that other lexical items might entail.

(2) [Mein Freund]$^{\text{POSSESSOR}}$ **hat** [einen guten Job]$^{\text{POSSESSION}}$.
'My boyfriend has a good job.'

The notion of an external perspective can be clarified when considering the third row of Table 1. Here, we have a negatively evaluated possessor (e.g. 'her worst enemy') and a positively evaluated possession (e.g. 'a good job'). If the relationship between the two is *have*, we would evaluate the situation negatively even though the possessor may actually be very happy about having a good job.

Of the functors that have been proposed by Anand and Reschke (2010), the main ones are those for possession (Table 1), existence (Table 2) and affectedness (Table 3).

4.1.1. Possession

This functor covers verbs entailing a state of possession, e.g. *spenden* 'to donate sth to sb' or *stehlen* 'to steal sth from sb'.

Possessor	Possession	*have*	*lack*
+	+	+	-
+	-	-	+
-	+	-	+
-	-	+	-

Table 1: Functor for verbs embedding a state of possession

4.1.2. Existence

This functor covers verbs embedding a state of existence/non-existence, e.g. *erbauen* 'to construct sth' or *zerstören* 'to destroy sth'.

Entity	E_{exist}	E_{non_exist}
+	+	-
-	-	+

Table 2: Functor for verbs embedding a state of existence

4.1.3. Affectedness

Verbs embedding a state of affectedness, e.g. *putzen* 'to clean sth' or *schlagen* 'to beat sb', are covered by this functor.

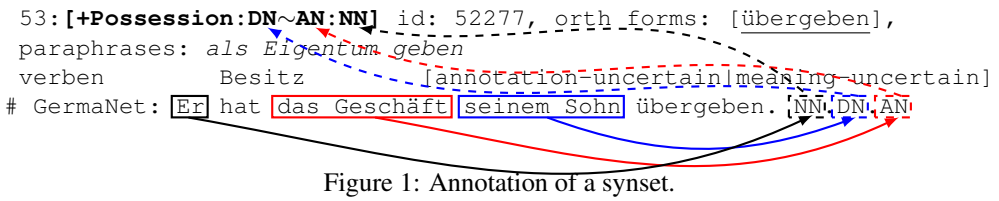

Figure 1: Annotation of a synset.

Entity	$E_{positive}$	$E_{negative}$
+	+	-
-	-	+

Table 3: Functor for verbs embedding a state of affectedness

In addition to the above three, we recognize further functors, including ones for possibility (Table 4), location (Table 5), similarity (Table 6), sentiment (Table 7), and scalarity (Table 8).

4.1.4. Possibility

This functor covers predicates embedding a state of possibility, e.g. *möglich* 'possible', *ermöglichen* 'to facilitate sth' or *verhindern* 'to prevent sth'.

Event	$E_{possible}$	$E_{impossible}$
+	+	-
-	-	+

Table 4: Functor for predicates embedding a state of possibility

4.1.5. Location

This functor covers predicates entailing a state of location, e.g. *auftragen* 'to spread sth' and *abräumen* 'to clear away (the table)'.

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

Table 5: Functor for predicates expressing location

4.1.6. Similarity

This functor covers predicates expressing similarity, e.g. *angleichen* 'to assimilate' and *unterschiedlich* 'different'.

4.1.7. Sentiment

This functor covers predicates expressing sentiment, e.g. *mögen* 'to like sb' and *verschmähen* 'to despise sb/sth'.

4.1.8. Scalarity

Reasoning along the lines of 'more is good, less is bad', this functor covers predicates embedding a state of occupying a position on a scale, e.g. *hoch* 'high', *erhöhen* 'to increase', and *verringern* 'to diminish sth'.

