Hope and Fear: Interpreting Perspectives by Integrating Sentiment and Event Factuality

Chantal van Son, Marieke van Erp, Antske Fokkens, Piek Vossen

VU University Amsterdam
De Boelelaan 1105 HV Amsterdam, the Netherlands
c.m.van.son@student.vu.nl, marieke.van.erp, antske.fokkens, piek.vossen@vu.nl

Abstract

Both sentiment and event factuality are fundamental information levels for our understanding of events mentioned in news texts. Most research so far has focused on either modeling opinions or factuality. In this paper, we propose a model that combines the two for the extraction and interpretation of perspectives on events. By doing so, we can explain the way people perceive changes in (their belief of) the world as a function of their fears of changes to the bad or their hopes of changes to the good. This study seeks to examine the effectiveness of this approach by applying factuality annotations, based on FactBank, on top of the MPQA Corpus, a corpus containing news texts annotated for sentiments and other private states. Our findings suggest that this approach can be valuable for the understanding of perspectives, but that there is still some work to do on the refinement of the integration.

Keywords: perspectives, sentiment, event factuality, corpus annotation, MPQA Corpus, FactBank

1. Introduction

Over recent decades, the style of news journalism has shifted from being mainly descriptive to becoming increasingly interpretive. When reporting about events, journalists are expected to go beyond the facts and provide analyses and interpretations, such as explanations of the reasons behind the events or speculations about future consequences. One of the strategies journalists use for this is to tell the story through the eyes of the (groups of) people that are involved in the chain of events. Their stance towards, or *perspective on* the events therefore play a central role in reporting events in newspapers. Consequently, the knowledge we acquire about situations in the world is based on their beliefs. Not only do the participants make claims about the truthfulness of a statement, they also express their view as to whether the statement should be considered as positive or negative. Both of these aspects of perspectives are therefore valuable information when trying to extract and interpret information about events from news texts.

Within the field of Natural Language Processing (NLP) the area that focuses on the identification of opinions, emotions and sentiments is referred to as *sentiment analysis* or *opinion mining* (Pang and Lee, 2008). The level of information that conveys whether an event mentioned in the text corresponds to the actual world or not is referred to as *veridicality* or *event factuality*. Most research so far has focused on either modeling factuality (Saurí and Pustejovsky, 2009) or opinions (Wiebe et al., 2005).

In this paper, we propose a model in which sentiment is combined with event factuality for the extraction and interpretation of perspectives expressed in news texts. By doing so, we can divide the information into negative and positive views on the actual world and negative and positive views on any speculated world, such as the future. Likewise, we provide a better understanding of the causes of an event or the possible consequences, as well as the different sides of the debate around reported events.

In this paper, we investigate the benefits that can be gained from bringing sentiment and event factuality together for the modeling of perspectives through a corpus annotation study. We describe an annotation scheme that we use to apply factuality annotations, based on the FactBank annotation scheme (Saurí and Pustejovsky, 2009), to a set of articles from a corpus that has already been annotated for opinions: the MPQA Opinion Corpus (Wiebe et al., 2005). By adding this semantic layer to the MPQA Corpus we aim to integrate two important aspects of perspectives that we believe are both fundamental for the reasoning about events in news texts. This is illustrated in sentences (1) and (2), which are taken from the MPQA Corpus.

1. [Zimbabwean President Robert Mugabe's reelection] has been **praised** by *the Organization of African Unity (OAU), African countries*, but **condemned** by some Western countries.

2. *The United Nations' secretary-general, Kofi Annan*, **urged** [Zimbabweans to refrain from violence].

In 2002, Mugabe won the heavily contested presidential election in Zimbabwe. After the results, many feared violence if voters thought the election was rigged. In the MPQA Corpus, the positive attitude of the African countries towards Mugabe's reelection is annotated in (1), as well as the negative attitude of some Western countries towards the same event. In (2), the positive attitude of Kofi Annan towards the Zimbabweans refraining from violence is annotated. What is not annotated in MPQA is that the reelection is an event that has actually happened, whereas the (refraining from) violence is merely a possibility in the future. Moreover, we are not able to extract from these annotations the information that Kofi Annan is only positive about the *refraining* of the violence, but is negative about the possibility of the *violence* becoming reality. By adding factuality information about the events, we aim to gain a more complete understanding of the perspectives held by

several (groups of) people on these events. However, since we will be working with a corpus that is not created with the intention to add factuality information at a later stage, it is to be expected that we will encounter some problems in the integration of the two dimensions.

