Published in: Dipper, Stefanie/Neubarth, Friedrich/Zinsmeister, Heike (eds.): Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016). Bochum: Ruhr-Universität Bochum, 2016. p. 226-235.
(Bochumer Linguistische Arbeitsberichte 16)

Verifying the robustness of opinion inference

Josef Ruppenhofer
Institute for German Language
Mannheim, Germany
ruppenhofer@ids-mannheim.de

Jasper Brandes
Hildesheim University
Hildesheim, Germany
brandesj@uni-hildesheim.de

Abstract

There is increasing interest in recognizing opinion inferences in addition to expressions of explicit sentiment. While different formalisms for representing inferential mechanisms are being developed and lexical resources are being built alongside, we here address the need for deeper investigation of the robustness of various aspects of opinion inference, performing crowdsourcing experiments with constructed stimuli as well as a corpus study of attested data.

1 Introduction

In recent years, sentiment analysis has seen increasing interest in inferring implicit opinions in addition to capturing explicit expressions of opinion. Work 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 pointed up the great potential of opinion inference: speakers and authors leave many implicit opinions for hearers to infer. In (1), we can infer, for instance, that the speaker felt negatively about having the flu, if we assume that she values herself and has a negative attitude towards the flu. Further, we can infer that she has a negative attitude towards the flu shot that she deems causally responsible for getting the illness.

(1) The last time I got a flu shot, it GAVE me the flu.

However, corpus annotation studies and subsequent efforts to acquire lexical acquisition for opinion inference have left certain questions about the robustness of the inferences unaddressed. The one that we take up here is the limited range of potential inference types that have been empirically evaluated so far. Existing studies have focused on predicates related to (1) creation/destruction, (2) possession/lack and (3) affectedness and they have typically tested inference about event evaluation given knowledge about participant evaluation. However, as argued by Ruppenhofer and Brandes (2016), additional classes of predicates give systematic rise to opinion inferences, for instance, predicates related to similarity and location. In our crowd-sourcing experiments, we include these new classes of predicates. Further, we look at inferences in the 'opposite' direction, going from event evaluation toward participant evaluation. We also explore inferences in several kinds of less prototypical constellations. For instance, we look at concessive situations, in which a good or bad situation fails to be prevented. Similarly, we explore whether inferences only arise when an event producing a resultant state is explicitly mentioned, or also when pure states are presented, as implied by the work of Reschke and Anand (2011).

2 Related work

There exist several related but distinct approaches to sentiment inference. Two key ones are the work of Klenner and colleagues on verb polarity frames (Klenner et al., 2014; Klenner, 2015; Klenner and Amsler, 2016) and the work by Wiebe and colleagues on effect-based inference (Deng et al., 2013; Choi and Wiebe, 2014; Deng and Wiebe, 2014). The work of Klenner and colleagues is focused on effects on participants, whereas we are interested in the evaluation of an event by external viewers. The work of Wiebe and colleagues shares our perspective but due to its specific approach has a limited coverage compared to the approach that we adopt, functor-based inference.

2.1 Opinion inference based on functors

Reschke and Anand (2011) explored the relationship between the lexical semantics of predicates and the attitudes that speakers are inferred to have towards the events referred to by those predicates. They treat predicates and their arguments as func-

tors that map tuples of argument and verb properties to evaluations. An example is given in Table 1. The first row 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*, that relation is valued positively (left grey cell). If the relation is *lack*, that relation is valued negatively (right grey cell). The table shows that the reasoning for *lack* also applies to events of withholding and depriving which result in lack. Note that the possessor x of *withhold* and *deprive* is the grammatical object of these verbs in active-form sentences rather than the subject as in the case of *have* and *lack*. However, this difference is unimportant to the logic that applies.

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

Table 1: Functors for verbs embedding a state of possession

Two considerations are important to keep in mind. First, the goal of the inference procedure is to assess the attitude of an **external viewer** on the event. For instance, while in (2) the external viewer Sue may feel negatively towards a situation where a person she dislikes, *x*, got something desirable, *y*, the relevant possessor, Peter, will most likely feel positively about the award he got.

(2) Sue is **disappointed** Peter WON the award.

Second, the inference procedure must be **context-dependent** and be capable of producing different results, at least under some circumstances. In other words, there cannot be an inference that always goes through and yields the same polarity. That would not be a contextual inference but simply part of inherent lexical meaning.

