A Perspectives Model Definition and Annotation

Version 0.1

(Work in progress)

VUA Perspectives Group

November 11, 2015

Contents

Chapter 1

Introduction

With the Internet having secured an increasingly prominent place in society, the current age is characterized by the ability of individuals to transfer and access information about the world. At the same time, people feel more and more encouraged to voice their opinion and debate with others about all kinds of topics on review sites, personal blogs, discussion boards and other social media platforms. In addition, even -or especially- news articles, which very function is to inform us about the world, inherently express conflict, resistance, engagement and political stance.[1] They reflect on-going debates in our society, stances on particular issues (e.g. abortion, vaccinations, etc.), and interpretative frames on events and their causes (e.g. conspiracy theories on 9/11). As such, the Internet is evolving into a rapidly increasing collection of unstructured data representing not only what is happening in the world, but also what people think about it.

In response, researchers in the fields of natural language processing (NLP), computational linguistics and information extraction (IE) have been developing text mining systems that are able to extract relevant pieces of information from natural language text, to filter the information for opinions, and possibly to classify them into different types of attitude and polarity. This field of research has been generally referred to as subjectivity analysis, sentiment analysis or opinion mining. It has, for example, been applied for automatic summarization of customer feedback in online product and service reviews. However, the complexity of the task greatly increases when the topic of interest is, for example, a contentious event or issue relating to politics and society (e.g. the Ferguson shooting, the European migrant crisis or health care reform). It is no surprise that the question "How do people feel about the European migrant crisis?" would lead to much more complicated answers than the question "How do people feel about the latest model of the iPhone?". The complexity of the task is inherent to the complexity of contentious text itself: each contentious event or issue involves a variety of (groups of) participants, and the arguments they put forward can be very complex and touch a range of distinct aspects of the debate (e.g. the insurance's costs in the health care reform debate). To understand one's opinion about a contentious event or issue is therefore to understand their *perspective*. That is, it is necessary to know which individuals or groups are involved in the debate and what they think is true, right and important.

Dealing with textual data, the only way to approximate an understanding of a person's perspectives is to analyze *what* is said and *how* this is said. The selection of information plays an important role because it reflects what someone considers relevant. For example, an article may report on someone's ethnicity, or experts' political preferences when citing them on a societal matter. Therefore, the assumption is that *all* information presented in text is fundamentally subjective. The way in which this information then is presented (i.e. how sources choose to picture something) also indicates the sources' point of view. Their

[1] "*militia est vita hominis super terram*" Book of Job, 7,1. http://www.wumingfoundation.com/giap/?p=22200

judgment with regard to what is true, right or important may be expressed in an explicit way (e.g. *I'm against vaccinations*) but also in a more implicit way through choice of words (e.g. calling someone a thief instead of saying he stole something).

This document describes an annotation scheme that can be used to analyze perspectives through what is said and how it is said by linking statements to their source while incorporating any additional information on the statement. It includes different aspects of linguistic encoding of perspectives, which in the last decade have been targeted as separated phenomena through series of different annotation initiatives. Among these, the most relevant to our work are:

- **Penn Discourse TreeBank (PDTB) v2.0 (Prasad et al., 2007) and its extension, the Penn Attribution Relation Corpus (PARC) v3.0 (Pareti, 2015)**: News texts annotated with *Attribution Relations (ARs)*, a linguistic phenomenon that is used to ascribe to someone the ownership of an attitude expressed in a text;
- **FactBank (Saurí and Pustejovsky, 2009)**: News texts annotated with *event factuality* on top of event annotations in TimeBank (Pustejovsky et al., 2006), which is defined as "the level of information expressing the commitment of relevant sources towards the factual nature of events mentioned in discourse" (Saurí and Pustejovsky, 2009, p. 231);
- **The MPQA Opinion Corpus (Wiebe et al., 2005):** News texts annotated with different types of opinions or *private states*, including sentiment, speculation, arguing and agreement;
- **The Language Understanding (LU) Corpus (Diab et al., 2009) and its extension in the DARPA DEFT project (Prabhakaran et al., 2015)**: News texts, training manuals, correspondences, interviews and discussion forum threads annotated with the *Level of Committed Belief*, which definition corresponds to that of event factuality.

The annotations described in this document are compliant with these annotation initiatives.

The remainder of this document is structured as follows. Chapter 2 gives a short overview of related work that has been done in the annotation initiatives mentioned above. Chapter 3 explains our basic approach; it clarifies why we make a distinction between *instances* (representing unique things in a possible world) and *mentions* (tokens referring to instances), describes the three main elements of perspectives (source, cue and target), justifies our multi-layered annotation approach, and introduces the notion of perspective scope. The remaining chapters are each devoted to the annotation guidelines for the separate perspective layers. So far, we have included the layers of event detection, attribution, belief/factuality, and sentiment/opinion. In the future, we might add additional layers such as emotion.

Chapter 2

Related work

2.1 Penn Discourse TreeBank and Penn Attribution Relation Corpus

The Penn Discourse TreeBank (PDTB) (Prasad et al., 2006) and the Penn Attribution Relation Corpus (Pareti, 2012) provide annotation for perspectivazation of information to distinguish between factual and non-factual information, and to identify the source of the information. Main distinctions with respect to previous approaches concern: a.) the definition of the attribution phenomenon which encompasses approaches based on sentiment analysis; b.) an extension of the annotation which goes beyond the sentence boundary, as in the MPQA schema, with a focus on more analytic semantic units, i.e. Abstract Objects (AOs).
Both PDTB and PARC define an attribution relation as a relation ascribing the ownership of an attitude towards some linguistic material, thus suggesting that an attribution relations will always provide access to subjective information.

Any attribution relation is composed by three elements:

- the source: it signals WHO is providing the information content to a receiver;
- a cue: it is a lexical items which signals the presence of an attribution relation. Cue words can be further exploited to classify the different levels of factuality of an attribution relations and possibly to distinguish different sources; and
- the content: it corresponds to the span of text attributed to the source (i.e. information content, or WHAT the source is stating/thinking/believing/promising etc.)

The PARC annotation scheme is an extension of the PDTB scheme for attribution relations annotation. In particular, in the PDTB attribution relations are annotated only when overlapping with a relation conveyed by a discourse connective. On the contrary, the PARC annotation scheme extend the lexicalized approach proposed in for the annotation of Discourse Relations in the PDTB (D-LTAG), assume that the annotation is independent from other discourse relations, and any kind of attributed material is considered.

Provided that the PARC scheme is an extension of the PDTB one, we will provide a description of the former as it is more relevant to our work. The PARC scheme inherits the three markables for cue, source and content from the PDTB and introduces a new markable: supplement. The supplement is used annotate additional information of an attribution relation. As reported in the PARC 3.0 Annotation

guidelines, a supplements are elements "concurring to the identification of the source and the provenance or mean by which the information was acquired (e.g. said on the phone); providing further specification of the attitude this holds (e.g. said with anger); the recipient of a reportive verb of the assertion type (e.g. told the jury) or of an eventuality (e.g. Mary expects John to do the shopping); and event specifications (e.g. said last week) providing context indications determinant to the interpretation and comprehension of the content". (Pareti, 2015, 8). Each attribution relation is composed by at least one content markable, one cue, and none or more source and supplement markables.

The PARC annotation guidelines inherits and extends attributed for the markables from the PDTB such as type of relations, type of source, scopal change, and factuality, they are not present in the released version of the corpus (PARC 3.0) due to low inter-annotator agreement.

The annotation of attribution relations following the PARC 3.0 guidelines consists of five subtasks: a.) cue identification; b.) source identification and c.) content identification; d.) supplement identification; and e.) relation identification, i.e. connecting the markables which compose an attribution relation.

2.2 TimeBank/FactBank

FactBank (Saurí and Pustejovsky, 2009) is a corpus in which *factuality* has been annotated on top of event annotations in TimeBank (Pustejovsky et al., 2006). Event factuality is defined as "the level of information expressing the commitment of relevant sources towards the factual nature of events mentioned in discourse" (Saurí and Pustejovsky, 2009, p. 231), where an *event* as specified in the TimeML annotation scheme refers to any situation that happens or occurs and may be expressed by means of a tensed or untensed verb, nominalization, adjective, predicative clause, or prepositional phrase.

Event factuality is modeled as the combination of *polarity* and *epistemic modality*. Polarity is a binary system with the values POSITIVE and NEGATIVE, referring to whether or not the event is negated by means of, for example, a negative polarity marker (e.g. *not*, *none*, *nobody*). Epistemic modality (or *certainty*) expresses the degree of certainty of a source regarding what is asserted. It is conveyed by, for example, modality particles (e.g. *could*, *probably*, *impossible*). In contrast to polarity, epistemic modality logically constitutes a continuum ranging from uncertain to absolutely certain, but FactBank distinguishes among three categories: POSSIBLE, PROBABLE and CERTAIN. In addition, FactBank adds UNDERSPECIFIED values on both axes to account for cases in which the source is fully or partially uncommitted to the factual nature of the event. This results in a total of eight factuality values:

Committed Values:

CT+ According to the source, it is *certainly* the case that X.
PR+ According to the source, it is *probably* the case that X.
PS+ According to the source, it is *possibly* the case that X.
CT- According to the source, it is *certainly not* the case that X.
PR- According to the source it is *probably not* the case that X.
PS- According to the source it is *possibly not* the case that X.