Our contributions are twofold. First, we define a theoretical framework that combines opinions and factuality (Section 3.). In this framework, we explain the way people perceive changes in (their belief about) the world as a function of a their fears of changes to the bad or their hopes of changes to the good. Second, we extend the MPQAv2.0 Corpus with factuality annotations (Section 4.) to investigate whether the dimensions can be effectively integrated and what we gain from this for the interpretation of perspectives on events. In Section 5., we discuss the results of our annotation efforts. In Section 6., we will discuss our findings and outline the next steps for future research.

2. Related work

Annotating text corpora with information of interest is a common step for the development and evaluation of NLP systems that automatically extract a variety of semantic information. These corpora do not only provide training and test material for statistical and machine learning approaches, but also give insight into the ways in which the information can be expressed in natural language and how it should be translated into a comprehensive but workable conceptual model.

The Penn Proposition Bank (PropBank) (Palmer et al., 2005) a corpus annotated with predicate argument structures was, for instance, created following a purely practical approach to semantic representation. The aim was to facilitate the automatic extraction of semantic roles by taking the variation in the syntactic realization of semantic arguments into account, but without the intention of confirming or disconfirming any particular semantic theory. Another example is TimeBank (Pustejovsky et al., 2003), which is annotated according to the TimeML scheme for annotating times, events and temporal relations in texts. The authors emphasize that, although TimeML may not be fully comprehensive yet, it does contribute to the study of how temporal information is expressed in language and it creates possibilities for experiments with automatic extraction systems.

Factuality or information that is related to factuality, such as epistemic modality or evidentiality, has been annotated in several corpora. This level of information has received much attention from NLP research on the domain of biomedical texts, which resulted in corpora such as BioScope (Vincze et al., 2008). Within the domain of news texts, there are other annotation studies that include factuality-related information as well. Rubin et al. (2006) and Rubin (2007) created a corpus of news reports and editorials annotated with explicit certainty and doubt markers in epistemically modalized statements. The MPQA Corpus (Wiebe et al., 2005), the Penn Discourse TreeBank (Miltsakaki et al., 2004) and TimeBank (Pustejovsky et al., 2003) contain some information that is related to factuality, but either factuality was not the focus of the annotation, or only possible factuality markers were annotated (and not the resulting interpretation). The corpus that comes closest to identifying the information in a way that we are looking for is FactBank (Saurí and Pustejovsky, 2009). Our annotation design for factuality is therefore based on the FactBank's annotation scheme. We will further discuss Factbank in Section 2.1.

Several attempts have been made to annotate opinions, emotions and sentiment in texts. The corpus created by Bethard et al. (2004) contains annotations of propositional opinions (i.e. sentential complements which contain the actual opinion for many verbs) and their sources. It was used to explore the automatic extraction of these opinion propositions. Another well-known corpus annotated for opinion expressions is the MPQA (Multi-Perspective Question Answering) Opinion Corpus (Wiebe et al., 2005). This corpus has been used in various studies for the automatic detection of opinion expressions (Breck et al., 2007; Wilson et al., 2004) and opinion holders (Choi et al., 2005; Kim and Hovy, 2005). We used this corpus to add information about factuality on top of the opinion annotations. The design of the MPQA Corpus is discussed in Section 2.2.

2.1. FactBank

FactBank (Saurí and Pustejovsky, 2009) is a corpus built on top of TimeBank (Pustejovsky et al., 2003) annotated with information concerning the factuality of events. Saurí and Pustejovsky (2009) define event factuality as the level of information expressing the commitment of relevant sources towards the factual nature of events mentioned in discourse. In their descriptive framework, event factuality is described by means of a double-axis scale as represented in Figure 1. The horizontal axis represents the binary distinction between positive versus negative polarity, and the vertical axis represents the continuous scale of certainty. According to this definition, events are either depicted as facts or counterfacts with absolute certainty, or, when there is a certain degree of uncertainty involved, as possibly factual or possibly counterfactual events.

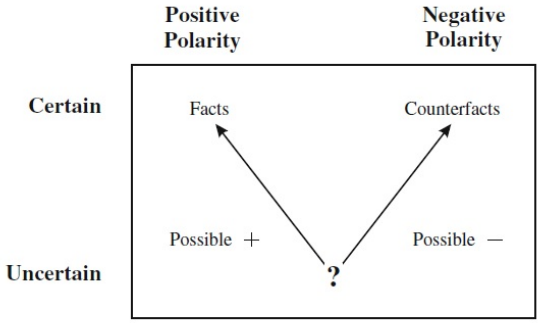

Figure 1: Double-axis representation of event factuality by Saurí and Pustejovsky (2009, p. 232)