Ruppenhofer and Brandes (2016) adopt the functor idea, proposing new functors for additional classes of verbs, among them predicates of location, similarity and sentiment.

Location This functor covers predicates entailing a state of location, e.g. *in/out of*; *at/away from* and *enter/exit*.

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

Table 2: Functor for predicates expressing location

Sentiment This functor covers predicates expressing sentiment, e.g. *love/hate* and *fall {in/out of} love*.

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

Table 3: Functor for predicates expressing sentiment

Similarity This functor covers predicates expressing similarity, e.g. *similar/different* and *assimilate/deviate*.

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

Table 4: Functor for predicates expressing similarity

We will use the Similarity functor in our crowdsourcing experiments. The functor underlies examples such as 3, which may be used to criticize the addressee for sharing traits with a parent.

(3) You're just `like` {your father/mother}!

2.2 Evaluation of functor-based inference

In the work of Reschke and Anand (2011), the usefulness of the predictions implicit in the proposed functors (existence, affectedness, possession) was tested 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 stimulus such as in (4) and asked them to assess

as *positive, negative* or *neutral* the author's overall evaluation of the event described in the sentence.

(4) The villain `murdered` the child.

Reschke and Anand (2011) report high inter-annotator agreement ($\kappa = 0.92$) 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". For the possession functor, results seemed to be less clear-cut ($\kappa = 0.68$) 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.

Ruppenhofer and Brandes (2016) report some results on crowd-sourcing experiments, in which they evaluate for several functors how consistently human ratings match the predictions that the functors make. In their experiments, the parameters that we vary in our experiments are stable: they always use eventive predicates, clearly biased sentence adverbs, canonical positioning of roles, and they always focus on the evaluation of the roles in the *entailed* relation. For instance, in the case of the possession functor, experiments test how well the functor predictions match human ratings for Possessor and Possession, but not for the Donor causing the entailed possession relation. Our experiments test the robustness of functor-based inference in a significantly broader range of constellations.

3 Experimental Design

While Reschke and Anand (2011) tested only the inference of event evaluation, we, like Ruppenhofer and Brandes (2016), run the functor-based inference process in the opposite direction: we fix the speaker's overall event evaluation but leave one or both of the participants of the functor predicates underspecified. Subjects are asked to guess the speaker's evaluation of one of the participants whose description is unbiased. Figure 1 shows a sample item illustrating this design. 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 could leave a comment for each judgment.

Depending on the specific functor at issue, we expect to find either a preference for a specific kind of polarity towards the participant in question, or to see considerable variation of the polarity, if the event evaluation is compatible with both a positive or negative attitude towards the participant in question. For the latter cases, we are interested to find out whether raters choose among the possible polarities with more or less equal likelihood or whether we can find evidence of biases or default preferences. Consider the example given in Figure 1 involving the sentiment functor, shown in Table 3 above. We can derive the expected evaluations(s) of the target role (i.e. the phrase in brackets) that we would expect to see as the responses of our subjects for the stimulus. 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 conveyed by the adverb (*fortunately*). The relevant argument to be judged by the study participants is in the *Arg1* position (*voters*). The sentiment functor in Table 3 shows that in theory both a positive and a negative evaluation of *Arg1* are compatible with the positive event evaluation (cf. second and fourth line).

The judgments we analyse are collected as part of larger surveys in which we also elicit intensity ratings for words and phrases. Each survey has about 40 utterances which are to be judged in terms of their evaluative stance. There is roughly the same number of intensity ratings in each survey. The two types of questions serve to mutually distract study participants from each other. Our items are randomized and presented singly to the participants. For each item, we collect judgments from 20 individuals.[1] We use a local installation of the LimeSurvey software and distribute the surveys to English native speakers registered in the US via the crowdsourcing website `prolificacademic.co.uk`. On average our surveys took between 10 to 21 minutes to complete. We paid each user between 2.40 Euro and 2.80 Euro, depending on the number of items in a survey. Each user could only participate in one survey in order to avoid learning effects or bias.

A full analysis of the factors impacting event evaluation would have to consider at least the parameters shown in Table 5, which we will briefly present. The `functor` parameter simply refers to which functor is relevant for the predicate at issue. `Functor polarity` refers to the question

[1] Due to a technical error, we sometimes received responses from one or two additional subjects.