(Partially) Uncommitted Values:

CTu The source knows whether it is the case that X or that not X.
Uu The source does not know what is the factual status of the event, or does not commit to it.

The factuality values are always relative to at least one relevant source. By default, the author of a text is taken to be a relevant source for each of the events mentioned in the text. In addition, the author can introduce an additional source as a factuality evaluator by means of *source-introducing predicates*

(SIPs). SIPs select an argument denoting an event of some sort, contribute a new source to the discourse that plays a role in assessing the factuality of the embedded event, and indicate to what extent the source commits to its factuality. Examples of SIPs are predicates of report (*say*, *claim*, *argue*), predicates of knowledge (*know*, *remember*, *admit*) and predicates of perception (*see*, *feel*, *hear*). If an event is embedded in an SIP, separate factuality values are assigned to each of the relevant sources. For example, the event *reported* in sentence 1 is annotated with a factuality value for the author only, while *contain* is annotated with two separate factuality values (one for the author, and one for *CBS News*):

1. CBS News first **reported**$_{e1}$ last night that the tomb may **contain**$_{e2}$ the remains of Air Force pilot Michael Blassie.

Relevant sources:	author, news_author
Factual assignments:	f(e1,author) = CT+
	f(e2,news_author) = PS+
	f(e2,author) = Uu

To help annotators mentally structure and comprehend the different information layers involved, the annotation task for creating FactBank was divided into three consecutive subtasks: (a) the identification of SIPs, (b) the identification of sources, and (c) the assignment of factuality values.

2.3 MPQA Opinion Corpus

The MPQA Opinion Corpus (Wiebe et al., 2005) is organized around the notion of *private states*, which in terms of their functional components are described as "(internal) states of *experiencers* holding *attitudes*, optionally towards *targets*" (Wiebe et al., 2005, p. 4). A distinction is made between DIRECT SUBJECTIVE FRAMES, which represent explicit mentions of private states as well as speech events expressing private states, and EXPRESSIVE SUBJECTIVE ELEMENT FRAMES, which represent so-called 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 sentence 2 there is a (negative attitude) of *Morgan Tsvangirai* expressed by *rejected*, which is an explicit mention of a private state. Sentence 3 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*. For this sentence, two private states would be created: a DIRECT SUBJECTIVE FRAME for *said* and an EXPRESSIVE SUBJECTIVE ELEMENT FRAME for *illegitimate*.

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

3. *He* **said** [the result] was **"illegitimate"**.

Other frames that have been defined in MPQA are the OBJECTIVE SPEECH EVENT FRAME to distinguish opinion-oriented material from material presented as 'factual' and the AGENT FRAME for representing the source of the attitude or speech event. An important property of sources in the annotation scheme is that they are *nested*. That is, one private state (or speech event) may be embedded in another. As such, the nested source of the private state expressed by *fears* in sentence 4 is represented as *<writer, Xirao-Nima, U.S.>* (Wiebe et al., 2005, p. 9). The notion of nested sources has been adopted in FactBank (Saurí and Pustejovsky, 2009) and will also be applied in future belief annotations of the DARPA DEFT project (Prabhakaran et al., 2015).

4. "The U.S. fears a spill-over," said Xirao-Nima.

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: the TARGET FRAME and the ATTITUDE FRAME (Wilson, 2008, p. 118-121). For the ATTITUDE FRAME, Wilson (2008) defined a new set of attitude types, which is represented in Table 2.1. The frame is further characterized by the intensity of the attitude type being expressed (*low, low-medium, medium, medium-high, high, high-extreme*) and by four additional properties that are used to mark particular characteristics of attitudes when they are relevant (*inferred, contrast, repetition, sarcasm*). TARGET FRAMES are linked to ATTITUDE FRAMES, which in turn are linked to DIRECT SUBJECTIVE FRAMES. Targets and attitudes are thus not annotated for EXPRESSIVE SUBJECTIVE ELEMENT FRAMES.

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

Table 2.1: The set of attitude types in MPQA 2.0 (Wilson, 2008, p. 116)

In MPQA 2.0, the annotations are entirely span-based. However, in a yet unreleased version of the corpus, MPQA 3.0, so-called *eTarget* annotations are being added to the MPQA 2.0 annotations in order to obtain a more accurate representation of *what* or *whom* an opinion is directed towards (Deng and Wiebe, 2015). An eTarget is an entity or event that is the target of an opinion; its annotation is anchored to the head word of the NP or VP that refers to the entity or event. For example, in sentence 5 the part between brackets is annotated as the target of the negative sentiment of *the Imam* in MPQA 2.0; in the new version, two eTargets are annotated within this target span: *Rushdie* himself and his act of *insulting*. This example also shows that there may be entities within a target span which the attitude is not directed towards, in this case, *the Prophet*.

5. When *the Imam* **issued the fatwa against** [Salman Rushdie for insulting the Prophet]...

2.4 Language Understanding Corpus and the DARPA DEFT project

The Language Understanding (LU) Corpus (Diab et al., 2009; Werner et al., 2015) aims at capturing the Level of Committed Belief (LCB), which is described as "a linguistic modality expressing a speaker or writer's (SW) level of commitment to a given proposition, which could be their own or a reported proposition." (Werner et al., 2015, p. 32). Initial work on LCB was undertaken by Diab et al. (2009), who used a 3-way distinction of belief tags to annotate the (original) LU-3 corpus: Committed Belief (CB), reflecting a strong belief of the SW towards the proposition, Non-committed Belief (NCB), reflecting a weak belief of the SW towards the proposition, and Non Attributable Belief (NA), reflecting another type of attitude of the SW towards the proposition (such as desire). However, the NCB category of the original LU tagging scheme in fact captured two different notions: that of uncertainty of the SW and that of belief being attributed to someone other than the SW. Recently, this category has been manually relabeled with the NCB tag (for uncertainty of the SW) and the Reported Belief tag (ROB), resulting in the LU-4 corpus (Werner et al., 2015).

Belief tags in LU-4 Corpus:	
Committed Belief (CB)	The writer strongly believes that the proposition is true.
Non-committed Belief (NCB)	The writer believes that the proposition is possibly or probably true, but is not certain.
Reported Belief (ROB)	The writer attributes belief to another person or group.
Non-belief propositions (NA)	The writer expresses some other cognitive attitude toward the proposition, such as desire or intention, or expressly states that s/he has no belief about the proposition.

Although the term *belief* in this framework essentially refers to the same phenomenon as *factuality* as defined by Saurí and Pustejovsky (2009), the annotations are quite different. In contrast to FactBank, which has fully annotated nested sources, the LU Corpus so far has only addressed the problem from the perspective of the speaker/writer. Furthermore, the LU corpus ignores negation and does not distinguish between possible and probable. Finally, there is a subtle difference in the targets of the belief/factuality annotations. Whereas FactBank has assigned factuality values to *events*, the LU corpus has assigned belief tags to the head words of *propositions*, leaving event-denoting noun phrases (e.g. *the **collapse** of the building*) out of consideration.

The 4-way distinction of belief tags as described above is currently being used in an ongoing annotation effort at the Linguistic Data Consortium (LDC) that has grown out of the DARPA DEFT project (Prabhakaran et al., 2015). Whereas the LU corpus covers different domains and genres (including news wire, blog data, email and letter correspondence, and transcribed dialogue data), this annotation project focuses on texts from discussion forum threads. While the current annotations only include beliefs of the writer, there are several plans to extend the annotations. These include: (1) the annotation of nested beliefs in a similar way as was done in FactBank, (2) the extension of the definition of target propositions by using semantic representations (as opposed to using the propositional head) and including the heads of noun phrases, (3) the identification of entities as targets of beliefs (referred to as the notion of belief "aboutness"), (4) combining belief with sentiment annotations, and (5) extending the annotations to Spanish and Chinese.

Chapter 3

Basic approach

3.1 Mentions vs. instances

A fundamental assumption in our representation is that we distinguish *instances* from *mentions*. Instances represent unique things in a possible world such as persons, locations, time expressions, events or concepts. Mentions are tokens that refer to these things.[1]

We use the Grounded Annotation Framework (Fokkens et al., 2013, GAF) to formally model the relations between instances and mentions. An instance can be `denotedBy` one or more references. When more than one mention refers to the same instance, these mentions corefer. As such, GAF provides a natural way to represent coreferential relations across entities and events. Because the instance representations exist completely independently from text (i.e. no anchor term is needed to represent a coreference chain), the approach is particularly suitable for cross-document coreference (Fokkens et al., 2013). An important aspect of coreference is the identity of things. For events, this may be complex since descriptions of events may be different while sources are still consider the same event.

In this document, we will not address coreference annotations. The annotations we describe cover information about sentiment analysis and opinions, factuality and sources of statements. In other words, we aim to annotate the perspective that a specific mention (or syntactic construction of mentions) displays on a specific entity. GAF and the instance representation may not be the main focus of this document, but they are fundamental to understand how our annotations ultimately contribute to a representation of perspectives. By investigating what sentiment, factuality and sources are associated with various mentions of an entity or an event, we can directly compare the different visions that are displayed about it in various texts.

3.2 Source, cue, target

Conceptually, we take perspectives to consist of three main elements: a *source* (an entity, such as a person or organization) having an *attitude* towards some *target*. The default source of a textual statement is its author (or the document in which in occurs, or its publisher), since all statements in a text can be attributed to the author. In many cases, this source will be implicitly present in the text, although it can be lexicalized through the first-person singular pronoun or expressions like *in **my** opinion*. Other possible

[1] Mentions can also be pictures, symbols, audio signals, etc. Only textual references are relevant for this document.

sources are specific entities that are introduced in the text (e.g. ***John*** *believes that...*) or arbitrary sources expressed by non-specific references in the text (e.g. *it is believed that...*).

