

# A Factuality Profiler for Eventualities in Text

## Dissertation Preview

Roser Saurí i Colomer

### Abstract

Basic tasks of Information Extraction, such as named entity recognition, event extraction, or even semantic role labeling, are at the core of a variety of NLP applications, ranging from those strictly oriented to building lexical resources, to others which require some degree of text understanding, such as Question Answering or Summarization. Current results in those IE tasks suggest that we are close to achieving a stage where the fundamental units for text understanding are put together; namely, predicates and their arguments. However, other layers of information, such as the factuality status of events, are essential as well. Having access to mentions of events in a document is of no use if we do not know whether they refer to situations that happened or hold in the real world. In other words, the inferences derivable from factual eventualities are different from those judged as possible or as non-existent.

This dissertation aims at designing and developing a *factuality profiler*, namely a tool devoted to identifying the factuality degree associated with each eventuality mentioned in text. Such a tool will be grounded on linguistic expressions we as speakers use to signal factuality information; for instance: polarity and modality particles, lexical triggers of non-existential presuppositions, and certain syntactic constructions. In that sense, it will disregard external factors, like event source trustworthiness or reader's beliefs, and instead it will adopt a neutral position, extracting factuality values assigned to events based on how they are presented in text. We will see that this approach confers a central role to sources. Different sources may introduce divergent, even contradictory, factuality values to the same event. Hence, event factuality degrees is always relative to a source.

## 1 Introduction

### 1.1 Motivation

Eventualities in discourse can be couched in terms of a veridicality axis that ranges from truly factual to counter-factual, passing through a whole spectrum of different modality shades, including:

1. **Degrees of possibility:** *These results indicate that Pb2+ **may** inhibit neurite initiation by inappropriately stimulating protein phosphorylation by CaM kinase.*
2. **Belief:** *Chinese analysts **believe** that the United States will continue to provoke North Korea.*
3. **Evidentiality:** *Subcomandante Marcos **said** that the Mexican government is not interested in putting an end to the conflict.*
4. **Expectation:** *Hans Blix **wants** the US to allow UN inspectors back into Iraq to verify any weapons found by coalition forces.*
5. **Attempting:** *George Mallory and Andrew Irvine first **attempted** to climb Everest in 1924.*
6. **Command:** *John Murtha **called** for the immediate withdrawal of U.S. troops from Iraq.*

Event factuality (or its complementary face, modality)<sup>1</sup> is a necessary component for representing eventualities in discourse, together with other levels of information such as argument structure or temporal information. Inferences derived from events that have not happened or have only possibly happened, are different from those derivable from events that are judged as factual in nature.

The need for a more sophisticated approach sensitive to this additional level of information is now becoming apparent in highly domain-oriented disciplines such as bioinformatics (Light et al., 2004). The modality marker *may* in item 1 above, for example, has effects on the pathway between bioentities that can be abstracted from the reported data.

Factuality information is also starting to be taken into account in more genre-oriented applications, such as Question Answering. For example, several of the systems that participated in the pilot evaluation on Text Entailment, developed under the Question Answering initiative program AQUAINT, attempted to cope with modality information to some degree. Consider example (1). Any QA system that would disregard the attempting context in which the event of Mallory and Irvine climbing Everest in 1924 is embedded, would be led to erroneously report (1c) as the answer to question (1b). Similarly, disregarding the reporting context in example (2) could mistakenly lead to the conclusion that no one from the White House was involved in the Watergate affair:

- (1) a. Of the 14 known ways to reach the summit, only the East Ridge route has never been successfully climbed since George Mallory and Andrew "Sandy" Irvine first attempted to climb Everest in 1924.
- b. When did George Mallory and Andrew Irvine first climb Everest?
- c. #In 1924.

---

<sup>1</sup>From this point forward, I will adopt the term factuality to refer to both faces: factuality and modality.

- (2) a. Nixon claimed that White House counsel John Dean had conducted an investigation into the Watergate matter and found that no-one from the White House was involved.
- b. What members of the White House were involved in the Watergate matter?
- c. #Nobody.

## 1.2 Goal of the Research

My dissertation aims at designing and developing a *factuality profiler*, namely a tool devoted to the identification of the factuality degree associated with each eventuality mentioned in a given text. The question from which to approach this goal is the following:

What support do we have, as readers of news text, about whether an event has happened or will take place, or whether a situation holds or will hold in the world?