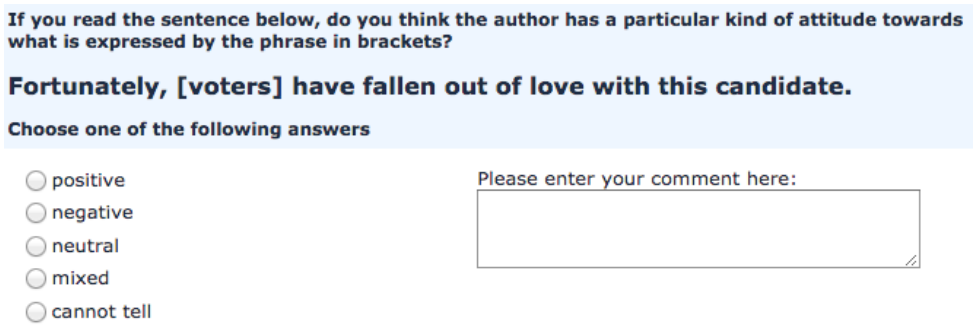

Figure 1: Screenshot of a survey item

Element	Values
Functor	Existence, Possession, Sentiment, Similarity, ...
Functor polarity	Affirmed, Denied
Event evaluation	Positive, Negative, Un(der)specified
Sentence polarity	Affirmed, Denied
Relevant role	Arg1, Arg2, Causal, Concessive
Placement of adjunct role	in place, fronted
Biassing of other arguments	Yes, No

Table 5: Element inventory for constructing the survey items

whether the state of affairs referred to by the functor is affirmed or denied. E.g. for the adjective *similar* the affirmed version of the similarity functor is relevant, for the adjective *different* the denied version. The parameter `event evaluation` keeps track of whether the event is explicitly evaluated negatively or positively (cf. (5)), or whether it is under- (cf. 6) or unspecified (cf. (7)).

(5) Unfortunately/Fortunately, John `got` the job.

(6) Surprisingly, John `got` the job.

(7) John `got` the job.

`Sentence polarity` refers to whether the predicate is within the scope of syntactic negation. Accordingly, (8) is a case of the denied possession functor, while the affirmed functor applies to (6).

(8) Surprisingly, John didn't `get` the job.

The parameter `relevant role` tracks for which role we are interested in the author's asssessment. Examples (9)–(12) illustrate the roles we consider here.

(9) Unfortunately, [John] didn't `get` the job. (Arg1)

(10) Unfortunately, John didn't `get` [the job]. (Arg2)

(11) Unfortunately, John didn't `get` the job [because of his uncle]. (Causal)

(12) Unfortunately, John didn't `get` the job [despite his uncle]. (Concessive)

The `placement` parameter lets us distinguish for adjuncts whether they are placed in their default location (in place) or whether they are fronted. Thus, the causal argument is fronted in (13) but in place in (11).

(13) Unfortunately, [because of his uncle] John didn't `get` the job . (Fronted)

The use of the `placement` parameter is motivated by the fact that in combination with negation, the interpretation of certain adjuncts is potentially ambiguous when they are in place. For instance, (11) may mean one of two things:

(14) John got the job but this is not so for reasons to do with his uncle. The speaker evaluates John's getting the job for the wrong reasons negatively. (Cause > Neg)

(15) John didn't get the job and this is so for reasons to do with his uncle. The speaker

evaluates John's missing out on the job negatively. (Neg > Cause)

Finally the `biassing` parameter indicates whether any of the other arguments or adjuncts are specified for polarity. (16) has such a biased argument (*stupid John*), unlike its counterpart with a neutral Arg2 (10).

(16) Unfortunately, stupid John didn't `get` [the job]. (Arg2)

Although there are many parameters that we would ideally all control for, here we will only look at several small, focused contrasts as we were not able collect the full data set needed for a global analysis, given the funds available to us. We construct the utterances for the survey items by varying the stimuli along some (but not all) of the dimensions shown in Table 5.

4 Results

We present the crowdsourcing results separately for each of our component studies. We always contrast *positive* vs. *negative* event evaluation, *affirmed* vs. *denied* functor polarity, and two settings for a third parameter. Often the third parameter concerns the argument roles (e.g. *Arg1* vs. *Arg2*) of the functor to be judged. Thus, for each functor, we investigate 8 sentences that differ in the realizations of these features.