Saurí and Pustejovsky (2009) translate the double-axis scale into discrete values for the annotation in the set of factuality values as represented in Table 1. The **CTu** value is used for cases where the source is certain about the fac-

	Positive (+)	Negative (-)	Underspecified (u)
Certain (CT)	Fact: CT+	Counterfact: CT-	Certain but unknown output: CTu
Probable (PR)	Probable: PR+	Not probable: PR-	(NA)
Possible (PS)	Possible: PS+	Not certain: PS-	(NA)
Underspecified (U)	(NA)	(NA)	Unknown or uncomitted: Uu

Table 1: The factuality values in FactBank (Saurí and Pustejovsky, 2009, p. 246)

tual nature of the event, but it is not clear what the output is (e.g. *John knows whether Mary came*). The **Uu** value is used when the source does not know what the factual status of the event is or does not overtly commit to it (e.g. *John does not know whether Mary came*).

Furthermore, the factual status that can be assigned to an event is always considered to be relative to at least one source that commits to it. The default source of an event mentioned in a text is the author of the text. However, the author can also introduce an additional source into the discourse by means of source introducing predicates (SIPs). SIPs select an argument denoting an event of some sort and contribute a new source to the discourse that plays a role in assessing the factuality of the embedded event. The SIP indicates to what extent the source commits to the factuality of the event. For example, in (3) the author introduces *Zimbabwe's President Robert Mugabe* as an additional source by means of the SIP *denied*, indicating that the event *rigging* is a counterfact according to Mugabe.

3. *Zimbabwe's President Robert Mugabe* on Monday denied$_{(s)}$ rigging$_{(e)}$ in last month's elections.

According to Saurí and Pustejovsky (2009), SIPs can be categorized into classes such as predicates of report (e.g. *say, tell, add*), predicates of belief and opinion (e.g. *think, predict, suggest*) or predicates expressing some psychological reaction as a result of an event (e.g. *regret, be glad/pleased (that), like (that)*).

Finally, Saurí and Pustejovsky (2009) appeal to the notion of *nested sources*. That is, in case of an SIP we can not be certain of the actual commitment of the additional source towards the event; all we know is what the author tells us. Thus in sentence (3) the rigging is taken as a counterfact by Mugabe according to the author, but the author stays uncommitted to the factuality of the rigging event (**Uu**). The chain of relevant sources for this event is represented as *mugabe_author*. Since a SIP can also be embedded in another SIP, it is possible to have multiple nesting levels.

2.2. MPQA Opinion Corpus

The MPQA Corpus (Wiebe et al., 2005) is a rich corpus containing news articles from a wide variety of news sources manually annotated for opinions and other *private states*, such as beliefs, emotions, sentiments and speculations. A private state is described as the state of an experiencer, holding an attitude, optionally toward a target. Three types of private state expressions are distinguished: explicit mentions of private states, speech events expressing private states, and expressive subjective elements (i.e. expressions that indirectly express private states through the way something is described or through a particular wording). For example, in (4) the experiencer is *Morgan Tsvangirai*, the (negative) attitude expressed by an explicit mention of a private state is *rejected*, and the target is *the outcome of the presidential poll*. Sentence (5) is an example of a private state expressed by the subjective speech event *said* and the expressive subjective element *"illegimate"*, expressing a negative attitude of *he* towards *the result*.

4. *Morgan Tsvangirai* has **rejected** [the outcome of the presidential poll].

5. *He* said [the result] was **"illegimate"**.

Similar to Saurí and Pustejovsky (2009), Wiebe et al. (2005) assume that there is a natural nesting of sources. However, Wiebe et al. (2005) only consider the 'main source' of the attitude. In (4), *Morgan Tsvangirai*) is relevant for evaluating the target; previous sources (like the author) are not considered relevant, except when there is an explicit disagreement between the sources. It should be noted that explicit mentions of private states and (subjective) speech events correspond to SIPs in many cases, because they can also convey the commitment of the source towards the factuality of an event.

The annotations in version 1.2 of the MPQA corpus were divided into four frames (Wiebe et al., 2005, p. 169-172):

Direct subjective frame representing subjective speech events or explicit mentions of private states

Expressive subjective element frame representing expressive subjective elements

Objective speech event frame representing speech events that do not express private states

Agent frame representing the source of the attitude or speech event

For version 2.0 of the MPQA Corpus, Wilson (2008) extended the annotation scheme by adding two more frames (annotated in around 73% of the articles in this version of the corpus) to better model attitudes and their targets (Wilson, 2008, p. 118-121):

Target frame representing the target of the attitude or speech event

Attitude frame representing the attitude specified for intensity (*low, low-medium, medium, medium-high, high, high-extreme*) and attitude type

Wilson (2008) also defined a new set of attitude types for these frames, which is represented in Table 2.