Knowing what events are mentioned in a text is of no use if we cannot know whether they have happened (or will happen) as well. To a certain extent, that information is a matter of perspective and depends on external elements such as: (a) who the reader is and how credulous s/he is; (b) what is the event provenance, i.e., who is the source; and finally (c) how much the reader trusts that source. That is, two different readers can have different takes on the situation being narrated. Take for instance the following text:

- (3) A U.N. war crimes tribunal official said Tuesday that the court was informed on several occasions that Slobodan Milosevic had regular access to unprescribed medication and alcohol smuggled into his prison cell.

The official, requesting anonymity because of the tribunal's strict confidentiality rules, told The Associated Press that the unit's prison warden had told the court that he could therefore no longer guarantee Milosevic's health. (CNN. March 14, 2006)

How certain are we, as simple newspaper readers, of the truthfulness of the U.N. official? How certain are Milosevic's followers? How certain can we be about what the prison warden reported? The present research will disregard external factors such as source reliability. Rather, it will focus on the linguistic expressions used in natural language to denote participants making judgements about the factuality nature of events, independently from their intentions and beliefs.

Indeed, the fact that an event is depicted in a text as happening or not happening in a particular time does not mean that this is the case in the world, but that this is how the author of the text encoded the factuality of that event. The factuality nature of an event is therefore always relative to at least one

*source*, the text author, but possibly more –for example, the U.N. official in (3). My factuality profiler will therefore focus only on what the author and other sources conveyed in the text, disregarding any reader’s particular take on the veridicality of each event.

The functionality of my factuality profiler will break down into two major tasks:

- I. Identifying and representing the information providing factuality judgments about eventualities in text. This will involve:
  1. Identifying the linguistic means that trigger that information: factuality markers and sources.
  2. Defining an adequate representation for them.
- II. Computing the factuality degree associated with each eventuality, based on the interaction of the different factuality markers ranging over it, and relative to the different sources mentioned in the text.

The factuality profiler will build on top of other basic applications, such as named entity recognizers and event extractors, and will be a support tool for tasks requiring some degree of text understanding, by providing information relative to the factuality status of events mentioned in discourse. Its input will be a text (or set of texts). As output, it will return each event entity in the text characterized with one or more pairs of a factuality value associated with a source.

In the design of the factuality profiler, I will adopt the following two basic constraints:

1. A factuality profiler must be grounded only on the linguistic expressions that signal factuality.
2. Whenever appropriate, a factuality profiler must be capable of representing different (and possibly contradictory) information about the factuality nature of a particular event.

## 2 Theoretical Settings

Event factuality is a convenient category I adopt for practical reasons. Broadly speaking, it expresses whether an event is presented in text as factual, uncertain, or counter-factual. In that sense, it is related to other notions already established in the literature, such as epistemic modality and evidentiality. In this section, I go over some examples in which events are characterized in different ways (section 2.1), and then present a first approximation to the category of event factuality (section 2.2).

Before embarking, we need to address a further terminological issue: the definition of *event*. Throughout this work it will be used as a synonym of the

ontologically broader notion of eventuality, thus including both states and dynamic events. Similarly, I will not distinguish between events (or eventualities) and other abstract types widely discussed in the literature, such as propositions, facts, possibilities, projective propositions (or outcomes), and situations (cf. Vendler, 1967; Barwise & Perry, 1983; Asher, 1993; Peterson, 1997; Ginzburg & Sag, 2000). Instead, I will take a more pragmatical approach and assume as events those situations in the world referred to by a linguistic expression, regardless of whether they are presented as propositions (as in contexts of belief), facts (factive contexts), or any other abstract type.

## 2.1 Did it happen? The way events are characterized in text

Consider the following examples (4-6), extracted from TimeBank1.2, the CNN archive, and BNC. Depending on how events (marked in bold face) are presented by the author, the reader will evaluate their factuality status differently.<sup>2</sup>

- (4) a. Five other U.N. inspection teams **visited** a total of nine other sites.  
 b. The size of the contingent **was not disclosed**.  
 c. The vines had just finished **flowering**, and tiny dark green grapes hung intermittently among the foliage.  
 d. As for the court, it managed with its splintered ruling **to cast doubt** both on the constitutionality of the immigration provisions it upheld and on its true commitment to fighting gender stereotypes.  
 e. She regretted **calling** the prime minister a traitor.

Events marked in bold face in (4) are all characterized as factual –or counterfactual (4b); that is, as events that took place at some point or that hold in the world –or that did not, in the case of counterfactuals. This is achieved through different mechanisms, such as by embedding the predicate as the complement of an aspectual predicate (4c), or a factive or implicative predicate (4d-4e) (Kiparsky & Kiparsky, 1970; Karttunen, 1970), or by means of a plain declarative sentence of positive or negative polarity (4a-4b).