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

Table 6: Functor for predicates expressing similarity

Experiencer	Stimulus	*love*	*hate*
+	+	+	-
+	-	-	+
-	+	+	-
-	-	-	+

Table 7: Functor for predicates expressing sentiment

Attribute	A_{high}	A_{low}
+	+	-
-	-	+

Table 8: Functor for verbs embedding a state of occupying a scalar position

A straightforward observation about the functors is that several of them are structurally identical. For instance, all the functors involving a single argument are the same (Tables 2, 3, 4, 8). We return to this point and also motivate the functor inventory in Section 5.

4.2. Explicit Functor Naming

To actually be able to use relational annotations of the form "DN~AN" illustrated in Figure 1, we need to know which functor to apply to them as not all functors involving two arguments are identical. Compare the similarity functor in Table 6 to that for possession in Table 1. For instance, when the first argument (Item1/Possessor) is evaluated negatively and the second argument is evaluated negatively (Item2/Possession), too, then events of possession or similarity are not evaluated the same, as shown by row 4 of the relevant tables. Knowing the functor is still not sufficient, however. For verbs of possession, we need to know whether they positively entail possession or its absence. Otherwise,

we could not distinguish between *have* and *lack*. Accordingly, we additionally annotate for each relational synset, whether it entails the holding of the relation or its not holding. In Figure 1 the "+" preceding the functor name indicates that the verb *übergeben* entails a positive possession relation. By contrast, the verb *vorenthalten* 'withhold' entails the absence of a possession relation.

4.3. Multiple Functors

As observed in the analysis of annotation disagreements by Ruppenhofer and Brandes (2015), certain predicates may be related to more than one functor. We encountered further evidence for this finding in the annotation experiment for this paper (cf. Section 6.). For instance, verbs of placing like *treiben* 'drive sth into sth' and *brennen* 'burn sth onto sth' can co-occur with arguments realizing the agent, the theme and the target location. There may be contexts in which the theme's creation is more salient, for example, when burning your initials into a tree. On the other hand, the correct placing of the theme in the target location may be more prominent in situations when considering whether bedrock is suitable for driving a tunnel into. These examples suggest that, if annotators can perceive such multiple possibilities based on the example sentences in GermaNet, they should not be forced to choose just a single functor.

4.4. Annotation of Causes

No previous effort related to opinion inference has dealt with the problem of how to identify causally responsible roles. The theory of Wiebe and her collaborators is focused on affected objects while the functor approach of Anand and Reschke does not label the participants as either affected or causally responsible: it simply maps the set of participants to an event evaluation polarity.

Trying to determine causally responsible participants heuristically is not trivial. First, we would need to know explicitly from information in the resource that there is a causally responsible party at all. Given the range of items we annotate, we cannot simply assume that this will be the case. For instance, if one finds a treasure by happenstance or one steals it, one ends up in possession of it. However, in the case of the accidental finding, nobody need have acted volitionally. Thus, although finding and stealing both have a possession entailment, the former has no causally responsible participant. Second, even if the predicates in a synset do not have core arguments expressing causally responsible participants, such participants can be introduced via adjuncts. Consider the set of paraphrases in examples (3)-(5). The causal role is an adjunct in the first example but an argument in the last two.

(3) stative *high.a*: [y <high>]
My rent is high/15% higher because of city taxes.

(4) inchoative *rise.v*: [BECOME [y <high>]]
My rent rose by 15% because of city taxes.

(5) causative *raise.v*:
[x CAUSE [BECOME [y <high>]]]
City taxes raised my rent by 15%.

A full treatment of opinion inference would thus also require that adjuncts expressing causal meanings are taken into account. Further, while we did not encounter such cases in our example sentences, note that in principle more than one causal expression may be present. For instance, a causal argument and a cause-related adjunct may co-occur as in (6) and (7), where both *he* and the *out of*-PP relate to the causation of the event.

(6) He stole the money out of greed.

(7) He stole the money out of love.

The above pair of examples illustrates that the interaction between a reason adjunct and an agent argument is not uniform: a bad reason reinforces the causal responsibility (culpability) of the agent, as in the case of (6), while a good reason would seem to reduce it, as shown by (7).