Sentiment	Agreement
Positive Sentiment	Positive Agreement
Negative Sentiment	Negative Agreement
Arguing	**Intention**
Positive Arguing	Positive Intention
Negative Arguing	Negative Intention
Speculation	**Other Attitude**

Table 2: The set of attitude types in MPQA (Wilson, 2008, p. 116)

Some of these attitude types are related to factuality information, such as speculation and arguing. However, the factual degree of *events* is not the focus of the annotation. Because we want to add more explicit information on event factuality and investigate how this integrates with positive and negative evaluations, we are only considering the sentiment related annotation of MPQA.

3. Theoretical framework

There are a few differences between our conceptualization and that of Saurí and Pustejovsky (2009) with regard to event factuality. According to Saurí and Pustejovsky (2009), events that are presented with absolute certainty are, depending on the polarity, either *facts* or *counterfacts*. Events that are presented with a degree of uncertainty on the other hand are either *possibly factual* or *possibly counterfactual*. However, we think that there is one drawback to this definition of factuality for our annotations in the MPQA Corpus. Consider the following sentences:

6. The *meeting*$_{(e)}$ was held yesterday in the Congressman's Washington office.

7. The *meeting*$_{(e)}$ will be held tomorrow in the Congressman's Washington office.

For Saurí and Pustejovsky (2009) the only information that is relevant for assessing the factuality of an event is whether the source commits to the event as a fact or not, irrespective of the temporality of the event. As a consequence, the event *meeting* in sentences (6) and (7) is in both cases equally depicted as a fact (**CT+**). But since any future event does imply some degree of uncertainty, even when the source presents it with absolute certainty (the meeting could be cancelled for some unexpected reason), we believe that there should be made a distinction between past or present events on the one hand and future events on the other hand.

Saurí and Pustejovsky (2009) have a good reason to ignore this factor in the definition of factuality; FactBank is built on top of TimeBank which already contains information on temporality. Unfortunately, we do not have this advantage with the MPQA Corpus. Therefore, we decided to include temporality in our definition as a third important aspect playing a role when determining the factuality of an event, next to polarity and certainty. By doing so, we do not ignore the central role of the stance of the source towards the event, but rather we assume that the source is aware that every future event does not yet correspond to the actual state of the world at the moment of speaking.

When incorporating this aspect into the definition of factuality, a different classification of events is realized:

- **Fact:** corresponds to the actual world
  polarity: YES
  certainty: CERTAIN
  temporality: PAST/PRESENT

- **Counterfact:** does not correspond to the actual world
  polarity: NO
  certainty: CERTAIN
  temporality: PAST/PRESENT

- **Possibility (uncertain):** could correspond to the actual world, but the source is not certain
  polarity: YES or NO
  certainty: UNCERTAIN
  temporality: PAST/PRESENT or FUTURE

- **Possibility (future):** could correspond to the actual world in the future
  polarity: YES or NO
  certainty: CERTAIN or UNCERTAIN
  temporality: FUTURE

We adopt the framework of Wilson (2008) for sentiment annotations, in which sentiment is divided into positive and negative sentiment and the intensity of the attitude can vary between low, low-medium, medium, medium-high, high, and high-extreme.

By combining factuality with sentiment, we can explain the way people perceive changes in the world (or more specifically, their beliefs of the world) as a function of a person's *fear* of changes to the bad or a person's *hope* of changes to the good. When talking about *changes*, we do not only refer to changes from the past or present to the future, but also to changes in our knowledge of events and belief about them. For example, (8) illustrates how hope can be held out for a change from a negative past/present (not finding the airplane) to a positive future (finding the airplane). However, the conceptual representation also applies to changes in our knowledge of the factuality of an event, as in (9). In this sentence, the past/present event of the airplane being hijacked is presented as a positive possibility (because then there is the chance that his/her loved ones are safe); the hope of the relative is that his or her knowledge about the factuality of this event changes from being merely a possibility to an indisputable fact.