- (5) a. Amir **may have been trying** to impress Har-Shefi by killing the prime minister.  
 b. United States **may extend** its naval quarantine to Jordan’s Red Sea port of Aqaba.  
 c. They **may not have enthused** him for their particular brand of political idealism.

---

<sup>2</sup>Here and throughout this work, events in the examples I present will be identified by marking only the verbal, nominal, or adjectival predicates expressing them, together with polarity particles and auxiliaries when deemed necessary. This goes along the lines defined in TimeML, the specification language to represent event and temporal information in text (Pustejovsky et al., 2003).

Sentences in (5) are also declarative, but in this case the events in question are characterized as uncertain, given the presence of the modal auxiliary *may*. The use of the future tense in (5b) colors the event as even less certain than the one in the previous sentence.

- (6) a. "We know a peaceful solution **was in reach**."
- b. They know that al-Qaida and the Taliban **are not the real face** of political Islam.
- c. Experts say the ground **is so saturated** it cannot absorb any more water.
- d. I believe they **have the culture** to adequately service high-net-worth individuals.

In examples (6a-6d), the factuality status of the events depends on a variety of factors. As can be observed, all of them are embedded as complements of other predicates (underlined). In a way comparable to modal auxiliaries, some of these embedding predicates seem to commit the author more strongly to the factual nature of the embedded event than others. Compare for instance the effect of *know* in (6b) with the effect of *believe* in (6d).

Equally important are the participants committing to event factuality. The same predicate *know* is used with a subject in third person in (6b), and a subject in first person in (6a). In none of these two examples does the subject (the referent of *they* and *we*, respectively) correspond to the author of the text. Interestingly however, in example (6b) the author commits to the veridicality of the event expressed by the complement, whereas in (6a) s/he remains neutral and presents the event as somebody else's opinion.

- (7) a. The analysis showed unemployment rates **dropped** steadily with higher levels of education.
- b. This would prove **positive** towards society in the future.
- c. He showed whether specific cells **undergo** programmed cell death as an integral part of the normal differentiation process.

Verbs of inference (7) do also tend to support strong commitment towards the factuality of their embedded event. We see in these examples, however, that the factuality degree of the embedded event can be affected by the presence of modal auxiliaries in the matrix clause (7b) or the type of the complement clause (7c).

- (8) a. The president also has offered to **help** offset Jordan's costs.
- b. "Anheuser is the biggest guy in the bar, and he just decided to **join** in the barroom brawl," said Joseph J. Doyle.
- c. It will give the rest of the world the view that Cuba is like any other nation, something the US has tried to **persuade** the world that it is not.

Finally, commissive predicates (8a-8b) or predicates of attempting (8c) characterize their event-denoting complements as counterfactual at the time the commitment or the attempt event takes place, but uncertain at a future time, unless more information is provided.

There are therefore different ways used to characterize the factuality nature of events in text. Sometimes the factuality status is introduced in absolute terms, as a non-negotiable feature: the event is asserted either as a fact in the world or as a counter-fact. This is the case with simple affirmative or negative declarative sentences, containing no modality information of any sort, but it can be observed in subordination contexts as well, such as those introduced by aspectual (4c), implicative (4d), or factive predicates (4e).

Some other expressions convey different degrees of uncertainty about whether something holds or not in the world; that is, whether an event can be considered factual or counter-factual. Modality operators are among them (5), but also commissive or volitional predicates (8), for instance, whose complement factuality is unknown unless more information is provided by the context.

Finally, event factuality is a more complex matter for those situations that are introduced in discourse subject to a particular source. This is so for contexts of evidentiality, belief, and knowledge, which are triggered by predicates of reporting (6c), inference (7), direct perception, knowledge (6a-6b), and belief (6d).

## 2.2 Event Factuality

Event factuality as understood here is the category in charge of the factual status of events. That is, it conveys whether event-denoting expressions are presented as corresponding to real situations in the world (i.e., as facts), to situations that have not happened or hold (counterfactuals), or to situations of uncertain status (possibilities). As such, it is a property assigned to events expressed in text.

In some contexts, the factual status of events is presented with absolute certainty. Events are therefore characterized as either *facts* (9a) or *counterfactuals* (9d). Other contexts, however, introduce different shades of uncertainty, which can be mapped into an axis of positive polarity, and then events are qualified as *possibly factual* (9b), or along an axis of negative polarity, marking events as *possibly counterfactual* (9c).

- (9) a. Five other U.N. inspection teams visited a total of nine other sites.  
b. United States may extend its naval quarantine to Jordan's Red Sea port of Aqaba.  
c. They may not have enthused him for their particular brand of political idealism.  
d. The size of the contingent was not disclosed.