For reasons of comprehensibility, we present two result tables with counts per study, one for *positive* and one for *negative* event evaluation. This allows us to represent the data in a two-dimensional fashion, as shown in Table 6.

event evaluation		Parameter X Value 1	Parameter X Value 2
Functor	Aff	pos/neg/unbiased	pos/neg/unbiased
	Den	pos/neg/unbiased	pos/neg/unbiased

Table 6: General format of a crowdsourcing result table

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

4.1 Causal and concessive adjuncts

Previous work on opinion inference has focused on the derivation of an event's evaluation from evaluations of its participants. Ruppenhofer and Brandes (2016) focused on inference about participants, given the event evaluation. However, they focused on the participants in the entailed relation. For example, for predicates with a possession entailment, they looked at evaluations of the possessor and the possession. By contrast, inferences about the donor were not tested.

Here, we look specifically at roles that have to do with the causation of the entailed relation. The simple case are expressions that refer to a causal force bringing about the event and thereby its entailed relation. In cases like (17), one can simply project the event's (positive) evaluation on the causal force that is responsible for bringing about the event.

Concessive expressions are more complicated. These refer to situations or events that took place, and whose taking place would ordinarily lead one to expect that the situation in the main clause does not hold. Nevertheless, the situation expressed by the main clause does hold. In other words, concessive expressions (clauses or prepositional phrases) talk about cases of failed prevention. For instance, in (18), one understands (i) that it is true that the immigrants did not assimilate; (ii) that the group's efforts were aimed at preventing the non-assimilation; and (iii) those efforts failed. The evaluation of the (failed) counter-force that is expressed in the concessive clause, thus, ordinarily should be the opposite of that of the event that took place (i.e. the event that was not prevented). For example (18), one should thus expect a negative judgment about *the group's efforts*, given that they were aimed at preventing an event that the speaker approves of.

(17) Fortunately the immigrants have `assimilated` to the surrounding culture [because of the group's efforts]. (affirmed, causal)

(18) Luckily the immigrants haven't `assimilated` to the surrounding culture [despite the group's efforts]. (denied, concessive)

Tables 7 and 8 show the elicited results. We report both the raw counts and three measures of entropy, in bits, that reflect the consistency of the crowd in a single number. The entropy is zero when one of the outcomes is certain, that is, when all responses agree. We report the overall entropy for the three possible responses (3-way); the entropy

pos.	cau	con
Aff	18/0/2	10/9/1
Den	15/3/2	6/11/3

(a) Counts

	3-way		P v N		U v ¬U	
pos.	cau	con	cau	con	cau	con
Aff	0.47	1.23	0	1.00	0.47	0.28
Den	1.05	1.41	0.65	0.94	0.47	0.61

(b) Entropy

Table 7: Evaluation of causal and concessive roles for *similarity* functor given positive event evaluation

neg.	cau	con
Aff	1/15/6	5/13/3
Den	1/13/7	6/7/8

(a) Counts

	3-way		P v. N		U v. ¬U	
neg.	cau	con	cau	con	cau	con
Aff	1.09	1.32	0.34	0.85	0.85	0.59
Den	1.17	1.58	0.37	1.00	0.92	0.96

(b) Entropy

Table 8: Evaluation of causal and concessive roles for *similarity* functor given negative event evaluation

of the probability distribution for just the positive - negative opposition (P v. N); and the entropy for the distribution of unbiased vs biased (U v. ¬U). For the 3-way entropy, the range of values is (roughly) [0,1.59); for the other two entropy measures, it is [0,1.0].

The responses are much as expected for causal roles: they are mainly rated positively or negatively in line with the specified event evaluation. For concessive roles, the situation is less clear. For instance, in response to stimulus sentence (18) we would have expected to see overwhelmingly negative judgments of the concessive role (i.e. *the group's efforts*). Yet, as the lower right cell in Table 7 shows, we find quite a few (6) positive judgments relative to the expected negative ones (11).