8. *Rescuers* still hold out **hope**$_{(s)}$ [to find$_{(e)}$ Malaysia Airlines Flight 370]. *(source: www.wltx.com)*

9. One relative told NBC News:"*I* **hope**$_{(s)}$ [that the plane was hijacked$_{(e)}$]." *(source: www.nbcnews.com)*

4. Annotating factuality in MPQA

The aim of this research is to investigate whether sentiment and event factuality can be effectively integrated for the interpretation of perspectives of people expressed in news texts. We decided to use the MPQA Corpus (version 2.0), which already contains the sentiment annotations we want to use, and add a semantic layer that represents factuality

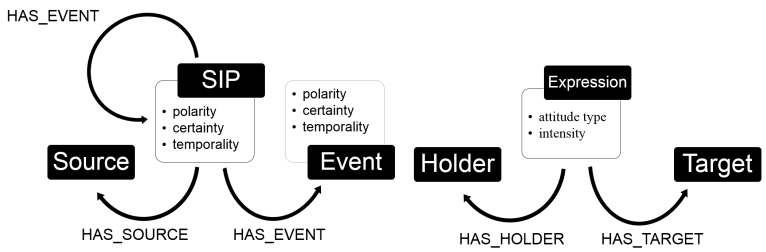

Figure 2: Annotation schema in CAT: factuality and sentiment

information. We selected a set of 31 articles (with a total of 17,710 words) from the MPQA Corpus that are all about the 2002 presidential election in Zimbabwe. In fact, there were 46 articles about this topic in the MPQA Corpus, but not all of them were annotated with the new attitude and target annotations defined in Wilson (2008). Because we needed the new attitude annotations to be able to focus only on *sentiment* and exclude all other attitude types (e.g. *arguing* or *speculation*), we decided to leave the other 15 articles out of the analysis. By selecting a set of articles about the same topic, we were able to concentrate on the aggregration of the perspectives of a limited set of people on a limited set of related events. In turn, this would give us some interesting material to analyze what exactly we could gain from the proposed approach.

We used the CELCT Annotation Tool (Lenzi et al., 2012) for our annotations, a general-purpose web-based tool for text annotation developed by the Center for the Evaluation of Language and Communication Technologies (CELCT). Users of CAT can define markables, attributes and relations between markables. We followed the following steps for the annotation task.

We first simplified the opinion annotations of the MPQA Corpus by translating the frames into opinion triples consisting of an opinion holder[1], an opinion expression and an opinion target. We then converted these opinion triples to the CAT format, which resulted in the markables, attributes and relations as represented in the right part of Figure 2.

The left part of Figure 2 shows the markables, attributes and relations we defined for the annotation of factuality. This annotation schema is similar to the one used for the annotation of FactBank in the sense that it identifies three basic elements: the (SIP-embedded) event, the SIP and the source that is introduced by the SIP. However, although we do believe that the notion of nested sources is relevant for the factual assessment of events, we only evaluated the factuality of the event according to the *main source* for the time being (i.e. only according to the source introduced by the SIP in which the event is embedded). We also use different attributes from the set of factuality values used for events and SIPs in FactBank:

- *polarity:* YES, NO, OTHER[2]
- *certainty:* CERTAIN, UNCERTAIN, OTHER
- *temporality:* PAST/PRESENT, FUTURE, OTHER
- *special cases:* general statements, conditionals

We not only added the attribute *temporality*, but also merged the PROBABLE and POSSIBLE values of the *certainty* attribute into the value UNCERTAIN. The value OTHER can be used for all attributes when the other values do not apply or the annotator is not certain about the value that should be assigned. Furthermore, we identify two types of special cases: general statements and conditionals. Sentence (10 is an example that contains an event that is part of a general statement, namely *elections*. These *elections* do not refer to a specific event than can be related to a specific time and place, but rather to the concept of a general event that occurs more often. Sentence (11) is an example of a conditional, in which sources do not refer to the factuality of an event in the actual world, but to the factuality of an event in a hypothetical world. Both X and Y in a conditional construction "if X, then Y" are marked as conditional events.

10. "...the electoral process$_{(e)}$ could not be said$_{(s)}$ to adequately comply$_{(e)}$ with the norms and standards for elections$_{(c)}$ in the SADC region," the forum said$_{(s)}$.

11. "If **sanctions**$_{(g)}$ are **imposed**$_{(g)}$ on Zimbabwe, it will not **survive**$_{(g)}$," he said$_{(s)}$.

5. Results and evaluation

In the following sections, we first report on the interannotation agreement for assessing the clarity of the annotation design for factuality. Second, the data distributions are discussed, which gives insight into the adequacy of the factuality classes. Finally, we discuss the added value of integrating factuality and sentiment found in our data.

5.1. Interannotation agreement

Most of the annotations were done by one person, but to gain insight into the clarity of the annotation design, 45 sentences were doubly annotated by a second annotator and another 26 sentences were doubly annotated by a third annotator. The annotation task can be divided into three subtasks, namely (1) the identification of the markables, (2) the annotation of the arguments of events and SIPs, and (3) the annotation of the relations between the markables. The agreements on the markables and their attributes are

[1] We use the term *holder* instead of *source* to avoid confusion between the source of an attitude and the source of an SIP.