Factuality can therefore be characterized by means of a double-axis scale. On the one hand, the axis of certainty, ranging from uncertain (or possible) to absolutely certain (or necessary). On the other hand, the axis of polarity,

consisting on two values: positive and negative. The resulting scale is illustrated in Figure 1.

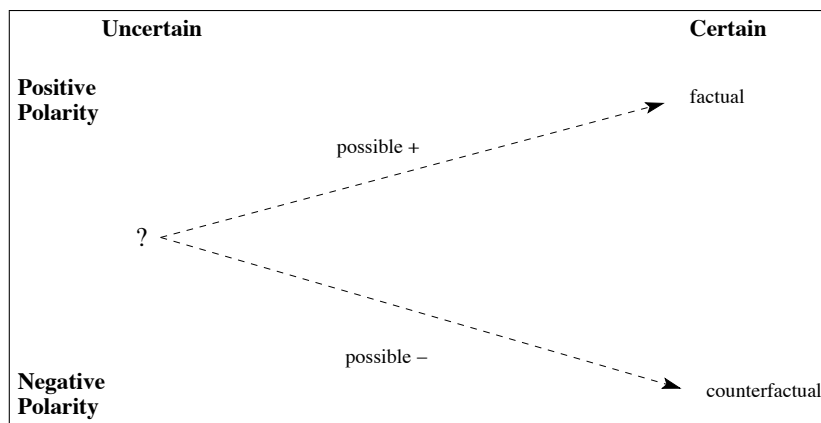


Figure 1: The double range of factuality force

As defined, factuality touches on epistemic modality, a category more standardly assumed in the linguistics literature. Chafe (1986) defines it as the degree of reliability of a situation being a fact, as assessed by authors. And according to Palmer (1986), one of the most complete works on the topic compiling data from multiple languages, it expresses the degree of commitment of the author towards the certainty of an uttered proposition. My notion of factuality stems from this category but it differs from it in that I will take event factuality as reflecting through grammaticalized but also not-grammaticalized devices. This broadens the notion of epistemic modality as explicitly defined in de Haan (1997). In this way, the factuality profiler can exploit the different relevant sources of information present in English: modal auxiliaries but also adverbials, verbs, nouns, and adjectives.

## 3 Basic Components

### 3.1 Factuality Markers

#### 3.1.1 The elements express event factuality

In the present section I examine the linguistic means that are used to express event factuality in natural language, which I call *factuality markers*. For practical reasons, my work will be based on English. That requires a thorough examination of how factuality is conveyed in this language.<sup>3</sup> Factuality markers

<sup>3</sup>Most of what will be presented here is extrapolative to at least other Indo-European languages, such as the Romance and Germanic ones.

can be lexical items or syntactic constructions.

**Polarity particles.** They express the positive or negative factuality of a given event represented in a text. Generally, positive factuality is denoted by the absence of any polarity particle. Polarity particles can be introduced at the local level of event-denoting expressions (that is, within its phrase scope) (10a), at their non-local level but still within the clause (10b), or at the level of a higher clause, in certain cases of subordination (10c). In the following examples the negative particle is in bold face, whereas the head of the event-denoting expression is underlined:

- (10) a. The size of the contingent [<sub>VP</sub> was **not** disclosed].  
b. **Nobody** [<sub>VP</sub> hurried her up].  
c. [Colin Powell **denied** [<sub>S'</sub>that Iraq had weapons of mass destruction]].

Polarity particles leave no room for uncertainty –unless they are combined with other elements, such as modality operators (see next section).

**Modality particles.** They contribute different shades of factuality to a given event. As such, they are based on a non-discrete scale and may leave some room for uncertainty, contrary to the case of polarity particles. In English there are two main sources of modality information: modal auxiliaries, which introduce modality at the syntactic local context of the event expression (in case of verbal-expressed events) (11), and clausal and sentential modal adverbials, which are expressed at a non-local level (12).

- (11) a. Possibility: *could, may*.  
b. Obligation: *must, have to*.  
c. Necessity: *need*.
- (12) a. Possibility: *probably, perhaps*.  
b. Frequency: *usually, always*.

**Lexically-triggered subordination contexts.** In many cases, event factuality is conveyed by predicates that select for a situation as (at least) one of its argument types. I will call them Situation Selecting Predicates (SSPs). SSPs can belong to different parts of speech. My work focusses on the following cases:

- **Verbal predicates** that subcategorize for a *that*-, gerundive, or infinitival clause. They can express:
  - Different degrees of certainty/possibility about the situation denoted by the embedded complement. These include verbs referred to by Hooper (1975) as weak assertive predicates (*think, believe, suppose*). In some cases, what she calls strong assertive predicates (*claim, say, affirm*) can be used in the same manner as well.