At this time, we limit ourselves to annotating GermaNet synsets relative to what is provided in the syntactic valence frames that come with the example sentences. However, we still want to distinguish between predicates that are strictly stative and therefore do not involve a causally responsible agent and predicates that do allow a causal agent but which may not be realized in the provided valence frames. For the former, we leave the annotation slot for the causal agent empty, while for the latter, we annotate "NI" to mark a non-instantiated causal agent (adjunct) for the predicate.

5. Towards a Typology of Functors

Comparing the location functor in Table 5 to that for possession in Table 1, we see that they are structurally identical. We might say that the Possessor is comparable to the Figure ('x') and the Possessum to the Ground ('y'). But actually, for the purposes of calculating event evaluation, we could switch what we treat as 'x' and 'y': the functor's output does not change because it fires based on sameness and difference. Generally, it is interesting that we observe a similarity between location and possession since this fits observations by other linguists of similarities between possession and location (e.g. Wunderlich (2012)). A classical case is the fact that many ditransitive verbs can occur with two different syntactic patterns: one where Recipient and Theme arguments occur in the so-called double object construction expressing possession (*give x y*) and another prepositional construction expressing location (*give y to x*), where the Figure occurs as object and the Ground as PP. Similarly, we can observe that the functors for similarity (cf. Table 6) and for sentiment (cf. Table 7) are the same. We could thus in theory collapse all functors sharing the same structure into a single more abstract functor. However, we refrain from doing so because the functor labels convey additional information that is useful for other purposes. For instance, events causing outright non-existence are probably judged more intensely positive or negative than events resulting in negative or positive affectedness.

More generally, given the structure of a functor for a predicate with 2 arguments, there are 4 feature values per predicate to fill. If we treat picking the values like tossing a coin, there are 2^4=16 possible feature combinations. Of these, we can eliminate half as being, for our purposes, negations of the remaining half.

Role1	Role2	?	?	?	?	?	?	*love*	*have*
+	+	+	+	+	+	-	+	+	+
+	-	+	+	+	-	+	+	-	-
-	+	+	+	-	+	+	-	+	-
-	-	+	-	+	+	+	-	-	+

Table 9: Possible functors for predicates entailing two-argument states (question marks represent unattested predicate types)

	Percent	Cohen's Kappa				
Synsets	**agreement**	**PHL**	**REL**	**CON**	**FUNC**	**CAUS**
1-50	0.66	0.91	0.93	0.90	0.86	0.93
51-100	0.58	0.82	0.84	0.84	0.85	0.90
101-150	0.66	0.90	0.88	1.0	0.76	0.92
151-200	0.66	0.96	0.92	0.97	0.91	0.85
201-300	0.70	0.88	0.86	0.90	0.85	0.94
1-300*	0.91	0.98	0.98	0.99	0.96	0.97

Table 10: Inter-annotator agreement. *Bottom row: agreement after adjudication.

Of the possible 8 output columns, we have found two to be attested (cf. Table 9). The functor exemplified by possessive *have* requires *congruence* to output positive evaluation. The functor exemplified by sentiment-bearing *love* represents cases where the evaluation of one of the two arguments wins out completely and the features of the other argument are ignored. A functor that always outputs positive (or negative) evaluation, as shown in column 3, would be a defective case, linguistically speaking, because the usual trigger factors would play no role: there would be nothing to implicate (for the speaker) or infer (for the hearer) in a context-dependent manner. And in fact, we think of verbs like *murder*, which would seem to fit the bill, as having lexically inherent negative evaluation.[2] It is an empirical question whether the other possible functors in columns 4-8 occur. What is important is that they, too, are cases where the polarity towards the (post)state does not simply depend on one of the two arguments. If we find instances of these cases in our data, we will need to expand our functor inventory to be able to deal with them.