[2] We avoid using the terms of *positive* and *negative* polarity to avoid confusion with positive and negative sentiment.

	Identification	Polarity	Certainty	Temporality
Source	0.74 / 0.52	-	-	-
Event	0.85 / 0.52	0.94 / 0.82	0.83 / 0.67	0.88 / 0.85
SIP	0.77 / 0.43	0.93 / 0.88	0.89 / 0.81	0.91 / 0.88

Table 3: The interannotation agreement scores for the markables and their attributes

evaluated with the Dice's coefficient. Table 3 presents the agreement on subtask (1) and (2) between annotator 1 and annotator 2, and between annotator 1 and annotator 3.

What is notable from these scores is that there is a clear difference in the agreement between annotator 1 and 2 in subtask (1) as opposed to the agreement between annotator 1 and 3. The latter is quite low. An analysis of the disagreement instances showed that annotator 1 took the overall concept of both events and SIPs broader than annotator 3 did, and that there was confusion in whether an event was a SIP or a 'normal' event. The disagreement in classifying events as SIP directly caused disagreement in the identification of sources as well, as sources are only identified if the SIP is identified. Sentence (12) is an example of a sentence that illustrates these disagreements. We think that there are many cases that are difficult to decide on because several interpretations are possible. In order to improve the clarity of the annotation design, the guidelines we are working on are extended with more explicit examples of the variations encountered in real texts.

12. **Annotator 1**: *Mr. Tsvangirai* might face$_{(s)}$ charges$_{(s)}$ of contravening$_{(e)}$ the Law and Order Maintenance Act for allegedly$_{(s)}$ threatening$_{(s)}$ to violently remove$_{(e)}$ President Mugabe from power$_{(e)}$ last year.

    Annotator 3: Mr. Tsvangirai might face$_{(e)}$ charges of contravening$_{(e)}$ the Law and Order Maintenance Act for allegedly threatening$_{(e)}$ to violently remove$_{(e)}$ President Mugabe from power last year.

With regard to subtask (2), the overall inter annotator agreement is relatively high. The disagreements that were found were all cases that were simply more complicated and needed an explicit consensus. We would expect that this could be solved by discussing such examples and developing guidelines that help to make clear decisions for difficult cases - and by a sufficient amount of practice.

5.2. Distributions

The majority of the 3,547 events and SIPs are classified as facts (2,481/70.0%), whereas only a small amount consisted of counterfacts (173/4.9%). Possibilities, including both the 'uncertain' and the 'future' possibilities, made up 16.8% (595 instances) of the events and SIPs. The remainder (289/8.4%) is a mixed group of conditionals, general statements, and the events and SIPs that were assigned one or more OTHER values.

A similar pattern can be seen in the distributions of the factuality values in FactBank, which also show a high frequency of the **CT+** value (58.1%) and an even lower frequency of the **CT-** value than we found (2.3%) (Saurí and Pustejovsky, 2009). The frequency of the possibilities found in our data, on the other hand, is much higher than the one found in FactBank (2.8%). This reflects a difference in the definitions of factuality. In the definition of Saurí and Pustejovsky (2009), the temporality of an event plays no role in the assesment of its factuality. As a consequence, future events can be facts as well as counterfacts. In our definition on the other hand, future events are inherently understood as possibilities, which results in a higher frequency of this class.

Nevertheless, our frequencies of facts and counterfacts are still higher than in FactBank. Indeed, one would expect them to be lower because some of the facts and counterfacts in FactBank are considered possibilities in our definition. In other words: why are *all* of the frequencies of facts, counterfacts and possibilities in FactBank lower than the ones we found? The reason most likely lies in the difference in the annotation designs: whereas we ignored the notion of nested sources in this annotation study, in FactBank each event is assigned the factuality values according to all relevant sources. Saurí and Pustejovsky (2009) explain how this leads to a very high frequency of the factuality value **Uu** (and consequently, a lower frequency of the other classes): in many cases that involve embedded events, the author remains uncomitted to the factuality of the event.

5.3. The gains of combining factuality and sentiment

As mentioned in Section 4., we chose to annotate a set of articles that are all about the 2002 elections in Zimbabwe (most of them are written after the elections). A set of articles around the same topic allows us to do a meaningful analysis on a limited set of people and events. We found that the combination of event factuality and sentiment provides us with some interesting insights. When analyzing the annotations, it is important to keep in mind that the layer of factuality is present in two parts of the opinions: in the opinion target, or in the opinion expression. For example, the event *election* in both (13) and (14) is part of the opinion target, whereas *condemned* in (13), and *free* and *fair* in (14) are part of the opinion expressions. We analyzed these two patterns separately, because their outcomes should be interpreted in different ways.