4.2 Stative versus eventive predicates

The results shown in Tables 7 and 8 in Section 4.1 came about in response to stimuli which expressed a change of state. Now, we want to test the assumption that a change of state is not necessary for the functor reasoning to apply. Accordingly, in Tables 9 and 10 we present results that are derived from stimuli that are parallel in all respects to those for which results are reported in Tables 7 and 8, except that they are based on *stative* predicates. In other words, rather than use the predicate *assimilate* and its negation, we use the adjectives *similar* and *different*.

pos.	cau	con
Aff	14/1/5	7/6/7
Den	15/1/4	6/6/8

(a) Counts

	3-way		P v. N		U v. ¬U	
pos.	cau	con	cau	con	cau	con
Aff	1.08	1.58	0.35	0.99	0.81	0.93
Den	0.99	1.57	0.34	1.00	0.72	0.97

(b) Entropy

Table 9: Evaluation of causal and concessive roles for *similarity* functor given positive event evaluation

neg.	cau	con
Aff	1/14/6	1/10/10
Den	2/8/11	4/6/11

(a) Counts

	3-way		P v. N		U v. ¬U	
neg.	cau	con	cau	con	cau	con
Aff	1.12	1.23	0.35	0.44	0.86	1.00
Den	1.34	1.46	0.72	0.97	1.00	1.00

(b) Entropy

Table 10: Evaluation of causal and concessive roles for *similarity* functor given negative event evaluation

We see the same pattern of results for stative predicates that we saw for eventive predicates. But it appears that the eventive predicates yielded somewhat clearer judgments, at least for the causal roles. There are fewer unbiased responses with eventive predicates (Tables 7–8) than with stative predicates (Tables 9–10), which is reflected by lower entropy values for U v. ¬U in the former tables.

4.3 Canonical placement versus fronting

As pointed out above, placement might play a role in how adjuncts are interpreted. Here we specifically consider instances of causal adjuncts, as illustrated above in (11) and (13). The predicates we

use have a possession entailment and we use both affirmed and denied instances in combination with positive or negative event evaluation. The results are shown in Tables 11 and 12.

pos.	can	fro
Aff	17/0/4	19/1/1
Den	11/1/9	13/2/6

(a) Counts

	3-way		P v. N		U v. ¬U	
pos.	can	fro	can	frol	can	fro
Aff	0.70	0.55	0	0.29	0.70	0.28
Den	1.22	1.27	0.41	0.57	0.99	0.86

(b) Entropy

Table 11: Evaluation of causal role for possession functor given positive event evaluation

neg.	can	fro
Aff	1/15/4	3/15/3
Den	1/18/1	1/19/1

(a) Counts

	3-way		P v. N		U v. ¬U	
neg.	can	fro	can	fro	can	fro
Aff	0.99	1.15	0.34	0.65	0.72	0.59
Den	0.57	0.55	0.30	0.29	0.29	0.28

(b) Entropy

Table 12: Evaluation of causal role for possession functor given negative event evaluation

The results are rather heterogeneous, with no clear picture emerging. We do not consistently observe lower entropy values for fronted placement.

4.4 Unbiased event evaluation

The experiments above and those of Ruppenhofer and Brandes (2016) all used *biased* event evaluation via sentence adverbs such as *unfortunately*, *luckily*, etc. Here, we report on a simple control experiment in which we use a sentence adverb that bears no inherent polarity, namely *surprisingly*. We are looking at affirmed and denied instances of the similarity functor, for two different causation-related roles, namely adjuncts expressing a means or a concessive. Two example sentences are given in (19) and (20).

(19) Surprisingly, the immigrants have/haven't `assimilated` to the surrounding culture by [adopting the local customs]. (affirmed/denied, means)

(20) Surprisingly, the immigrants have/haven't `assimilated` to the surrounding culture [despite the party's efforts]. (affirmed/denied, concessive)

Given the unbiased nature of the sentence adverb, we predict responses to be neutral in the main, and to vary randomly between positive and negative among the non-neutral responses.

neu.	means	concessive
Aff	9/0/11	9/6/5
Den	4/3/13	5/4/11

(a) Counts

	3-way		P v. N		U v. ¬U	
neu.	mea	con	mea	con	mea	con
Aff	0.99	1.54	0	0.97	0.99	0.81
Den	1.28	1.44	0.99	0.99	0.93	0.99

(b) Entropy

Table 13: Evaluation of means and concessive roles for similarity functor given unbiased event evaluation

The first prediction that non-biased responses are in the majority is borne out for three of our constellations, as can be seen from Table 13. The exception are affirmed cases, where we ask about concessives. The preference for non-biased responses is, however, not very pronounced as shown by the high entropy values for U v. ¬U.