6. Annotation Experiment

The authors of this paper independently annotated 300 GermaNet synsets from scratch. The synsets were annotated in four rounds of each 50 synsets and one final annotation round of 100 synsets. After each round, we discussed differences and adjudicated the annotations. For the inter-annotator agreement (cf. Table 10), we compute the percent agreement and Cohen's κ (Cohen, 1960) along the lines of Ruppenhofer and Brandes (2015). In addition to phrase label (PHL), relational (REL) and contextual (CON) annotations as described in that paper, we also report on agreement results for the functor (FUNC) and causal entity (CAUS) annotations. Recall that this is the first effort of effect annotations that also considers the functor applied and the entity causally responsible.
We achieved good agreement across all types of annotations.

7. Crowdsourcing Opinion Inferences

In order to empirically assess the utility of the functors presented in Section 4, we conducted a crowdsourcing experiment in which we asked speakers to judge the inferred evaluative stance of an author towards a participant of an event. In previous work, Reschke and Anand (2011) tested the predictions implicit in their proposed functors (existence, affectedness, possession) using constructed sentences in which the participants in the argument slots of each predicate are canonically positive (e.g. *hero, cathedral*), negative (e.g. *villain, torture chamber*), or neutral (e.g. *man, building*). The authors presented annotators with a sentence such as in (8) and asked them to assess the overall evaluation of the author towards the event described in the sentence as being either *positive, negative* or *neutral*.

(8) The villain murdered the child.

Reschke and Anand (2011) report high inter-annotator agreement (κ = 0.92) for the predictions related to the affectedness and existence functors: "that is, killing was judged more positive when the entity losing existence was an enemy and judged more negative when it was an ally". With regard to the possession functor, results seemed to be less clear-cut (κ = 0.68) for positively evaluated possessors possessing a positively evaluated possession (e.g. "a hero gaining a valuable watch") and negative possessors showing evaluations similar to neutrally judged possessors.
In contrast to Reschke and Anand (2011), we run the inference process in the opposite direction: we provide the speaker's overall event evaluation but leave the participants of the functor predicates underspecified. Subjects are then asked to guess the speaker's evaluation of one of the participants. Depending on the details of the specific functor at issue, we expect to find either a preference for a specific kind of polarity towards the participant in question, or considerable variation of the polarity, if the event evaluation is compatible with both a positive or negative towards the participant in question. With the latter cases, we are interested to find out whether raters choose among the two possible polarities with more or less equal likelihood or whether we can find evidence of biases or default preferences.

7.1. Study Design

We collect our judgments as part of larger surveys in which we also elicit intensity ratings for complex phrases. Each survey contains approximately 40 utterances which are to be judged for their evaluative stance. There is roughly the same number of intensity ratings in each survey. The two types of questions serve to distract study participants from focusing too hard on a single main task within the survey. Items are randomized and presented individually to the participants. For each item, we elicit judgments from 20 individuals.[3] We use a local installation of the open source survey software LimeSurvey and distribute the surveys to

[2] The verb *kill* would be a different case.

[3] Due to a technical error, we obtained judgments from 21 subjects for some items.

English native speakers located in the US via the crowdsourcing website `prolificacademic.co.uk`. Average completion time for the surveys varies from 10 to 21 minutes and we pay each user between 2.40 Euro and 2.80 Euro, depending on the number of items in a survey. Each user can only participate in one survey so as to avoid any learning effects or biases.

We test the predictions of four functors, the existence and possession functors by Anand and Reschke (2010) and two of the functors proposed in this paper: sentiment and similarity. We construct the utterances for the survey items by varying the stimuli along the dimensions shown in Table 11. However, for cost reasons we did not elicit data for all the combinatorial possibilities.

Element	Values
Functor	Existence, Possession, Sentiment, Similarity
Functor polarity	Affirmed, Denied
Event evaluation	Positive, Negative
Relevant argument	Arg1, Arg2

Table 11: Dimensions of variation of the survey items

An utterance from one of the surveys is presented in (9). In this example, we assess the *denied* version of the *similarity* functor with the predicate ('to assimilate'). The overall event evaluation is *explicitly positive* as expressed by the adverb 'luckily'. The relevant argument of the predicate is in the *Arg2* position ('the surrounding culture') and subjects are asked to provide the speaker's implicit evaluation of this participant.