13. [Zimbabwe's presidential election$_{(e)}$] was **condemned**$_{(s)}$ by *some Western countries*.

14. *The Tanzanian and Democratic Republic of Congo observer teams* [...] said$_{(s)}$ [the election]$_{(e)}$ had been **free**$_{(e)}$ and **fair**$_{(e)}$.

The relevance of factuality information in opinions is clearest when the event is part of the opinion target. To gain knowledge about the different perspectives on the events around the elections, we splitted the opinions about events into positive and negative opinions about factual events, counterfactual events, future events and uncertain

	Facts	Counterfacts	Possibilities	Other	Total
Event	1,514	103	450	226	2,293
SIP	967	70	145	72	1,254
Total	2,481	173	595	298	3,547

Table 4: The distribution of the factuality classes

events. Not surprisingly, the (conduct of the) election itself was the factual event that was most frequently mentioned as a target of the opinions expressed by the different groups. This event was referred to by words as *election(s)*, *poll*, *vote*, and *conduct*. There was a very clear split in the perspectives of different people on this event: all the people that were on Mugabe's side were positive about the event, whereas all the people that were on Tsvangirai's side were negative about the event. The same groups could be seen in the perspectives on the event of Mugabe winning the elections, referred to with words as *victory*, *reelection* and *win/won*.

With regard to future events, which are considerably less frequent in the corpus, the pro-Tsvangirai perspectives are much more represented than the pro-Mugabe perspectives. This group expresses positive views on possible future events such as *suspension*, *sanctions*, *action (on the situation)*, *refrain from violence* and *calm*, but negative views on events such as *violence*, *(reckless) action* and *confrontation (with the state)*. In other words: they want to make a change, but in a peaceful manner. On the contrary, Mugabe is making clear that he is already moving forward focusing on people's *acceptance* of the results and the hope of *stabilizing* the economy.

The factuality of events that were part of the opinion expressions should be interpreted in another way. In the previous cases, there was no disagreement in the factuality of the target event between the groups. That is, they both agreed that it is a fact that Mugabe won the election, and they made no effort to prove otherwise. However, our analysis revealed that with events that are part of opinion expressions, there is a difference in the assessment of the factuality reflecting different perspectives. For example, opponents of Mugabe argued that it was a fact that the election was *rigged*, *flawed*, *illegitimate* and marred by *violence*, *intimidation* and *manipulation*. On the other hand, they denied that the election was *free*, *fair*, *peaceful* and *transparant*, and that it *met international standards*. Mugabe and his allies disagreed on both points.

	pro-Mugabe		pro-Tsvangirai	
	Fact?	Positive?	Fact?	Positive?
election (T)	Yes	Yes	Yes	No
rigged (E)	No	No	Yes	No

Table 5: Contrasting perspectives

What we learn from this is that people can express their perspectives on events in two ways in newspapers. If the event is part of the target within the opinion, the person positively or negatively evaluates the event without arguing about the factuality of it. Contrasting perspectives will disagree about the evaluation of the event. If the event is part of the expression within the opinion, the person is arguing about the factuality of the event, not about the positiveness or negativeness of it. Contrasting perspectives will disagree about the factuality. This difference in the sentimental evaluation and factuality assessment of target events and expression events is illustrated in Table 5.

However, there were also cases where the annotation of factuality on top of the sentiment annotation led to wrong interpretations. In sentence (15) for example, a positive attitude *hope* of *Mugabe's allies and detractors* is annotated towards the part between brackets as a target. The events and SIPs annotated in this opinion are underlined. The events *situation*, *fears* and *concerns* are problematic: these are all annotated as facts, but obviously the holder of this opinion is not positive towards these events being facts. This problem is caused by the fact that the annotation design of MPQA has not been designed to be integrated with event factuality from the very beginning, and we have not yet accounted for cases like this in our current annotation design.

15. *Mugabe's allies and detractors* may all **hope**$_{(s)}$ [that he is spending$_{(e)}$ the time reviewing$_{(e)}$ the tense situation$_{(e)}$ in the country and working out$_{(e)}$ how best to calm$_{(e)}$ fears$_{(e)}$ and concerns$_{(e)}$ and rally$_{(e)}$ Zimbabweans together].