The attitude of the source is expressed by some cue, which is the lexical anchor that links the source to the target. The most typical perspective cues are verbs that in an active sentence take the source as their subject and the target as their direct object. These include, for example, verbs of reporting (e.g. *say, write, claim*), verbs of cognition (e.g. *think, wish, forget*), and verbs of perception (e.g. *see, hear, feel*). However, there is a range of linguistic expressions that can function as a perspective cue. For example, uncertainty can be expressed by, among others, modal auxiliaries (e.g. *might*, *may*) and modal adverbs (e.g. *perhaps*, *maybe*). In turn, one cue can express a range of different attitudes. For example, a verb like *hope* expresses a positive sentiment and uncertainty towards what is stated at the same time.

The last element of a perspective is the target. In general, the semantic target can be either an (abstract) entity or an event. Entities are usually expressed as (proper) nouns or pronouns (e.g. *I don't like* ***John/dogs/it***), while events may be expressed as verbs (e.g. *I think he will* ***come***), nouns (e.g. *the* ***attack*** *was scary*), pronouns (e.g. ***it*** *was scary*), adjectives (e.g. he seemed ***angry***) and prepositional phrases (e.g. I think John is ***at home***). In addition, a perspective may target specific aspects of events or entities, which may be expressed as propositions. For example, in *I hope John was killed fast*, the positive sentiment is most likely directed towards the Arg-MNR (*fast*) of the killing event only. In this document, we use the term *perspective scope* to refer to those specific aspects of an event or entity that are the target of a perspective (see Section 3.4). Finally, not all types of attitudes can have the same types of targets; while sentiment (see Chapter 7) may target both events and entities, factuality (see Chapter 6) can only target events.

In sum, we annotate the following three elements in perspective relations:

- **Source:** The lexical elements that refer to the entity which the perspective is attributed to (optional).
- **Cue:** The lexical elements that signal the presence of a perspective (non-optional).
- **Target:** The lexical elements that refer to the specific entity or event that the perspective is about, or those semantic elements that comprise the perspective scope (optional).

Note that both the source and the target are optional; this is because we take a lexicalized approach to annotation. That is, although we assume that all three elements are always conceptually present in a perspective relation, we only annotate them if we can anchor them to lexical elements in the text. The cue is the only element that is required to know that there is some perspective expressed; the source and target may very well not be lexicalized. In addition, we adhere to the surface form of the document, excluding any information that may be retrieved from other sources (e.g. world knowledge) from the annotation.

3.3 Multi-layered annotation

An important property of the annotation approach presented in this document is that it is *multi-layered*. Perspective annotation involves capturing a variety of semantic and pragmatic phenomena, such as attribution, factuality and sentiment. However, through previous experience we have found that asking even experienced annotators to analyze texts for various dimensions of information constitutes a high risk for information overload and confusion. By splitting the annotations in subtasks, we avoid the problem of

overloading annotators with the complex interactions between the different information layers involved. Instead, annotators can focus on one information layer at a time, while still being able to use the annotations in the previous layers. In addition, it enables us to better monitor the interactions between the layers.

We consider *events* to be the most basic semantic units that may give rise to or be involved in perspectives. In a text, events function as hubs of information: they inform on the participants, are associated with a time and location, express the attitude of the source, and are building blocks for more complex representations such as TimeLines or Storylines. Therefore, our first layer of annotation corresponds to *event detection*. The annotation of events is described in detail in Chapter 4.

The second layer of annotation corresponds to the identification of *attribution relations* (ARs), which indicate who has expressed a particular stance towards some information. In this layer, we identify the three elements of ARs (source, cue, target) *without* further characterizing it. Chapter 5 gives a detailed description of the annotation of ARs.

The actual characterization of ARs in terms of stance or attitude is done in separate layers. One of them is the layer of event factuality. The guidelines for the annotation of event factuality are provided in Chapter 6. Another layer that is used to characterize the relation between source and target is that of sentiment. Chapter 7 is devoted to the annotation of this fourth layer of perspectives.

3.4 Perspective scope

Section 3.2 briefly introduced the phenomenon of perspective scope, where only specific semantic aspects of an event (or entity) are the target of a perspective. To clarify, first consider Sentence 6 below. In this sentence, *thinks* expresses uncertainty of *Sally* towards the *killing* event. We can interpret this sentence in two ways: either Sally is uncertain about what happened to John (killed or not), or she is uncertain whether it was John who was killed or someone else. Either way, we can be more specific by stating that the uncertainty is directed towards the semantic relation between *John* and *killed*, comprising the whole proposition expressed by *John was killed*.

6. Sally thinks that John was killed.

Now consider Sentence 7. The event targeted by *thinks* is still *killed*, but it does not seem to be the case that Sally questions whether it was John who was killed (or that someone was killed at all); rather, she is uncertain whether he was killed by Kim or by someone else, which interpretation is strengthened by the contrasting perspective of Harry. In other words, the uncertainty only targets a specific participant of the event, namely the KILLER (the Arg0).

7. Sally thinks that John was killed by Kim, but Harry thinks that he was killed by Tom.

Perspective scope is relevant for both factuality and sentiment; for example, in Sentence 8 the positive sentiment (and uncertainty) expressed by *hope* is most likely directed towards *fast* only as the MANNER (or Arg-MNR) of the killing event.

8. Sally hopes that John was killed fast.

It can be a matter of interpretation which aspects of an event or entity are affected by a perspective cue; although less likely, it is also possible to interpret Sentence 8 as if Sally hopes that it was *John* who was killed fast (as opposed to someone else). It is assumed that there are default or preferred interpretations for sentences in isolation; for sentences in context, it is assumed that readers can derive the correct interpretation from that (global) context.

Only when needed, we represent the perspective scope in the annotations. The details of how this should be done for factuality and sentiment can be found in Chapters 6 and 7 respectively.

Chapter 4

Event detection

4.1 Event: A Definition

This section describes the annotation guidelines for marking up all mentions of events in documents. The guidelines are based on the TimeML Annotation Guidelines (Pustejovsky et al., 2006), the NewsReader Annotation Guidelines (Tonelli et al., 2014) and the ECB+ Annotation Guidelines (Cybulska and Vossen, 2014). The guidelines applies to English data and at document and mention levels (see Section 3.1 for details on the distinction between mentions and instances).

TimeML defines events as "situations that happen or occur" (Pustejovsky et al., 2006). In TimeML, *events* is used as a cover term for describing both dynamic, static, durative and non-durative situations. States are additionally defined as "circumstances in which something obtains or holds true" (ibid.). Every event mention, regardless of the fact that it is specific (i.e. referring to an event description associated with a time and place) or generic (i.e. an a-temporal event description or lacking a spatial anchors) are annotated.

Following the ECB+ and the NewsReader guidelines, we include in the event annotation task also the annotation of the participants. The access to this type of information will play a major role in the perspective scope annotation (see Section 3.4).

4.2 <EVENT>: Annotation Guidelines

The tag <EVENT> is used to annotated all linguistic representations of event mentions. It extends the set of eligible linguistic descriptions with respect to the TimeML definition following the proposal in NewsReader. It corresponds to the NewsReader tag <EVENT_MENTION>. In the examples, which will be used to clarify and illustrate the use of the markable, the target linguistic element realizing an event mention is underlined.

4.2.1 <EVENT>: tag extent

Following the surface-oriented TimeML annotation philosophy, event annotation is based on the notion of *minimal chunks*, because higher constituents may contain more than one event expression. This means that only the head of the event denoting chunk will be marked up with the tag. Auxiliaries, polarity mark-

ers, particles, modifiers, complements and specifiers are not part of the extent of the tag. Nevertheless, to be more informative at the semantic layer, the minimal chunk rule is applied in a more flexible manner. **Multitoken event mentions** are allowed though restricted to: a) idioms, b) phrasal verbs, c) prepositional phrases, and d) proper nouns. Multitokens events are always annotated with a unique <EVENT> tag even when their surface realization is discontinuous.

Events can be syntactically realized by the following parts-of-speech:

- **Verbs**: both finite and non-finite verb forms are annotated. The event token corresponds to the verbal head of the VP. Auxiliaries (both temporal and modal) are excluded from the tag extent;

  1. *People are buzzing about the release of the new iPhone.*
  2. *I might come with you if you don't mind.*

  Present and past participles in pre-modifier position must be annotated as events both if the modified elements denotes an entity or an event:

  1. *The crying baby had a high fever.*

- **Nouns**: events realized by nouns can correspond to:

  - nominalizations;
  - event nouns;
  - contextual event readings.

  The "contextual event reading" of nouns refer to type-coercion phenomena due to cases such as the co-occurrence of event denoting objects in argument position of verbs requiring an event as their argument (e.g. *beginning a book*), metonymy (e.g. *The bomb ended the manifestation.*) and similar.