(9) Luckily, the immigrants haven't assimilated to [the surrounding culture].
(similarity; positive; denied; Arg2)

Let us, for the sake of clarity, repeat the similarity functor from Table 6 in Table 12. Using this functor, we can now predict the implicit evaluation by the speaker in example (9) towards the phrase in brackets ('the surrounding culture'). From the fact that the similarity functor is denied ('haven't assimilated'), we know that we first need to look into the fourth, bold-faced column of Table 12. The positive evaluation ('luckily') by the speaker towards the event as a whole indicates that we now need to focus on the second and fourth rows of the table (i.e. the dot-circled rows). This information allows us to derive the speaker's negative implicit attitude towards the item in brackets which equates to 'Item2' in Table 12 (i.e. the shaded cells).

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

Table 12: Functor for predicates expressing similarity

When guessing the speaker's evaluation of the participant in brackets, study participants could choose between five possible responses: 'positive', 'negative', 'neutral', 'mixed', and 'cannot tell'. In addition, users may leave a comment for each judgment. Figure 2 displays a screenshot of a survey item as it was presented to the participants of the study.

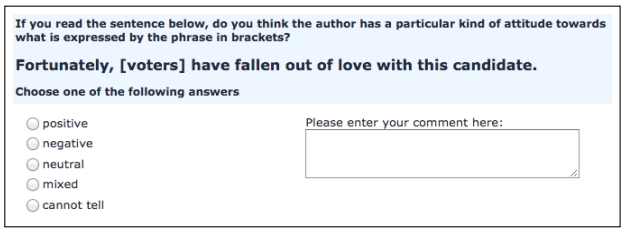

Figure 2: Screenshot of a survey item

The utterance in Figure 2 is an example that contains a predicate expressing sentiment. Again, we can clarify the composition of the utterance by referring to Table 11. Here, we have an eventive predicate and the sentiment functor is *denied* ('have fallen out of love'). The overall event evaluation by the speaker is *explicitly positive* as realized by the adverb ('Fortunately'). In contrast to example (9), the relevant argument to be judged by the study participants is in the *Arg1* position ('voters'). The sentiment functor in Table 7 shows that both a positive and a negative evaluation of *Arg1* are compatible with the positive event evaluation (cf. second and fourth line).

7.2. Results

We present the crowdsourcing results separately for each functor for the features presented in Table 11. Hence, we contrast *positive* vs. *negative* event evaluation, *affirmed* vs. *denied* functor polarity, and *Arg1* vs. *Arg2* position of the functor participant to be judged. Thus, for each functor, we investigate 8 sentences that differ in the realizations of these features.

For ease of exposition, we oppose *positive* against *negative* event evaluation. This allows us to represent the data in a two-dimensional fashion as in Table 13 in which, for a given overall event evaluation, we contrast the argument position of the relevant functor participant (in columns) against the *affirmed* (Aff) and *denied* (Den) functor (in rows).

The cells in this table contain, in that order, the response frequencies for a *positive, negative*, and *unbiased* evaluation by the author of the utterance towards the phrase in brackets. Note that for the *unbiased* category, we conflate the three responses 'neutral', 'mixed', and 'cannot tell'.

event evaluation		Argument Arg1	Arg2
Functor	Aff	pos/neg/unbiased	pos/neg/unbiased
	Den	pos/neg/unbiased	pos/neg/unbiased

Table 13: General format of the crowdsourcing results

7.2.1. Existence

The upper left cell in Table 14 captures the evaluation of an agent or causal force in a positively evaluated act of creation such as the one reported in (10).

(10) Luckily, John built a swimming-pool.