6. Discussion and future work

We have suggested to combine the dimensions of factuality and opinion for the interpretation of perspectives on events. Although our findings suggest that this approach could prove fruitful for modeling perspectives in news, it also appeared that not all perspectives in MPQA were interpreted correctly. This was caused by the fact that the separate models for factuality and sentiment have not been constructed to be integrated from the very beginning.

Furthermore, the annotation of event factuality is in itself still an open issue. Although we do believe that polarity, certainty and temporality are the three main aspects for the assessment of factuality, there are still cases in which it is difficult to decide what value should be assigned. Further examination of problematic cases is needed to provide better guidance in the annotation process.

Especially the aspect of temporality needs some further consideration, because of the difficulty that is inherent to the concept. For example, we can talk about the present as the future when talking about our beliefs from the past (*as a kid I believed that...*), and compare those beliefs with the beliefs we hold at the present (*but now I know that...*). Therefore, not only events should be place on a timeline, but the perspectives we have on them should be placed on a timeline as well.

It would also be worthwile to carry out a deeper analysis on the concept of events itself. In this study, we made no distinction between *events* and *states*. Also, we ignored the words that do imply an event but do not refer to the event itself, such as entities that are the agent of an event (*winner*) or words that refer to the end or result of an event (*results, outcome*). It would be interesting to see what can be done with this kind of variation in the representation of events.
In our ongoing work, we are investigating more refined annotations of perspectives, where event factuality and sentiment are effectively integrated. The aforementioned issues will be considered in the development of these annotations. The ultimate goal is to create high performance factuality and opinion classifiers which we will apply to large daily news streams to model perspectives in the news. Our dataset is available through `http://www.newsreader-project.eu/results/data/`.

Acknowledgments

This research was funded by the European Union's 7th Framework Programme via the NewsReader (ICT-316404) and OpeNER (ICT-296451) projects.

References

Bethard, Steven, Yu, Hong, Thornton, Ashley, Hatzivassiloglou, Vasileios, and Jurafsky, Dan. (2004). Automatic extraction of opinion propositions and their holders. In *AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications*.

Breck, Eric, Choi, Yejin, and Cardie, Claire. (2007). Identifying expressions of opinion in context. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-2007)*.

Choi, Yejin, Cardie, Claire, Riloff, Ellen, and Patwardhan, Siddharth. (2005). Identifying sources of opinions with conditional random fields and extraction paterns. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT/EMNLP)*, Vancouver, Canada.

Kim, Soo-Min and Hovy, Eduard. (2005). Identifying opinion holders for question answering in opinion texts. In *Proceedings of AAAI-05 Workshop on Question Answering in Restricted Domains*, Pittsburgh, PA, USA.

Lenzi, Valentina Bartalesi, Moretti, Giovanni, and Sprugnoli, Rachele. (2012). CAT: the CELCT Annotation Tool. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, May 23-25.

Miltsakaki, Eleni, Prasad, Rashmi, Joshi, Aravind, and Webber, Bonnie. (2004). The Penn Discourse TreeBank. In *Proceedings of the Language Resources and Evaluation Conference (LREC'04)*, Lisbon, Portugal.

Palmer, Martha, Gildea, Dan, and Kingsbury, Paul. (2005). The Proposition Bank: A corpus annotated with semantic roles. *Computational Linguistics*, 31(1).

Pang, Bo and Lee, Lillian. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

Pustejovsky, James, Hanks, Patrick, See, Roser Saurí Andrew, Gaizauskas, Robert, Setzer, Andrea, Radev, Dragomir, Sundheim, Beth, Day, David, Ferro, Lisa, and Lazo, Marcia. (2003). The TIMEBANK corpus. In *Proceedings of Corpus Linguistics 2003*, pages 647–656.

Rubin, Victoria, Liddy, Elizabeth, and Kando, Noriko. (2006). Certainty identification in texts: Categorization model and manual tagging results. In *Computing attitude and affect in text: Theory and applications*, pages 61–76. Springer.

Rubin, Victoria. (2007). Stating with certainty or stating with doubt: Intercoder reliability results for manual annotation of epistemically modalized statements. In *Proceedings of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*, pages 141–144.

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

Vincze, Veronika, Szarvas, György, Farkas, Richárd, Móra, György, and Csirik, János. (2008). The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(S9).

Wiebe, Janyce, Wilson, Theresa, and Cardie, Claire. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165–210.

Wilson, Theresa, Wiebe, Janyce, and Hwa, Rebecca. (2004). Just how mad are you? Finding strong and weak opinion clauses. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence*, San Jose, California, USA, July 25-29.

Wilson, Theresa Ann. (2008). Representing attitudes and targets. In *Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States*. University of Pittsburgh.