  2. *The flight was scheduled at 08.00 p.m.*
  3. *The meeting lasted 3 hours.*
  4. *A panic attack can be dangerous.*

  Special cases are events realized by proper nouns (e.g. *World War II*). Events realized by proper nouns require the entire noun phrase realizing the event description to be annotated with a single event tag. As multiple information can be associated with an event noun, we claim that different information should be annotated in different layers. For instance, if a date is used to refer to an event, we require that the event description to be annotated in the event layer and the temporal expression to be annotated in the temporal expressions layer (if present or available in the annotation guidelines):

  5. *1972 Summer Olympics*

  Event-denoting nouns in pre-modifier position are never to be annotated as event mentions:

  6. *The election defeat was annoying.*
  7. *The panic attack was scaring.*

- **Adjectives**: they normally denote stative events. Adjectives are to be annotated only when in predicative position, i.e. adjectives that function as the predicative complement of a verb. Only the adjectival head of the predicative construction must be annotated. We further restrict predicative constructions to:
  - copular constructions (*be*, *seem*, etc.);
    8. *The crowd seems angry.*
  - inchoative predicates;
    9. *The crowd became angry.*
  - aspectual predicates;
    10. *The crowd keeps angry.*
  - causative predicates;
    11. *The police made the crowd angry.*
  - perception predicates;
    12. *The crowd looked angry.*
  - evaluative and descriptive predicates.
    13. *The crowd is often described as angry.*

  We also include all predicates which denote or express a change of state in general. In case of doubt DO NOT mark the adjective.

- **Prepositional Phrases**: they denote an event only when in predicative constructions. Only the head of the prepositional phrase (i.e. the preposition) is to be annotated.

  8. *While Paula was at home during the evening, I was at the theater.*

  Prepositional phrases are to be marked up ONLY IF the verb, noun or adjective within the prepositional phrase does NOT denote an event:

  9. *The show has begun with an interview to the president.*

  In case a prepositional phrase is an entry in the British or American version of the Collins English Dictionary online[1], the minimal chunk rule does not apply and the extent of the tag corresponds to the whole expression.

  10. *According to the president you must resign.*
  11. *All people on board died.*

- **Pronouns**: they are to be annotated when they are used to denote coreferential mentions of events.

  12. *The attack killed 10 people and it brought to an end the peace process.*

Multitokens verbal events: Strict application of the minimal chunk rule is not always functional from a semantic point of view. The same rule as for prepositional phrases is applied: the extent of idioms and phrasal verbs are to be annotated as the whole expression when they appear as an entry in the British or American version of the Collins English Dictionary online[2]. In case the components of the multitoken events are discontinuous, they must all be annotated with a unique tag:

[1] http://www.collinsdictionary.com/dictionary/english/
[2] http://www.collinsdictionary.com/dictionary/english/

9. *He switched off the light.* (phrasal verb)
10. *He switched the light off.* (phrasal verb)
11. *My computer has kicked the bucket.* (idiom)

Complex Event Constructions: Special rules apply to complex event constructions.

- Light verb constructions: they are predicates involving a verb with a null or very low semantic content (e.g. *make, get, take, put, have,* etc.) and a noun as its argument. We annotate both the light verb and the noun with two separate event tags:

  12. *He took a shower.*
  13. *I got more support from my boss.*

- Aspectual constructions: predicates involving an aspectual verb (*begin, start, keep, continue, end, terminate,* etc.) requires the aspectual verb and its argument to be annotated with two separate event tags:

  14. *He finished the book.*
  15. *I conclude my talk with these remarks.*

- Inchoative constructions: they express the coming into existence of a situation. Inchoative constructions require that both the predicate and its argument expressing the resulting state must be annotated with two separate event tags:

  16. *He finished the book.*
  17. *I conclude my talk with these remarks.*

- Copulative constructions: they are predicate realized by a copular verb (*be, seem* etc.) followed by an NP, AP or PP. The copular verb must always be annotated. As for the complement NP, PP or AP, general rules specified for nouns, adjectives and prepositional phrases apply:

  18. *Hillary Clinton is the most successful candidate for the 2016 election. - the most successful candidate for the 2016 election* is an entity, not an event.
  19. *The crowd looked angry.*

- Causative constructions: causative constructions are normally composed by a triplet: the logical subject, the causal expression and the event complement. Following NewsReader Guidelines, all elements involved in a causative construction must be annotated with separate event tags:

  20. *The rains caused the floods.*

4.2.2 <EVENT>: attributes

The annotation of event mentions requires the assignments of the values to the following attributes. Most of these attributes are inherited from the NewsReader Guidelines.

- id: unique id automatically generated by the annotation tool;

- pred: the lemma of the event mentions;
- pos: it encodes the different parts-of-speech of event mentions;
- tense: it encodes standard distinctions in the grammatical category of tense;
- aspect: it encodes standard distinctions in the grammatical category of viewpoint aspect;
- nwr_class: it encodes event classes as specified in the NewsReader Guidelines
- timeML_class: it encodes event classes as specified in the TimeML Guidelines;
- modality: it encodes surface distinction concerning the degrees of factivity associated to an event mentions;
- comment.

The *pred* attribute

The attribute corresponds to the lemma associated with the event mention.

The *pos* attribute

The attribute capture the different parts-of-speech which can realize an event mention. Its values are based on standard criteria for part-of-speech classification in linguistics:

- VERB: finite and non-finite verb forms;
- NOUN: both common nouns and proper names;
- ADJECTIVE: adjectives;
- PREPOSITION: prepositions;
- PRONOUN: pronouns.

The *tense* attribute

The attribute encodes the surface values of category of tense. The values are: PRESENT, PAST, INFINITIVE, PASTPART, PRESPART, FUTURE, and NONE.

TimeML annotation guidelines for this attribute apply.

The *aspect* attribute

The attribute encodes the surface values of category of viewpoint aspect. The values are: PROGRESSIVE, PERFECTIVE, PERFECTIVE_PROGRESSIVE, and NONE.

TimeML annotation guidelines for this attribute apply.

The *nwr_class* attribute

The attribute nwr_class assigns at mention level a class value. The attribute is inherited from the News-Reader Guidelines. The attribute values are:

- SPEECH_COGNITIVE: this value is assigned to all event mentions which denote a reporting event (e.g. *say, tell, declare, deny, explanation* etc.) or a cognitive event (mental states, mental acts, mental or cognitive processes; e.g. *think, believe, know, remember, prefer, decide, decision* etc.)
- GRAMMATICAL: this value applies to all events which are semantically dependent on another event or a content noun. Grammatical events:
  - do not introduce other participants;
  - do not have any time span outside the content verb or noun;
  - they do not introduce any change of state that i not already expressed in the governing content verb or noun.

  Some cases of grammatical events:
  - light verbs or copula verbs;
  - aspectual verbs and nouns;
  - verbs or nouns which express causal and/or motivational relations (e.g. *cause, result, lead to, induce, produce, secure* etc.);
  - verbs and nouns expressing occurrence of an event (e.g. *happen, occur, take place, occurrence* etc.)
- OTHER: this value applies to all other events which do not fall into the SPEECH_COGNITIVE or GRAMMATICAL classes.

The *timeML_class* attribute

The values of this attribute are inherited from the TimeML Annotation Guidelines. Its values are:

- REPORTING: "Reporting events describe the action of a person or an organization declaring something, narrating an event, informing about an event, etc" (ISO (2008): 48)
- PERCEPTION: "Events involving the physical perception of another event" (ISO (2008): *ibid.*);
- ASPECTUAL: these events code information on a particular phase or aspect in the description of an event. They are a grammatical device which code a kind of temporal information and focus on different facets of the event history.
- I_ACTION: "I_ACTION stands for intensional action. I_ACTIONs describe an action or situation which introduces another event as its argument, **which must be in the text explicitly**". (ISO (2008): 49).
- I_STATE: "They are similar to the events in the previous class. I_STATEs also select for another event as their argument, but contrary to I_ACTIONs, they denote stative situations" (ISO (2008): *ibid.*).

- OCCURRENCE: This class includes all other types of events describing situations that happen or occur in the world.
- STATE: States describe **circumstances** in which something obtains or holds true. The class of STATE does not contain any instance of I_STATEs which in case of absence of event argument are marked as OCCURRENCE.

The *modality* attribute

This attribute is filled only when a modal word (i.e. a modal auxiliary) modifies the event mention. The value of the attribute correspond to the lemma of the modal word.

4.2.3 BNF description of the <EVENT>: tag

```
attributes ::= id pred pos tense aspect nwr_class timeML_class [modality] [comme
id ::= <integer>
pred ::= CDATA
pos ::= VERB | NOUN | ADJECTIVE | PREPOSITION | PRONOUN
tense ::= FUTURE | PAST | PRESENT | PASTPART | PRESPART | INFINITIVE | NONE
aspect ::= PROGRESSIVE | PERFECTIVE | PERFECTIVE_PROGRESSIVE | NONE
nwr_class ::= SPEECH_COGNITIVE |  GRAMMATICAL | OTHER
timeML_class ::= REPORTING | PERCEPTION | ASPECTUAL | I_ACTION | I_STATE | OCCUR
modality ::= CDATA
comment ::= CDATA
```

Chapter 5

Attribution Relations

The Penn Discourse TreeBank (PDTB) (Prasad et al., 2006) and the Penn Attribution Relation Corpus (Pareti, 2012) are annotated with attribution relations. Provided that the PARC scheme is an extension of the PDTB one, we take as a reference the second. We follow very closely the PARC annotation guidelines (Pareti, 2015), as described below.

Both PDTB and PARC define an attribution relation as a linguistic phenomenon that is used to ascribe to someone the ownership of an attitude expressed in a text. The attitude can be an utterance, a belief or knowledge, or an intention. An attribution relation is typically composed by three elements:

- Source, the owner of the attitude.
- Cue, the linguistic material used to ascribe the attitude to its owner.
- Content, the attitude that is ascribed.