With regard to the second prediction, that the biased responses would be split rather evenly between positive and negative, this is borne out in most cases. The clear exception are affirmed cases in which we ask about the means role. Here, no negative evaluations of the means of assimilation were produced, resulting in an entropy of 0 for the P/N opposition. Potentially, the problem here lies with our stimulus: our raters might intrinsically all have favored the idea of immigrants assimilating and thus projected that positive attitude onto the means by which the assimilation is accomplished.

5 Corpus study

In sections 3 and 4, we investigated the robustness of opinion inference experimentally. In this section, we want to shed some light on attested instances

Count	*relieved*	*glad*
Existence	34	17
Location	24	27
Possession	16	21
Possibility	7	5
Sentiment	5	16
Affectedness	4	5
n/a	60	59
Total	150	150

Table 14: Functors embedded under *relieved*

of opinion inferences in corpora. To that end, we analyze clauses embedded under the predicates *relieved* and *glad*, which both provide positive event evaluation towards the situations expressed by embedded predicates. Example (21) shows an instance of the predicate *present*, which has an existence entailment that is negated in context, embedded under *glad*; example (22) illustrates a case where a predicate with a negative possession entailment, *conclude*, is embedded under the predicate *relieved*.

(21) I 'm **glad** that didn't present insurmountable problems as , although having suffered over the final volumes of the original " Dune " series I somehow was n't expecting too much , it turned out to be an extremely enjoyable story .

(22) Joan Keane , GMB Regional Organiser , said : " Whilst Mr Williams is **relieved** that the matter is now concluded , he has endured years of bullying and harassment by his colleagues ...

Table 14 shows the distribution of functor types embedded under 150 instances of each of the two predicates *glad* and *relieved*. The instances were randomly sampled from the uKwaC corpus (Ferraresi et al., 2008) and classified by the first author. The table shows that while the Existence and Possession functors proposed by Anand and Reschke (2010) are frequent, the Location functor is, too. Of the other new functors introduced by Ruppenhofer and Brandes (2016), only Sentiment is attested in the sample, but not, for instance, Similarity. [2]

The crowdsourcing experiments suggested that speakers employ certain defaults when reasoning about constellations of event and participants evaluations, where the latter are unspecified. Accordingly, it is interesting to ask if the default interpretations observed in the experiments match those that apply to naturally occurring instances where participant evaluation is unspecified.

We begin by considering the instances of the Possession functor in our samples for *glad* and *relieved*. For both predicates, of the four possible constellations that are compatible with positive event evaluation (cf. gray shaded cells in Table 15), only two occur, with a stark frequency difference among them: the constellation of positive evaluation for both participants and the event itself clearly predominates, which matches the results that Ruppenhofer and Brandes (2016) got when eliciting judgments for parallel, artificially constructed stimuli where the participants were described neutrally and only the event evaluation was explicitly biased.

		relieved		*glad*	
Possessor	P.ion	*have*	*lack*	*have*	*lack*
+	+	+/14	-	+/19	-
+	-	-	+/2	-	+/2
-	+	-	+/0	-	+/0
-	-	+/0	-	+/0	-

Table 15: Possession functor instances embedded under *relieved* and *glad*

We find similar asymmetries for the Location functor as we saw for the Possession functor. As Table 16 shows, the first constellation, where a positively valued Figure is at a positively valued Ground, predominates. However, we seem to find more variety than for Possession.

		relieved		*glad*	
Figure	Ground	*in*	*out of*	*in*	*out of*
+	+	+/13	-	+/22	-
+	-	-	+/8	-	+/3
-	+	-	+/2	-	+/2
-	-	+/1	-	+/0	-

Table 16: Location functor instances embedded under *relieved* and *glad*

For predicates with an Existence entailment, the distribution is as shown in Table 17. For the instances embedded under *glad*, we find a stark asymmetry, as we had before for the other functors. The constellation where existence of a positively valued

[2] The category "n/a" is assigned to instances where the main predicate of the embedded clause cannot be assigned to one of the known functors.

entity is valued positively is much more frequent. There is only instance of the other pattern, where non-existence of a negatively valued entity is evaluated positively by the external viewer, namely (21). By contrast, for *relieved*, the distribution of instances among the positively evaluated constellations is much more even and instances such as (22) are much more common.

	relieved		*glad*	
Entity	*exist*	*not exist*	*exist*	*not exist*
+	+/19	-	+/16	-
-	-	+/15	-	+/1

Table 17: Existence functor instances embedded under *relieved* and *glad*

The difference between *glad* and *relieved* is amenable to explanation. *Relieved* references a situation where an Experiencer feels positively about the fact that something (more) positive rather than something (more) negative happened. In talking about relief, one can highlight either the negative event that *did not* happen or the positive event that *did* happen, but the other viewpoint is always presupposed. Accordingly, we find many more references to instances of the negative functors (not being at a place, not existing) for *relieved* than for *glad*: the latter has no presupposition that a potential negative situation did not come to pass.