- Factuality of the situation expressed by the argument: *regret, forget*.
  - Evidentiality: *say, report*.
  - Command: *force, order*.
  - Commitment: *offer, propose*.
  - Volition: *want, wish*.
  - Agreement/plan: *decide, agree, plan*.
  - etc.
- **Adjectival predicates** subcategorizing for a *that*- or an infinitival clause, and which express:
    - Possibility: *possible that*.
    - Inclination: *willing, ready, eager, reluctant*.
  - **Nominal predicates** related to verbs and adjectives in the two previous bullets.

These predicates, or subsets of them, have been analyzed from multiple perspectives (e.g., Vendler, 1967; Kiparsky & Kiparsky, 1970; Karttunen, 1970, 1971; Hooper, 1975; Bach & Harnish, 1979; Ballmer & Brennenstuhl, 1981; Wierzbicka, 1987; Bergler, 1992; Asher, 1993; Dor, 1995; Peterson, 1997; Ginzburg & Sag, 2000; Koenig & Davis, 2001; Jackendoff & Culicover, 2003). Furthermore, previous work within the Systemic Functional Grammar provides an exhaustive characterization of such markers in English (e.g., Halliday, 1985; Eggins, 1994).

**Constructionally-based subordination contexts.** Some subordination relations are not brought about by a specific predicate but by a discourse relation between two clauses. Many of these constructions introduce factuality information of some sort. The following list, although not exhaustive, gives an indication of how pervasive this phenomenon is.

- **Relative clauses:** The event denoted by the relative clause (underlined in the following example) is presupposed as true.
 

(13) Rice, who became secretary of state two months ago today, took stock of a period of tumultuous change.
- **Cleft sentences:** The event of the embedded clause is presupposed as true.
 

(14) It was Mr. Bryant who, on July 19, 2001, asked Rep. Bartlett to pen and deliver a letter to him.
- **Purpose clauses:** The event denoted by the clause is intensional in nature.

(15) The environmental commission must adopt regulations to ensure people are not exposed to radioactive waste.

- **Conditional constructions:** The event denoted by the consequent clause (underlined) is intensional and dependent on the factuality of the event denoted in the antecedent clause (bold face), which is also intensional.

(16) On Dec. 2 Marcos promised to return to the negotiating table if the conflict zone was **demilitarized**.

- **Subordinated temporal clauses:** The event in the temporal clause (underlined) is presupposed as true.

(17) While Chomsky was revolutionizing linguistics, the rest of the social sciences was asleep.

### 3.1.2 Percolation of factuality values

In order to compute the appropriate factuality force of eventualities in text, defining the semantics of factuality markers is not enough, since in most cases event factuality is established through the interaction of more than one factuality marker. For example, the proposition in (18a) conveys its referred situation (the police being tipped off) as a fact. But when embedded within the SSP *deny* (18b) this value is overwritten.

- (18) a. Drug dealers were tipped off before the operation.  
b. Yesterday, the police **denied** that [drug dealers were tipped off before the operation].

Sometimes, we find several levels of embedding, each of which may affect the factuality force of any of the eventualities under their scope. Take for instance the sentence in (19), original from the BNC. As it is, it expresses that there are currently no fire brigade inspections allowed by the Royal Family –thus, the event defined as  $e3$  receives a counterfactual force.

- (19) The Royal Family will **continue** <sub>$e1$</sub>  to **refuse** <sub>$e2$</sub>  to **allow** <sub>$e3$</sub>  detailed fire brigade **inspections** <sub>$e4$</sub>  of their private quarters.

Nonetheless, suppressing some of the embedding levels of this eventuality or changing one SSP for another one, will lead to a different inference: in (20b) the state of the Royal Family allowing inspections does not hold either, but it is however a fact in (20a) and (20c).

- (20) a. The Royal Family **allows** <sub>$e3$</sub>  detailed fire brigade **inspections** <sub>$e4$</sub>  of their private quarters.  
b. The Royal Family **refuses** <sub>$e2$</sub>  to **allow** <sub>$e3$</sub>  detailed fire brigade **inspections** <sub>$e4$</sub>  of their private quarters.

- c. The Royal Family will **stop**<sub>e1</sub>, **refusing**<sub>e2</sub> to **allow**<sub>e3</sub> detailed fire brigade **inspections**<sub>e4</sub> of their private quarters.