From these elements, the content always has to be expressed for an attribution relation to hold. The source and the cue might be omitted.

Four types of attribution are annotated:

- Assertions: acts of communication.
- Beliefs: the expression of a mental process.
- Facts: the content is presented as a fact.
- Eventualities.

ADD EXAMPLES

The annotation process proceeds as follows:

- Identification of the attribution relation.
- Identification of the constitutive elements of the attribution relation.

- Selection of the text spans that will be labelled as markables. Deciding what is in the scope of the attribution relation
- Linking together all the markables in an attribution relation.

5.1 Attribution elements

5.1.1 Source

The source span should include all the words relevant to the identification the entity having this role. The entity can be named (Mr. Smith) or unnamed (e.g. a man), animate or inanimate (e.g. The White House/ the article) or even implicit (e.g. It was reported that...).

The source markable should always comprehend the full noun phrase expressing it, including appositives (12) or relative clauses (13) referring to the entity in the noun phrase. If the source is represented by an adjective (e.g. the presidential report) the full noun group should be annotated. If it is represented by a possessive pronoun (14), only the possessive pronoun should be annotated[1].

Implicit sources do not have a corresponding markable since they are not expressed in the text. Null or missing subjects, having no corresponding span, should also not be marked.

12. said [Sterling Pratt, wine director at Schaefer's in Skokie, Ill., one of the top stores in suburban Chicago].
13. says [Warren H. Strother, a university oficial who is researching a book on Mr. Hahn].
14. [His] advice: Don't panic.

5.1.2 Cue

The cue can be expressed by a considerable number of elements, although in most cases it is expressed by different types of verbs. Occasionally cues are expressed by other elements.

- Reporting verbs (e.g. say, write, confirm, think)
- Manner verbs (e.g. shrug, beam)
- Other verbs (e.g. add, continue)

The full verbal group should be annotated, including auxiliaries, modals and negative particles. Adverbials adjacent to the cue (e.g. she said angrily) need to be included, since they can modify the verb.

Cues can also be expressed by prepositions or prepositional groups (e.g. according to, for, in the eyes of) and nouns (e.g. report, idea, fear) as in (15).

15. a. Our hope that the Senator and other members of the congressional left.

    b. Even the volatility created by stock index arbitrage and other computer driven trading strategies isn't entirely bad, in Mr. Connolly's view .

[1] In PARC the full NP is annotated.

5.1.3 Content

The content is the span of text that is attributed to the source. The annotation of the content should be limited to the fragments of text about which there is not doubt that they are to be attributed to the source. This means that the content span should not include utterances of uncertain attribution due to syntactic ambiguities. For example, if there is a coordination, the second coordinated clause can be annotated only if it is also surely attributed to the source, as in (16), otherwise it could represent material added by the writer.

16. Still, without many actual deals to show off, Kidder is left to stress [that it finally has "a team" in place, and that everyone works harder].

The content span can be discontinuous. When the content span is separated by an incidental phrase or clause, it should be annotated as a single markable.

17. ['There's no question that some of those workers and managers contracted asbestos–related diseases,"] said Darrell Phillips, vice president of human resources for Hollingsworth & Vose. ["But you have to recognize that these events took place 35 years ago. It has no bearing on our work force today."]

The complementizer that should always be included in the content span, together with the quotation marks. Punctuation at the end of a content span should only be included if part of the content itself. This means that for example a full stop at the end should be included when the content is expressed by a full sentence, a question mark when the content itself is a question and so forth (or when inside the quotation marks).

5.1.4 Supplement

Supplement are additional elements of the attribution relation, which, although not fundamental in an attribution relation, do carry useful information. These can be: concurring to the identification of the source and the provenance or mean by which the information was acquired (e.g. said on the phone); providing further specification of the attitude this holds (e.g. said with anger); the recipient of a reportive verb of the assertion type (e.g. told the jury) or of an eventuality (e.g. Mary expects John to do the shopping); and event specifications (e.g. said last week) providing context indications determinant to the interpretation and comprehension of the content.

Chapter 6

Belief/factuality

6.1 Factuality: A definition

The current chapter provides a definition of event factuality and describes the guidelines for its annotation, which are based on the FactBank Annotation Guidelines (Saurí, 2008) and the NewsReader Annotation Guidelines (Tonelli et al., 2014).

In FactBank, factuality is defined as "the level of information expressing the commitment of relevant sources towards the factual nature of events mentioned in discourse" (Saurí and Pustejovsky, 2009, p. 231). In other words, it expresses "whether events in discourse refer to real situations in the world (facts), have no real counterpart (counterfacts), or are of an uncertain nature (possibilities)" (Saurí, 2008, p. 2). The events to be annotated with factuality are defined in Chapter 4.

The three main elements that form the basis of all perspective relations have the following meaning with respect to factuality:

- **Source:** The lexical elements that refer to the entity committing to the factual nature of the event;
- **Cue:** The lexical elements that express the commitment of the source towards the factual nature of the event;
- **Target:** The lexical elements that refer to the event which factual status is evaluated by the source. It has three attributes:
  - CERTAINTY: expresses the degree of certainty of the source (certain, probable or possible);
  - POLARITY: expresses whether the event is affirmed or negated;
  - TIME: expresses the actual temporal anchoring of the event (future or non-future).

In the examples we use to illustrate factuality annotations, target events are underlined, cues are represented in bold face, and sources are italicized.

6.2 Factuality target

The target of a factuality relation is an event as defined in Chapter 4. Events have already been identified in the event layer. The factual nature of events, which is annotated in this layer, is annotated by using

three attributes: certainty, polarity and time (van Son et al., 2014; Tonelli et al., 2014).[1] For all these attributes counts that the annotators should base their assessment uniquely on the knowledge available in the sentence expressing the event without using world knowledge or other knowledge taken from the text. In *Obama will be re-elected as president in 2016*, for example, *re-elected* is presented as a CERTAIN, AFFIRMATIVE, FUTURE event and should be annotated as such, even though the annotators know from their world knowledge that this is not true.

6.2.1 Certainty

The certainty attribute is used to express the degree of certainty of the source regarding the factuality of the event. The possible values are:

- CERTAIN; the source is certain that the event did or will (not) take place.
- PROBABLE; the source thinks it is probable that the event did or will (not) take place.
- POSSIBLE; the source thinks it is possible that the event did or will (not) take place.
- UNDERSPECIFIED; the source does not know whether the event did or will take place, or it is not expressed.

The default value is CERTAIN; events can only be POSSIBLE or PROBABLE if they are affected by some factuality cue in the text (e.g. *likely, hope, maybe, might*). Factuality cues are discussed in more detail in Section 6.3.

18. Harry will come to the party.
    come = CERTAIN
19. Harry will probably come to the party.
    come = PROBABLE
20. Harry might come to the party.
    come = POSSIBLE
21. Mary does not know whether Harry will come to the party.
    come = UNDERSPECIFIED

POSSIBLE vs. PROBABLE
We follow the guidelines from FactBank (Saurí, 2008) to distinguish between POSSIBLE and PROBABLE events. The idea behind the distinction is that while an event can be *possibly true* or *possibly not true* at the same time, it cannot be *probably true* and *probably not true* at the same time. Whether an event is possible or probable can be tested by denying the original statement using a marker of probability in a context of opposite polarity. If the resulting statement is logical or semantically valid, the correct certainty is POSSIBLE. If it is not, the certainty should be annotated as PROBABLE.

22. Harry might come to the party.
    TEST: *Harry might come to the party, but he probably will not.*
    come = POSSIBLE

[1] The combination of the certainty and polarity attributes corresponds to factuality values defined in FactBank. For example, the combination CERTAIN/AFFIRMATIVE corresponds to CT+ in FactBank.

23. I think that Harry will come to the party.
    TEST: **I think Harry will come to the party, but he probably will not.*
    come = POSSIBLE

6.2.2 Polarity

The polarity attribute is used to distinguish between events that are affirmed and events that are negated. There are three possible values:

- AFFIRMATIVE; the source thinks/knows that the event did or will take place.
- NEGATIVE; the source thinks/knows that the event did or will **not** take place.
- UNDERSPECIFIED; it is not possible to deduce whether the event is affirmed or negated.

The default value is AFFIRMATIVE; the NEGATIVE value can only be assigned if the event is affected by a polarity cue (e.g. *not, never*) or an attributional cue with negative meaning (e.g. *forget, deny*).

24. Harry will come to the party.
    come = POSITIVE
25. Harry will **not** come to the party.
    come = NEGATIVE
26. Harry **forgot** to come to the party.
    come = NEGATIVE
27. Mary **does not know whether** Harry will come to the party.
    come = UNDERSPECIFIED

6.2.3 Time

The time attribute is used to express the temporal interpretation of the event. Its possible values are:

- NON-FUTURE; according to the source, the event did (not) take place in the past or present.
- FUTURE; according to the source, the event will (not) take place in the future.
- UNDERSPECIFIED; it is not possible to deduce the temporal interpretation of the event.

Please note that the time attribute does **not** make explicit the surface tense form value. However, in many cases it is possible to *deduce* the value of a verbal event from the tense of the verb.