Overall, our preliminary corpus study supports the idea that not all constellations covered by a functor are equally frequent and that speakers and hearers may operate with default interpretations in elicitation tasks. The contrasts observed between *glad* and *relieved*, however, suggest that there may not be a global default that applies regardless of the specifics of the embedding predicate that specifies event evaluation. Stimulus construction for experimental tasks thus needs to pay attention to the rich lexical semantics of embedding predicates.

6 Conclusion and Future Work

In this work, we performed several crowdsourcing experiments in which we explicitly evaluated several key aspects of the functor-based framework for opinion inference. First, we established the relevance of the newly introduced similarity functor to opinion inference. Second, we tested opinion inferences that start with given event evaluations and target the evaluation of specific participants/roles. Here, we looked specifically at causal and concessive adjuncts, finding that the latter were less reliably evaluated in the way we had predicted than causal adjuncts. Other results gave evidence that opinion inference does indeed apply both to stative and their related eventive predicates alike, thus confirming Reschke and Anand (2011)'s intuition to that effect. We also performed a control experiment confirming that, given unbiased event evaluation, participant evaluation is either unbiased or varies more or less randomly between positive and negative polarity.

However, much remains to be done to firmly establish how reliable opinion inferences are, or what factors impinge on them. For instance, our experiments on the fronted or regular placement of causal and concessive adjuncts offered no real support for the idea that causal roles in fronted position lead to more consistently biased responses than in canonical (typically, final) position. Likewise, the results of all our experiments show, as does the work of Ruppenhofer and Brandes (2016), that the inferences produced for denied functors (e.g. *different*, *not assimilate*) tend to be less clear than those for affirmed functors (e.g. *similar*, *assimilate*). This is unexpected since, in terms of the logic of functors, the denied cases equally lead to predictable results. Both these last two findings may have resulted from our artificial setting, where context was lacking, even though both the use of fronted placement and negation are very much context-dependent. Testing on naturally occurring instances sampled from corpora might help resolve these and other questions.

To complement our crowdsourcing results, we performed a small corpus study to investigate the question where the default interpretations come from that were observed both in the study of Ruppenhofer and Brandes (2016) and in the present work. Our results suggest that the default values in the elicitation settings may derive from the usage patterns in naturally occurring, contextualized instances of opinion inference. However, the contrasts that we observed between our two event evaluation predicates, *glad* and *relieved*, suggest that there may be slightly different patterns of default reasoning used for different classes of embedding predicates that express event evaluation.

Acknowledgements

The authors were partially supported by the German Research Foundation (DFG) under grants RU 1873/2-1.

References

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

Yoonjung Choi and Janyce Wiebe. 2014. +/-EffectWordNet: Sense-level Lexicon Acquisition for Opinion Inference. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1181–1191, Doha, Qatar, October. Association for Computational Linguistics.

Lingjia Deng and Janyce Wiebe. 2014. An investigation for implicatures in chinese : Implicatures in chinese and in english are similar ! In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 8–17, Baltimore, Maryland, June. Association for Computational Linguistics.

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.

Adriano Ferraresi, Eros Zanchetta, Marco Baroni, and Silvia Bernardini. 2008. Introducing and evaluating ukwac, a very large web-derived corpus of english. In *Proceedings of the 4th Web as Corpus Workshop (WAC-4) Can we beat Google*, pages 47–54.

Manfred Klenner and Michael Amsler. 2016. Sentiframes: A resource for verb-centered german sentiment inference. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France, may. European Language Resources Association (ELRA).

Manfred Klenner, Michael Amsler, and Nora Hollenstein. 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.

Manfred Klenner. 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.

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

Josef Ruppenhofer and Jasper Brandes. 2016. Effect functors for opinion inference. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia, may. European Language Resources Association (ELRA).

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