Tackling the issue as a presupposition problem, Karttunen (1973) already points out the effect of different types of clause embedding predicates (what I call SSPs) in projecting the factuality value of the embedded event. More than 30 years later, the interest for this issue is being renewed and it is currently explored from a more applied perspective, mainly within the context of textual entailment.

Worth mentioning is the work by Nairn et al. (2006), which puts forward an algorithm for computing the projection of implications through embedded contexts, involving mainly *that* and infinitival clauses. The work is a promising step, but it is limited in that it disregards basic groups of predicates, it obviates the role of modality, and it does not commit to the presence of factuality sources. In other words, it does not tackle the issue from a general approach that contemplates it as the problem of handling modality and evidentiality information in natural language text.

## 3.2 Sources: The Factuality Evaluators

### 3.2.1 Factuality sources as basic discourse participants

Events mentioned in discourse, be it oral or written, have always an implicit source who commits to their factuality force. By default, the author of the text (a speaker or a writer) is making such a commitment, which is expressed by means of factuality markers as those introduced in the previous section. In (21a), for instance, the positive polarity of the sentence commits the author (CNN) to the factuality of Milosevic being on trial. Nonetheless, sentences of evidential type have an additional explicit source (underlined in example (21b)), relative to which the factuality of a given event is also assessed. Here there are two different source at play: the author, as in the previous sentence, and Milosevic’s lawyer.

- (21) a. Milosevic was on trial for 66 counts of genocide, crimes against humanity and war crimes in the Balkans during the 1990s. (CNN, March 13, 2006)
- b. Milosevic’s lawyer, Zdenko Tomanovic, on Monday denied news reports that the ex-president altered his medication. (CNN, March 13, 2006)

For our purpose, event participants are as important as factuality markers. Different participants may have different perspectives about the factuality nature of the very same event, as illustrated by example (25). In (25a), Condoleezza Rice and Colin Powell present the state of Iraq having weapons of mass destruction as counterfactual, whereas in (22b), a different participant, President Bush, presents the same event as factual.

- (22) a. In mid-2001, Colin Powell and Condoleezza Rice both publically **denied** that Iraq had weapons of mass destruction.
- b. Six months later, President Bush claimed that Iraq had WMD.

Thus, the factuality degree of events, as expressed through discourse, is always relative to the participants or what from now on I will refer to as *factuality sources*.<sup>4</sup>

Distinguishing the different layers of sources that assess the factuality of every event in text is of interest here because they may bear disagreement among them. Combining the source parameter with that of the time of commitment (that is, the point in time where a source commits to a factuality value for a given event), the following situations may hold:

**Situation 1** Within the same sentence, two sources present different views about the factuality degree of the event.

- (23) Sen. Barbara Boxer of California **said** Rice has not **acknowledged** those **deaths**. (CNN, January 19, 2005)

Example (23) illustrates this situation. The sources involved in the statement are the following three: the author of the report (say, CNN), Barbara Boxer, and Condoleezza Rice. There are also three events expressed here, marked in bold face. What is of interest here is the disagreement between the factuality status of the deaths events. The use of the predicate *acknowledge* commits Boxer to their factuality while expressing that Rice remain uncommitted to it.

**Situation 2** The same participant can present a different view at different points in time. The following sentence, for example, distinguishes between a time in which Van Anraat was ignorant about a factual event, and a time, presumably now, in which he is aware of it.

- (24) Van Anraat's attorneys argued he did not know the chemicals he provided to Iraq would be used in attacks. (CNN, December 23, 2005)

The disagreement can also be brought across different sentences or documents. The following two assertions can belong for example to two different documents of our knowledge textual base:

- (25) a. In mid-2001, Colin Powell and Condoleezza Rice both publically denied that Iraq had weapons of mass destruction. (North Texas Indy Media Center, Oct 17 2005)

---

<sup>4</sup>Because of the inclusion of sources, my work partially overlaps with the one developed in the Multiple Perspective Question Answer framework (Wiebe et al., 2005). In particular, the overlap is on the subset of opinions about whether something has happened, happens, or will happen –in other words, opinions about the factuality nature of events.

- b. Secretary of State Colin Powell Thursday defended the Bush administration's position that Iraq had weapons of mass destruction. (CNN, January 8, 2004)

In (25b), the main clause presenting Colin Powell claiming the factuality of Iraq having weapons of mass destruction is anchored at a temporal point in 2004 (that of the document creation time), three years later than the anchor for Colin Powell sustaining the opposite in sentence (25a).