28. Harry came to the party.
    came = NON-FUTURE
29. Harry will come to the party.
    come = FUTURE

In other cases, the value of the time attribute cannot be deduced from the syntactic tense of the event. This is the case for: (a) infinitive verbs, (b) verbs preceded by a modal word (e.g. *would*, *may*), and (c) nouns. For these type of events, the annotator should use the context to decide whether the event is NON-FUTURE or FUTURE.

30. Harry might forget to come to the party.
    come = FUTURE
31. Harry would come to the party.
    come = NON-FUTURE
32. Harry came to the party.
    party = NON-FUTURE

Remember that the value of the time attribute is always relative to the relevant source. In the previous examples, the relevant source was the author of the text; in these cases, the value of the time attribute is relative to the time of writing. However, when the event is embedded in an attribution relation, its time should be interpreted in relation to the time of the attribution cue. For example, in Sentence 33 the event *come* is embedded in the attribution relation signaled by *said*. Even though the *coming* event might already have taken place, it should be annotated as FUTURE because, at the time, Sally made a statement about the future.

33. Sally **said** that Harry would come to the party.
    said = NON-FUTURE
    come = FUTURE
34. Harry **decided** to come to the party.
    decided = NON-FUTURE
    come = FUTURE
35. Harry **wants** to come to the party.
    wants = NON-FUTURE
    come = FUTURE

6.2.4 Special cases

Section following soon.

- General statement (GEN);
- Main clause of a conditional construction (COND_MAIN_CLAUSE);
- If-clause of a conditional construction (COND_IF_CLAUSE).

6.3 Factuality cue

We distinguish between two types of factuality cues: attributional cues and non-attributional cues. Attributional cues have two functions: first, they qualify the factual nature of the embedded event; second,

they introduce the sources that commit to that factual nature of the event (therefore, they are also known as source-introducing predicates (SIPs) in the framework of FactBank). For example, the verb *expect* always expresses some degree of uncertainty towards the event according to a specific source (the subject of *expect*). In contrast, non-attributional cues influence the factual nature of the event without introducing an additional source to the text. For example, modal auxiliaries (e.g. *might, may*) express uncertainty about an event, but do not explicitly attribute the uncertainty to a source. All factuality cues have an effect on the event within their immediate local scope, but may contribute as well to characterizing the factuality of events expressed at a non-local level of embedding.

6.3.1 Attributional cues

Attributional cues are factuality cues that contribute an additional source to the text relative to which the factuality of the embedded event is assessed. For example, in the sentences below, both *say* and *suspect* introduce *Western countries* as a relevant source for assessing the factual nature of the event *rigged*. Whereas *suspect* depicts the event as merely a possibility according to Western countries, *say* presents it as a fact.

36. *Western countries* **say** that the election was rigged.
37. *Western countries* **suspect** that the election was rigged.

Per definition, each cue identified in the attribution layer (where statements are linked to their sources through the lexical items that signal the presence of an attribution relation, see Chapter 5) is an attributional factuality cue. Therefore, we do not need to identify them in the factuality layer again.

6.3.2 Modality cues

Modality cues affect the *certainty* of an event. Following (Saurí, 2008), we distinguish between the following modality cues:

- **Modal auxiliaries:** They introduce modality at the local context of a verbal event expression. They include: *can, could, may, might, must,* and *should*.
- **Clausal and sentential modal adverbs:** They introduce modality at a non-local level. Examples are: *apparently, certainly, necessarily, presumably, probably, possibly, reportedly, supposedly*.
- **Modal adjectives:** They are able to introduce modality both at a local level and at a non-local level, since they include both adjectives that function as predicative complements and take a clausal complement (e.g. *it is **possible/probable/likely** that...*), and attributive noun modifiers (e.g. *the **supposed** decline*).
  Examples are: *apparent, certain, hypothetical, improbable, impossible, likely, necessary, presumed, probable, possible, reported, supposed, sure, uncertain, unlikely, unsure*.

6.3.3 Polarity cues

Polarity cues (or negation cues) affect the *polarity* of an event. In general, unmarked contexts are associated with affirmative statements with a positive polarity (e.g. Sentence 38), and polarity cues shift the

polarity into negative (e.g. Sentence 39). However, when the polarity of an event is set to negative by another polarity cue or an event-selecting predicate with negative meaning, such as *denied* in Sentence 40, polarity cues can also shift the polarity back to positive.

38. Mary won the game.
39. Mary did **not** win the game.
40. Paul **denied** that Mary did **not** win the game.

Following (Saurí, 2008), we distinguish between the following polarity cues:

- **Polarity adverbs.** Examples are: *no, nor, neither, never, almost, too*.
- **Polarity determiners.** Examples are: *no, non, neither, little*.
- **Polarity pronouns.** Examples are: *none, nobody, nothing, nowhere*.
- **Polarity affixes.** Examples are: *in-, un-, non-, de-, dis-, a-, anti-, im-, il-,* and *ir-*.

These polarity cues can be introduced at different structural levels: at the clausal level of the event (immediately scoping over the event-referring expression), at the subclausal level (affecting one of the arguments of the event), or at the lexical level (by means of affixes such as *un-* in *unable*) (Saurí, 2008, p. 44-45).

Negating the predicate expressing the event. X

41. Mary did **not** win the game.

Negating one of the arguments of the event. This can be done using polarity pronouns (e.g. *nothing, nobody*) in the head position of the argument, or by means of *no* or *neither* as determiners in the NP expressing the argument. The following examples illustrate negation of the subject:

42. **Nobody** listened to the teacher.
43. **Neither** option was acceptable.

The same applies to the direct or indirect object:

44. The waiter received **no** tips.
45. The teacher gave **nobody** a second change.

Adverbial modification. Certain adverbial constructions negate the event as well. For example, time or place negative adverbials, infinitival constructions dependent on the degree adverb *too*, or adverbs such as *almost* directly modifying the event denoting expression.

46. The children **never** listen to the teacher.
47. The key was **nowhere** to be found.
48. He was **too** proud to even try.
49. Mary **almost** won the game.

6.4 Factuality source

The factual status of a event in text is always relative to a specific source committing to that status. By default, event mentions have an implicit source that corresponds to the author of the text. For example, if the following sentence occurs in a news article, the source that commits to the factual status of the *release* event is the journalist that wrote the article:

41. Apple will release a new version of its iPhone next week.

    Factuality annotations:

source	**cue(s)**	**target**	**certainty**	**polarity**	**time**
author	-	release	CERTAIN	AFFIRMATIVE	FUTURE

Except when the author is lexicalized through the first personal pronoun (*I*, *in **my** opinion*), we do not annotate this source since we cannot anchor it to a lexical item in the text. Instead, by default we take the author to be the source, unless otherwise specified. The latter is the case when the author uses an attributional factuality cue (Section 6.3) to introduce another source to the text that has a stance towards the factual nature of the event. In Sentences 42 and 43, for example, *said* and *fears* introduce *the children* and *Naomi* respectively as relevant sources relative to which the factuality of *ran* and *come* should be evaluated.

42. *The children* **said** that the dog ran away.

    Factuality annotations:

source	**cue(s)**	**target**	**certainty**	**polarity**	**time**
the children	said	ran	CERTAIN	AFFIRMATIVE	NON-FUTURE

43. *Naomi* **fears** that Harry may not come anymore.

    Factuality annotations:

source	**cue(s)**	**target**	**certainty**	**polarity**	**time**
Naomi	fears, may, not	come	POSSIBLE	NEGATIVE	FUTURE

There is a natural *nesting* of sources in texts. For example, the reader of Sentence 43 above only learns the stance of *Naomi* towards the *come* event according to what the author of this sentence asserts. Therefore, strictly speaking, we should say that the relevant source for evaluating the factual status of *come* is *Naomi according to the author*, which we could represent as $\{author\_Naomi\}$. The author himself remains uncommitted to the factual status of *come*. Similarly, in Sentence 44 below, where one attributional cue (*fears*) is embedded in another (*said*), the correct nested source of the *come* event is $\{author\_Paul\_Naomi\}$. In contrast to FactBank, in principal we only annotate factuality values of events according to their most nested source (e.g. *Naomi* for *come*). The reason for this is that we can assume that in most cases nesting sources will remain uncommitted to the factual status of the event.

44. *Paul* **said** that *Naomi* **fears** that Harry **may not** come anymore.

Factuality annotations:

source	cue(s)	target	certainty	polarity	time
author	-	said	CERTAIN	AFFIRMATIVE	NON-FUTURE
Paul	said	fears	CERTAIN	AFFIRMATIVE	NON-FUTURE
~~author~~	-	~~fears~~	~~UNDERSP.~~	~~UNDERSP.~~	~~UNDERSP.~~
Naomi	fears, may, not	come	POSSIBLE	NEGATIVE	FUTURE
~~Paul~~	~~said~~	~~come~~	~~UNDERSP.~~	~~UNDERSP.~~	~~UNDERSP.~~
~~author~~	-	~~come~~	~~UNDERSP.~~	~~UNDERSP.~~	~~UNDERSP.~~

An exception to this rule is when there is an explicit disagreement expressed between the nested source and the nesting source. For example, *unnecessarily* in Sentence 45 expresses that Paul disagrees with Naomi about the factual status of *come*. In these cases, we also annotate the factuality values according to the nesting source (in this case, *Paul*).

45. *Paul* **said** that *Naomi* unnecessarily **fears** that Harry **may not** come anymore.