Unsurprisingly, the main response here was positive evaluation of that actor. The second most common response was neutral evaluation, which may seem somewhat surprising if causal forces are usually evaluated in the same way as the resulting state. However, recall that what functors provide are defeasible implicatures rather than firm entailments. It is thus not too surprising when speakers choose to refrain from inferring a biased evaluation towards a participant in the absence of any contextual support. Further, there may have been an issue that some raters did not strongly perceive causal responsibility and intentionality on the part of the agent, given that the items were rated in isolation, without further context. Had we used a graduated scale allowing for more or less strong positive and negative evaluation, we might have obtained further non-neutral responses.
The lower left cell in Table 14 displays the ratings of an agent or causal force in a positively evaluated act of non-creation such as the one reported in (11). The results parallel those in the left upper cell: if there is a valenced assessments, then non-actors/refrainers tend to get credit for not bringing about negatively evaluated states of affairs.

(11) Luckily, John didn't build a swimming-pool.

The upper right cell of Table 14 shows the evaluation of *Arg2*, the created entity, when the overall event's evaluation is positive (cf. (10)). The results are as expected.
The lower right cell of Table 14 shows the evaluation of *Arg2* in a non-creation act when the overall event's evaluation is positive (cf. (11)). The results here are not as clear: the functor predicts that we should find overwhelmingly negative assessments of *Arg2*, but we find quite a few positive assessments and many neutral ones. It is not clear, if the positive responses result from a positivity bias introduced by the adverb *luckily*; inattention/random response; or some implicit context that those raters assumed. For instance, if one believes that John doesn't do a good job of building things, then one may be happy he didn't build a swimming-pool while maintaining a positive view of swimming-pools.
We now turn to the evaluation of participants in negatively evaluated acts of creation. The results in Table 15 display the expected symmetry to the results for the positively evaluated acts shown in Table 14. For instance, the upper right cell of Table 15 shows that the created entity tends to be negatively evaluated if that is true of the event overall. As with the positively evaluated events, we find unexpected results for *Arg2* in the lower right cell, that is, when its non-creation is negatively evaluated. We would have expected overwhelmingly positive responses but in fact find those (7) to be slightly outnumbered by the negative responses (9).

pos.	Arg1	Arg2
Aff	11/1/8	18/0/2
Den	10/2/8	4/6/10

Table 14: Existence; positive event evaluation

neg.	Arg1	Arg2
Aff	0/11/10	0/18/3
Den	1/12/8	7/9/5

Table 15: Existence; negative event evaluation

7.2.2. Possession

Results for positively evaluated states of possession are shown in Table 16. According to the possession functor, a positive evaluation of an event, as in (12), is compatible with a positively valued possessor possessing a positively valued possession (cf. row 1 of Table 4) as well as with a negatively valued possessor possessing a negatively valued possession (cf. row 4 of Table 4). However, lacking context, our raters seem to default to the positive constellation, overwhelmingly rating both Arg1 and Arg2 positively.

(12) Luckily, Pat got a stereo.

For positively evaluated situations where somebody did not obtain a possession, as illustrated by (13), the functor again predicts that two different constellations are conceivable: (a) a deserving possessor having something negatively evaluated, or (b) an undeserving possessor having something desirable. The ratings for the possessor in the lower left cell suggest that speakers favored interpretation (a) over (b). The ratings for the possession in the lower right cell show more balanced results.

(13) Luckily, Pat didn't get a stereo.

We now turn to the negatively evaluated cases in Table 17. In row 1, where the state of possession does hold, possessors and possessions tend to be rated negatively. But note that the upper left and the upper right cell represent distinct and incompatible constellations: negativity towards the possessor corresponds to line 3 of the possession functor, where the possession is positively valued. By contrast, negativity towards the possession corresponds to line 2 of the possession functor, where the possessor is positively valued. Thus, the raters were not consistent across the two aligned ratings of the two participants.

pos.	Arg1	Arg2
Aff	14/1/5	18/0/2
Den	10/3/7	7/6/7

Table 16: Possession; positive event evaluation

neg.	Arg1	Arg2
Aff	1/11/8	1/16/3
Den	4/7/9	8/5/7

Table 17: Possession; negative event evaluation

For negatively evaluated situations in which somebody lacks a possession, as illustrated by (14), we again get rather mixed results, as shown by row 2 of Table 17.