**Situation 3** There is indeed a further situation in which there are different views about the factuality nature of the same event, triggered not only by different participants but also by a difference in the time when the views are expressed. This possibility is subsumed by situations 1 and 2 above.

### 3.2.2 Linguistic means for agreement and disagreement

Linguistically, factuality disagreements between different sources can be prompted lexically, using for instance predicates of belief or knowledge. The choice of one or another predicate conveys a different commitment of the author. Contrast for instance the following two fragments:

- (26) There are certain shameless politicians that will play<sub>e1</sub> that card because they<sub>s1</sub> **know**<sub>e2</sub> that it resonates<sub>e3</sub> with certain people in their constituency. (CNN. February 21, 2006)
- (27) Members of the sect are so devoted<sub>e1</sub> to tradition they even oppose<sub>e2</sub> the state of Israel, because they<sub>s1</sub> **believe**<sub>e3</sub> that biblically Jewish sovereignty over the ancient Land of Israel can come<sub>e4</sub> only with the Messiah. (CNN. April 25, 2006)

In the first one, by using the predicate *know* the author (source *s0*) is not only stating that the resonating event *e3* is a fact according to source *s1* (the so-called shameless politicians), but also according to himself. By contrast, the agreement between the two sources is not possible in (27): the choice of *belief* by the author expresses that it is only according to *s1* (in this case, the members of the sect) that the complement event, *e4*, is a fact.

In other words, predicates of belief do not commit the author of the text to the factuality of the event denoted by the complement, but complements of knowledge predicates do. This is however the case only in affirmative sentences. The use of knowledge predicates in negative sentences may also trigger disagreement between sources.

- (28) Van Anraat's attorneys<sub>s1</sub> argued<sub>e1</sub> he<sub>s2</sub> (Van Anraat) **did not know**<sub>e2</sub> the chemicals he provided<sub>e3</sub> to Iraq would be used<sub>e4</sub> in attacks. (CNN, December 23, 2005)

By using the SSP *know* in a context of negative polarity, Van Anraat’s attorneys are not only asserting their customer had no knowledge of the event denoted by the complement, but that they themselves are aware that such event is a fact.

A second common way of expressing disagreement between sources is at the discourse level, by establishing opposition relations between clauses. Consider the following example, where the factuality marker is a predicate of perception:

- (29) Yesterday , the police **denied** that [drug dealers were tipped off before the operation]. However, it emerged last night that a reporter from London Weekend Television unwittingly tipped off residents about the raid when he phoned contacts on the estate to ask if there had been a raid —before it had actually happened .

In spite of the pervasivity of these sort of oppositions throughout discourse, the present work will obviate this type of factuality disagreements and will focus only on those that are triggered by lexical means, mainly SSPs of knowledge, belief and report.

### 3.3 The role of Time

I argued in the previous section that disagreements between factuality values of the same event can be due to the different points in time in which sources (the same ones or different) commit to the factuality of the event. Time-stamping each factuality commitment act with the time in which it takes place is therefore a fundamental capability that any factuality profiler must have.

### 3.4 Factuality profiles

Every time an eventuality is mentioned in discourse, there is an act of commitment towards its factuality by the author or any other source. In some cases it is an implicit act, as when using plain declarative sentences to assert a fact in the world. But in others situations, the act of commitment is made explicit, as in reporting or knowledge contexts, which provide some sort of evidentiality towards the factuality of the event.

We will define *factuality commitment acts* as involving four components:

1. The event at focus, **e**.
2. The source assigning a factuality value to that event, **s**.
3. The factuality value been assigned.
4. The time of assignment of this factuality value, **t**.

We saw that different sources can commit to the factuality of the same event, or that the same source can do it at a different points in time. Therefore, every event expressed in discourse can have assigned to it one or more factuality

commitment acts. The *factuality profile* of an event will be the set of factuality commitment acts it has assigned to it. Formally, we can represent it as follows:

$$(30) \text{ fp}(\mathbf{e}) = \{ \langle \mathbf{s}, \mathbf{t}, \mathbf{f} \rangle \mid \mathbf{s} \text{ is a source, } \mathbf{t} \text{ is a temporal expression, and } \mathbf{f} \text{ is a factuality value} \}$$

## 4 Designing a factuality profiler

The goal of a factuality profiler is identifying the different degrees of factuality force associated with event expressions in text. This requires, first, recognizing the linguistic components that trigger that information, and second, organizing the different layers of information contributed and modeling their interaction.

The factuality profiler will take a document (or set of documents) as input, and it will apply the following steps:

1. Identification of the basic elements contributing to the event factuality layer: events, factuality markers, factuality sources, and temporal expressions.
2. For each event mentioned in the input text, assignment of its factuality profile (the set of factuality commitment acts).