Factuality annotations:

source	cue(s)	target	certainty	polarity	time
author	-	said	CERTAIN	AFFIRMATIVE	NON-FUTURE
Paul	said	fears	CERTAIN	AFFIRMATIVE	NON-FUTURE
Naomi	fears, may, not	come	POSSIBLE	NEGATIVE	FUTURE
Paul	said, fears, may, not, unnecessarily	come	CERTAIN	AFFIRMATIVE	FUTURE

6.5 Perspective scope

The notion of perspective scope was introduced in Section 3.4, where it was defined as the phenomenon where only specific semantic aspects of an event (or entity) are affected by a perspective cue. In the case of factuality, it means that the identified factuality cues (see Section 6.3) affect only certain aspects of events. In most cases, these aspects are expressed as arguments of an event predicate.

The perspective scope of factuality is only annotated when it is needed. More specifically; we annotate it only when there are multiple predicate-argument relations expressed in the sentence that are candidates for the target of a factuality cue and only one or some of them are affected by this cue. Consider Sentence 46 below; semantically, the uncertainty of *Sally* expressed by *thinks* is directed towards the specific relation between *kill* and its Arg1, *John*. However, it is not needed to annotate this specific relation as the target of the uncertainty expressed by *think*, since there is no other argument expressed in the sentence that is a candidate target as well. Instead, it suffices to directly annotate *kill* (representing the whole proposition) as the target of factuality. This target is represented as f(e2) below.

46. *Sally* thinks$_{e1}$ that John was killed$_{e2}$.

Factuality annotations:

source	target	certainty	polarity	time
author	f(e1)	CERTAIN	AFFIRMATIVE	NON-FUTURE
Sally	f(e2)	PROBABLE	AFFIRMATIVE	NON-FUTURE

In contrast, in Sentence 47 the predicate *killed* has two arguments. The most likely interpretation of the sentence is that it has been established by Sally that John was killed, but that there is uncertainty about

who was the killer. So although there are two candidate arguments in the sentence, the factuality cue *thinks* only affects the certainty of the argument that fulfills the Arg0 role of the predicate *kill*. Therefore, we need to assign different factuality values for each of the target predicate-argument relations, which are represented as f(e2,Arg0) and f(e2,Arg1) below.

47. *Sally* **thinks**$_{e1}$ that John was killed$_{e2}$ by Kim. (But Harry thinks that he was killed by Tom.)

Factuality annotations:

source	target	certainty	polarity	time
author	f(e1)	CERTAIN	AFFIRMATIVE	NON-FUTURE
Sally	f(e2,Arg0)	PROBABLE	AFFIRMATIVE	NON-FUTURE
Sally	f(e2,Arg1)	CERTAIN	AFFIRMATIVE	NON-FUTURE

More examples are shown below; we illustrate the factuality annotations of only those events that require specification on the perspective scope.

48. *I* **think**$_{e1}$ that the mood is fairly gloomy$_{e2}$, and *I* **think**$_{e3}$ it's not going to change$_{e4}$ for a couple of years.

Factuality annotations:

source	target	certainty	polarity	time
I (author)	f(e4,Arg1)	UNDERSP.	NEGATIVE	FUTURE
I (author)	f(e4,ArgM-TMP)	PROBABLE	NEGATIVE	FUTURE

49. By mid afternoon, *official Serb sources* were **saying**$_{e1}$ the operation$_{e2}$ was over$_{e3}$, but that has not yet been **confirmed**$_{e4}$ from *Belgrade*, the capital of Serbia, which is where the whole attack$_{e5}$ is **thought**$_{e6}$ to have been planned$_{e7}$.

Factuality annotations:

source	target	certainty	polarity	time
gen_author	f(e7,Arg1)	CERTAIN	AFFIRMATIVE	NON-FUTURE
gen_author	f(e7,ArgM-LOC)	PROBABLE	AFFIRMATIVE	NON-FUTURE

50. *Nhek Bunchhay* **said**$_{e3}$ *he* now **believed**$_{e3}$ Howes had been killed$_{e3}$ within a week of his capture$_{e3}$ by a Khmer Rouge faction loyal to Pol Pot.

Factuality annotations:

source	target	certainty	polarity	time
he/Nhek Bunchhay	f(e3,Arg1)	CERTAIN	AFFIRMATIVE	NON-FUTURE
he/Nhek Bunchhay	f(e3,Arg0)	PROBABLE	AFFIRMATIVE	NON-FUTURE
he/Nhek Bunchhay	f(e3,ArgM-TMP)	PROBABLE	AFFIRMATIVE	NON-FUTURE

51. The World Court Friday rejected$_{e1}$ *U.S. and British* **objections**$_{e2}$ to a Libyan World Court case$_{e3}$ that has blocked$_{e4}$ the trial$_{e5}$ of two Libyans **suspected**$_{e6}$ of blowing$_{e7}$ up a Pan Am jumbo jet over Scotland in 1988.

Factuality annotations:

source	target	certainty	polarity	time
gen_author	f(e7,Arg0)	POSSIBLE	AFFIRMATIVE	NON-FUTURE
gen_author	f(e7,Arg1)	CERTAIN	AFFIRMATIVE	NON-FUTURE
gen_author	f(e7,ArgM-TMP)	CERTAIN	AFFIRMATIVE	NON-FUTURE
gen_author	f(e7,ArgM-LOC)	CERTAIN	AFFIRMATIVE	NON-FUTURE

52. I don't remember$_{e1}$, maybe Obama was born$_{e2}$ in 1961.

Factuality annotations:

source	target	certainty	polarity	time
I (author)	f(e7,Arg0)	CERTAIN	AFFIRMATIVE	NON-FUTURE
I (author)	f(e7,ArgM-TMP)	POSSIBLE	AFFIRMATIVE	NON-FUTURE

6.6 BNF descriptions

6.6.1 BNF description of the <SOURCE>: tag

```
attributes ::= id [comment]
id ::= <integer>
comment ::= CDATA
```

6.6.2 BNF description of the <CUE>: tag

```
attributes ::= id attributional modifies_modality modifies_polarity [comment]
id ::= <integer>
attributional ::= ATTRIBUTIONAL | NON-ATTRIBUTIONAL
modifies_modality ::= YES | NO
modifies_polarity ::= YES | NO
comment ::= CDATA
```

6.6.3 BNF description of the <TARGET>: tag

```
attributes ::= id certainty polarity time special_cases [comment]
id ::= <integer>
certainty ::= CERTAIN | POSSIBLE | PROBABLE | UNDERSPECIFIED
polarity ::= AFFIRMATIVE | NEGATIVE | UNDERSPECIFIED
time ::= NON_FUTURE | FUTURE | UNDERSPECIFIED
special_cases ::= NONE | GEN | COND_MAIN_CLAUSE | COND_IF_CLAUSE
comment ::= CDATA
```

Chapter 7

Sentiment/opinion

7.1 Opinion: A Definition

This section describes the annotation guidelines for marking up opinions in text. The guidelines are based on the MPQA guidelines (Wiebe et al., 2005), and the OpeNER guidelines for touristic reviews (REF) and opinions in news (REF). Wiebe's definition of *private state*, widely used in opinion mining, describes an opinion as a kind of inner state that is not open to verification (REF Quirck). This notion includes beliefs, thoughts, feelings, emotions, goals, evaluations and jugdments. We adopt a narrower definition and consider only private states that express a **negative** or **positive** attitude as an opinion. These opinions include both the attitudes of the speaker or writer as the attitude attributed to some agent in the text.

7.2 Opinion triplet

Following many others (Wiebe et al., 2005; Toprak et al., 2010), we regard the opinion as consisting of 3 elements:

- holder (who): an entity (person or organisation) who holds the opinion (optional)
- target (about what): the object the opinion is targeted at (optional)
- opinion: a positive or negative attitude

7.2.1 Opinion

The opinion expressions annotations label the words in the sentence that express the opinion. They can be verbs, nouns, adverbs , adjectives, interjections, as illustrated by the following examples. Moreover, opinion expressions may consist of constituents or other pieces of text such as, for example, *the high volume of interest* in ex. (56) and (57).

53. Angela Merkel's immigration policy is **insane**, Donald Trump says.
54. According to Elton John Pope Francis is a **hero**

55. I **thank** him for his work.

56. Due **to the high volume of interest** in the Elton John tickets we have **unfortunately** sold out of tickets.

57. Despite just two albums to her name Amy Winehouse is **one of the biggest music icons** in British history.

Also from a semantic point of view, opinions can be lexicalized by words of different categories, such as emotions (*She **hates** John)*, stance taking expressions (I will **support** you) , agreeing expressions (*They **fulfilled** her wishes*), evaluating expressions (*An **amazing** house*) and others.
Opinion expressions have the following attributes:

semantic orientation/polarity

The opinion expression refers to the span of words that indicate a positive or negative attitude of the holder towards the target. The positiveness or negativeness of the opinion is called semantic orientation. The semantic orientation is identified while taking into consideration negation (cf. ex. 60) and polarity shifters (cf. ex. 61).

58. He is a **hero** - positive

59. I **like** this house - positive

60. I **don't like** this house - negative

61. I think they are **too kind** -negative

strength

The strength of an opinion is indicated by two values: average(default) and strong. It is identified while taking into consideration intensifiers and weakeners

62. He is a **real hero** - strong positive

63. I **adore** this house - strong positive

direct and indirect expressions Following Wiebe Wiebe et al. (2005), we distinguish between two types of opinion expressions direct opinion expressions and indirect opinion expressions[1]

Direct opinion expression denote the opinion of some agent in the text and simultanuously attribute this opinion to him. Indirect opinion are merely connoted or illicted by the way something is described or by a particular wording.