(14) Sadly, Pat didn't get a stereo.

7.2.3. Sentiment

Table 18 shows the results for positively evaluated (positive) sentiment, as illustrated by (15).

(15) Fortunately, voters have fallen in love with this candidate.

As predicted by the functor, Arg2, is mostly understood to be evaluated positively (upper right cell) when positive sentiment holds. While the functor does not uniquely predict positive or negative evaluation for the combination of positive event evaluation and positive evaluation of Arg2 (cf. lines 1 and 3 in Table 7), raters favor a positive interpretation.
Row 2 of Table 18 represents the situation where the not holding of positive sentiment is positively evaluated, as illustrated by (16).

(16) Fortunately, voters have fallen out of love with this candidate.

The ratings suggest that the raters by default assume the constellation where a positively evaluated Experiencer (Arg1) rightly withholds positive sentiment from an undeserving Stimulus (Arg2), rather than the case where a negatively evaluated Experiencer withholds positive sentiment from an undeserving Stimulus.

pos.	Arg1	Arg2
Aff	13/1/7	17/2/1
Den	14/0/7	3/15/2

Table 18: Sentiment; positive event evaluation

neg.	Arg1	Arg2
Aff	0/16/4	3/15/2
Den	3/9/8	10/8/2

Table 19: Sentiment; negative event evaluation

Turning now to the negatively evaluated situations shown in Table 19, we observe that for cases where the Experiencer's positive sentiment toward a Stimulus is negatively evaluated from the external point of view, both Experiencer and Stimulus tend to be negatively evaluated by the raters (cf. line 4 of Table 7). The alternative constellation where the Experiencer is still viewed positively while the Stimulus is evaluated negatively seems not to have been salient (cf. line 2 of Table 7).

In the case of negative evaluation of a situation where positive sentiment is withheld (row 2 of Table 19), the results are more mixed, which is as expected for the Experiencer but not so for the Stimulus (cf. lines 1 and 3 of Table 7).

7.2.4. Similarity

The results for positively evaluated similarity and difference are shown in Table 20, those for the negatively evaluated counterparts in Table 21.

When similarity is positively evaluated, raters assume that both participants themselves must be positively valued, too (cf. row 1 of Table 20), even though the functor does not require that of Item1 (cf. lines 1 and 3 of Table 6). When difference between Item1 and Item2 is positively evaluated (cf. row 2 of Table 20), as in (17), Item2 is predicted to be negatively evaluated by the functor for similarity (cf. lines 2 and 4 Table 6). However, we do find surprisingly many positive ratings (6) for Item2.

(17) Fortunately, the immigrants haven't assimilated to the surrounding culture.

When similarity is negatively evaluated, raters assume that both participants themselves must be negatively valued, too (cf. row 1 of Table 21), even though the functor does not require that of Item1 (cf. lines 2 and 4 of Table 6). When difference / lack of similarity is evaluated negatively, as in (18), we would expect to find Item2 positively evaluated. However, that is not what we found. Instead, Item2 was clearly rated negatively by the majority of raters.

(18) Unfortunately, the immigrants haven't assimilated to the surrounding culture.

pos.	Arg1	Arg2
Aff	14/0/6	16/1/3
Den	13/5/2	6/9/5

Table 20: Similarity; positive event evaluation

neg.	Arg1	Arg2
Aff	0/16/4	2/13/5
Den	2/11/7	4/11/5

Table 21: Similarity; negative event evaluation

7.3. Summary

Our crowdsourcing results suggest several points worth considering in future working on opinion functors. First, the results for affirmed situations (e.g. states of possession) typically are clearer than for denied situations (e.g. states of lack). This may reflect a difference in complexity between processing affirmed and negative propositions. Second, even though many binary functors yield the same event evaluation for two distinct constellations of how the participants are rated, we find that speakers in the rating task often seem to default to one of these constellations. It would be interesting to test on corpus data if these defaults reflect actual tendencies in regular language use. Third, we found some completely unexpected results. It is not clear if this is a) a result of using non-experts; b) owed to our stimuli; or c) to additional assumptions that raters may introduce in their understanding because of the lack of context. Controlling for such factors should shed further light on when the functors allow reliable conclusions and when not.