The difference between these two types of expressions is illustrated in the following examples:

64. Amnesty International has *criticized* the sentencing of 13 Catholic activists to up to 13 years in prison. (direct)

[1] similar to Wiebeś direct subjective expressions and expressive subject elements

65. Amnesty International called the punishment *part of an escalating government crackdown on freedom of expression.*(indirect)

In both sentences there is a negative attitude expressed of the opinion holder Amnesty International. In ex. 64 the negative attitude towards the target is expressed by the verb *criticized*, which directly denotes and expresses the type of attitude held by the opinion holder. Also, these kind of attitude expressions often attribute to the opinion holder by means of their syntactic properties: there needs to be someone that *is criticizing*. These kind of expressions are called direct expressions. In ex. 65 on the other hand, the negative attitude is expressed by the words between quotation marks: *part of an escalating government crackdown on freedom of expression*. In this case, the negative attitude can be inferred from the content of the words. The inferred attitude would be something like disapproval. We will call these kind of opinion expressions indirect expressions.
Direct expressions can be verbs (e.g.*blame*), nouns (*refusal*) or adjectives(*willing*) but they must refer to a cognitive process.

66. they demolish archeological sites
    Opinion annotations:

holder	**opinion**	**target**
SW	indirect-negative (demolish)	archeological sites

67. the radio station slanders the prime minister
    Opinion annotations:

holder	**opinion**	**target**
radio station	direct-negative (slander)	prime minister
SW	indirect-negative (slander)	radio station

7.2.2 Holder/Source

The holder of an opinion is the entity whose opinion is expressed. It can be, for example, a person or group of persons (cf. ex.68, 70, 71:*Mbeki*), an organisation (cf. ex.69) and their metonomical counterparts (cf. ex.(71:*newspaper report*, 72). Holders are often realised as the subject of the opinion verb, but they can also be found in other syntactic constructions (cf. ex. 70:*his* and ex.71:*newspaper report*).

68. **He** does not like this house.
69. **The school** refused to admit pupils with special needs.
70. **They** were happy with **his** refusal to join the army.
71. **Mbeki** questions the spendings on HIV/AIDS , according to **a newspaper report**
72. Gay rights is EU entry criterion, **Brussels** says
73. Unfortunately, the concert is cancelled. (**SW**)

The holder may be an agent in the discourse as in examples (68) to (72), but it can also be the speaker or writer(SW) who expresses his opinion (cf. ex. 73). Moreover, people may express opinions about other people's opinions leading to nested sources in one sentence. For example, sentence 70 includes two holders: *they* being happy and *his* who is refusing. The second source is nested as it is only according to *they* that 'he' is refusing. Likewise ex. (72) presents the two holders *newspaper report* and *Mbeki*. The following examples show in more detail how these cases are annotated:

74. Mbeki questions the spendings on HIV/AIDS , according to a newspaper report
    Opinion annotations:

holder	**opinion**	**target**
Mbeki, newspaper report	direct-negative (questions)	spendings

75. They were happy with his refusal to join the army.
    Opinion annotations:

holder	**opinion**	**target**
they	direct-positive (happy)	to join
his,they	direct-negative(refusal)	army

A special kind of holder is the 'anybody*'. This label used for holders that cannot be identified as a specific agent in the text nor as the SW (cf. ex.), but still seem to express an opinion.

76. I don't like this so-called beautiful house
    Opinion annotations:

holder	**opinion**	**target**
I	direct-negative (don't like)	house
I	indirect-negative (so-called)	house
anybody*, I	indirect-positive (beautiful)	house

7.2.3 Target

The target of an opinion can vary from a one word entity to a whole sentence as illustrated by the following sentences. Semantically speaking, the target may be an entity (ex.77) or an event with or without participants (ex.78).

77. I like **John**.
78. I definitely support that **John will be part of the organisation of this meeting**.

A sentence may contain different opinions directed at the same target (79:*Internet Service*).

79. One of the most annoying things is the poor Internet Service.
    Opinion annotations:

holder	**opinion**	**target**
SW	indirect-strong negative (One of the most annoying things)	Internet Service
SW	indirect-negative (poor)	Internet Service

7.3 Lexical elements with multiple functions

Lexical elements may have more than one function. Consider, for example, ex. (80) where the expression *boasting* is the opinion in two separate opinions with different holders and targets. Moreover, the semantic orientation of the verb *boasting* changes according to who is the holder. Example (81) shows that the target and opinion may be lexicalised by the same cue, i.e. *idiot*.

80. Damilola's killers were boasting about his murder..
    Opinion annotations:

holder	**opinion**	**target**
SW	indirect-negative (boasting)	killers
killers	direct-positive (boasting)	murder

81. What an idiot!
    Opinion annotations:

holder	**opinion**	**target**
SW	indirect-strong-negative (**idiot**)	**idiot**

7.4 Some examples

82. When the Imam issued the fatwa against Salman Rushdie for insulting the Prophet
    Opinion annotations:

holder	**opinion**	**target**
Imam	issued the fatwa against	Salman Rushdie
Imam	negative (insulting)	Salman Rushdie
Salman Rushdie, Imam	negative (insulting)	the prophet

7.5 Description of the opinion tags

```
holder::=SW|anybody|CDATA
target::=CDATA
expression::=CDATA
expression attributes::=strength orientation expressionType1 expressionType2
strength::=STRONG|AVERAGE(default)
orientation::=POSITIVE|NEGATIVE
expressionType1::=DIRECT|INDIRECT
```

Bibliography

Cybulska, A. and Vossen, P. (2014). Guidelines for ecb+ annotation of events and their coreference. Technical report, Technical report, Technical Report NWR-2014-1, VU University Amsterdam.

Deng, L. and Wiebe, J. (2015). MPQA 3.0: Entity/event-level sentiment corpus. In *Proceedings of NAACL-HLT (short paper)*.

Diab, M., Levin, L., Mitamura, T., Rambow, O., Prabhakaran, V., and Guo, W. (2009). Committed belief annotation and tagging. In *Proceedings of the Third Linguistic Annotation Workshop*, pages 68–73, Suntec, Singapore. Association for Computational Linguistics.

Fokkens, A., van Erp, M., Vossen, P., Tonelli, S., van Hage, W., Serafini, L., Sprugnoli, R., and Hoeksema, J. (2013). GAF: A grounded annotations framework for events. In *Proceedings of the first Workshop on Events: Definition, Detection, Coreference and Representation*, Atlanta, USA.

ISO, S. W. G. (2008). *ISO DIS 24617-1: 2008 Language resource management - Semantic annotation framework - Part 1: Time and events*. ISO Central Secretariat, Geneva.

Pareti, S. (2012). A database of attribution relations. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC-2012)*, pages 3213–3217.

Pareti, S. (2015). Annotation schema and guidelines for the annotation of PARC 3.0. Technical report, School of Informatics, University of Edinburgh.

Prabhakaran, V., By, T., Hirschberg, J., Rambow, O., Shaikh, S., Strzalkowski, T., Tracey, J., Arrigo, M., Basu, R., Clark, M., Dalton, A., Diab, M., Guthrie, L., Prokofieva, A., Strassel, S., Werner, G., Wiebe, J., and Wilks, Y. (2015). A new dataset and evaluation for belief/factuality. In *Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics (*SEM)*, Denver, USA.

Prasad, R., Dinesh, N., Lee, A., Joshi, A., and Webber, B. (2006). Annotating attribution in the Penn Discourse TreeBank. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, pages 31–38. Association for Computational Linguistics.

Prasad, R., Dinesh, N., Lee, A., J. A., and Webber, B. (2007). Attribution and its annotation in the Penn Discourse TreeBank. *Traitement Automatique des Langues, Special Issue on Computational Approaches to Document and Discourse*, 47(2):43–64.

Pustejovsky, J., Verhagen, M., Saurí, R., Littman, J., Gaizauskas, R., Katz, G., Mani, I., Knippen, R., and Setzer, A. (2006). TimeBank 1.2. Technical report, Linguistic Data Consortium (LDC), Philadelphia, PA.

Saurí, R. (2008). FactBank 1.0 annotation guidelines. Technical report, Brandeis University.

Saurí, R. (2008). *A factuality profiler for eventualities in text*. PhD thesis.

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

Tonelli, S., Sprugnoli, R., Speranza, M., and Minard, A.-L. (2014). NewsReader Guidelines for Annotation at Document Level. Technical report, Technical report, Technical Report NWR-2014-2, Fondazione Bruno Kessler, Trento.

Toprak, C., Jakob, N., and Gurevych, I. (2010). Sentence and expression level annotation of opinions in user-generated discourse. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 575–584. Association for Computational Linguistics.

van Son, C., van Erp, M., Fokkens, A., and Vossen, P. (2014). Hope and fear: Interpreting perspectives by integrating sentiment and event factuality. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC-2014)*, pages 3857–3864, Reykjavik, Iceland.

Werner, G., Prabhakaran, V., Diab, M., and Rambow, O. (2015). Committed belief tagging on the FactBank and LU corpora: A comparative study. In *Proceedings of NAACL Workshop on Extra-propositional aspects of meaning in computational linguistics (ExProM)*, Denver, USA.

Wiebe, J., Wilson, T., and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2-03):165–210.

Wilson, T. (2008). *Fine-grained subjectivity and sentiment analysis: recognizing the intensity, polarity, and attitudes of private states*. PhD thesis, University of Pittsburgh.