8. Conclusion and Future Work

We have extended the theory of effect functors for opinion inference, introducing new functors and developing a typology of possible functors. We have reported on the first effort to comprehensively annotate effect-relevant predicates with (i) the functor applied, (ii) the affected, and (iii) the causal entity. This information is essential for any rule-based opinion inference machinery to work effectively.[4] We have demonstrated through an inter-annotator-agremeent study that we can annotate the various functors with good agreement. Finally, we performed a crowdsourcing experiment in which we tested the utility of various functors for opinion inference in a novel setting, where event evaluations were given, the arguments carried no obvious sentiment and subjects had to infer the evaluation of a specified participant.

In terms of resource-building, we plan to use our gold standard annotations for approaches that automatically annotate previously unlabeled synsets by either leveraging the GermaNet graph or using methods from distributional semantics. With respect to the effect functors themselves, we intend to perform additional experiments, including ones with contextualized sentences, to gain a better sense of why certain inferences are less robust than expected.

9. Acknowledgements

Part of this work was supported by the German Research Foundation (DFG) under grant RU 1873/2-1.

[4] With the final version, we will release our annotations of all 1667 synsets.

10. Bibliographical References

Anand, P. and Reschke, K. (2010). Verb classes as evaluativity functor classes. In *Proceedings of Verb 2010 - Interdisciplinary Workshop on Verbs*, pages 98–103.

Choi, Y., Deng, L., and Wiebe, J. (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 (WASSA)*, pages 107–112. Association for Computational Linguistics.

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

Deng, L. and Wiebe, J. (2014). Sentiment propagation via implicature constraints. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 377–385, Gothenburg, Sweden, April. Association for Computational Linguistics.

Deng, L., Choi, Y., and Wiebe, J. (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.

Hamp, B. and Feldweg, H. (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. Association for Computational Linguistics.

Klenner, M., Fahrni, A., and Petrakis, S. (2009). Polart: A robust tool for sentiment analysis. In *Proceedings of the 17th Nordic Conference on Computational Linguistics (NODALIDA 2009)*, pages 235–238.

Klenner, M., Amsler, M., and Hollenstein, N. (2014). Verb polarity frames: a new resource and its application in target-specific polarity classification. In *Proceedings of the 12th Edition of the Konvens Conference*, pages 106–115. Universität Hildesheim.

Klenner, M. (2015). Verb-centered sentiment inference with description logics. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 134–139, Lisboa, Portugal, September. Association for Computational Linguistics.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. (1990). Wordnet: An on-line lexical database. *International Journal of Lexicography*, 3:235–244.

Remus, R., Quasthoff, U., and Heyer, G. (2010). SentiWS – a publicly available German-language resource for sentiment analysis. In *Proceedings of the 7th International Language Resources and Evaluation (LREC'10)*, pages 1168–1171.

Reschke, K. and Anand, P. (2011). Extracting contextual evaluativity. In *Proceedings of the Ninth International Conference on Computational Semantics*, pages 370–374. Association for Computational Linguistics.

Ruppenhofer, J. and Brandes, J. (2015). Extending effect annotation with lexical decomposition. In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 67–76, Lisboa, Portugal, September. Association for Computational Linguistics.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., and Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2):267–307.

Waltinger, U. (2010). GermanPolarityClues: A lexical resource for German sentiment analysis. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC)*, Valletta, Malta, May.

Wiebe, J. and Deng, L. (2014). An account of opinion implicatures. *CoRR*, abs/1404.6491.

Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 347–354. Association for Computational Linguistics.

Wunderlich, D., (2012). *The Oxford handbook of compositionality*, chapter Lexical Decomposition In Grammar, pages 307–327. Oxford University Press.