Examples (31-34) below illustrate what the factuality profiler will return as output. In each of the sentences, the different basic elements, events and sources, are numbered and identified: events are in bold face whereas sources are underlined. Note that the author of the text is taken also as an (implicit) source,  $\mathbf{s}_0$ , and that the act of reporting the news is assumed to take place during the document creation time,  $\mathbf{t}_0$ . No other temporal expressions appear in the text, and thus some of the commitment acts will have their temporal reference underspecified. For now, the factuality values (expressed as the third element in each tuple) are just made up for the sake of the exemplification.

$$(31) \text{ a. } \underline{\text{They}}_{\mathbf{s}_1} \text{ **know**}_{\mathbf{e}_2} \text{ that al-Qaida and the Taliban **are not the real face**}_{\mathbf{e}_3} \text{ of political Islam.}$$

$$\text{ b. } \mathbf{e}_2: \textit{know} \quad \text{fp}(\mathbf{e}_2) = \{ \langle \mathbf{s}_0, \mathbf{t}_0, \text{FACT} \rangle \}$$

$$\mathbf{e}_3: \textit{are not the real face} \quad \text{fp}(\mathbf{e}_3) = \{ \langle \mathbf{s}_1\text{-}\mathbf{s}_0, -, \text{FACT} \rangle, \langle \mathbf{s}_0, \mathbf{t}_0, \text{FACT} \rangle \}$$

The formulas in (31b) give the factuality profiles for the events in (31a), denoted by *know* ( $\mathbf{e}_2$ ) and *are not the real face* ( $\mathbf{e}_3$ ).  $\mathbf{e}_2$  is asserted as a fact by  $\mathbf{s}_0$  (the author of the text) at time  $\mathbf{t}_0$ , the document creation time. On the other hand, there are two factuality commitment acts for  $\mathbf{e}_3$ , established through the SSP *know*: a first one performed by source  $\mathbf{s}_0$  at time  $\mathbf{t}_0$ , presupposing  $\mathbf{e}_3$  to be a fact, and a second one in which  $\mathbf{s}_1$  (according to  $\mathbf{s}_0$ , hence the embedded representation:  $\mathbf{s}_1\text{-}\mathbf{s}_0$ ) commits to the factuality of  $\mathbf{e}_3$  at an underspecified time.

- (32) a. Analysts<sub>s1</sub> **said**<sub>e1</sub> they<sub>s2=s1</sub> **knew**<sub>e2</sub> a peaceful solution **was in reach**<sub>e3</sub>.
- b. e1:*said*            fp(e1) = { <s<sub>0</sub>, t<sub>0</sub>, FACT> }  
       e2:*knew*            fp(e2) = { <s<sub>1-s0</sub>, -, FACT>, <s<sub>0</sub>, t<sub>0</sub>, uncom> }  
       e3:*was in reach*    fp(e3) = { <s<sub>1-s0</sub>, -, FACT>, <s<sub>0</sub>, t<sub>0</sub>, uncom> }

Sentence (32) is presented as a way of contrast with (31). In this second example, the event denoted by *was in reach* (e3) is also embedded by an SSP of knowledge (e2). Nevertheless, the two sources, s<sub>0</sub> and s<sub>1-s0</sub>, are now in disagreement about the factuality value of e3 because the knowledge predicate is embedded by the reporting verb *said*.

Another contrast pair is (33-34). In both cases, there is only one source (the author of the text) and therefore the sets of factuality commitment acts contain only one tuple each. Of interest here is the different factuality values of *allow* (e3) and *inspections* (e4), due to different embedding configurations.

- (33) a. The Royal Family will **continue**<sub>e1</sub> to **refuse**<sub>e2</sub> to **allow**<sub>e3</sub> detailed fire brigade **inspections**<sub>e4</sub> of their private quarters.
- b. e1:*continue*        fp(e1) = { <s<sub>0</sub>, t<sub>0</sub>, FACT> }  
       e2:*refuse*        fp(e2) = { <s<sub>0</sub>, t<sub>0</sub>, FACT> }  
       e3:*allow*         fp(e3) = { <s<sub>0</sub>, t<sub>0</sub>, COUNTERFACT> }  
       e4:*inspections*    fp(e4) = { <s<sub>0</sub>, t<sub>0</sub>, COUNTERFACT> }

- (34) a. The Royal Family will **continue**<sub>e1</sub> to **allow**<sub>e3</sub> detailed fire brigade **inspections**<sub>e4</sub> of their private quarters.
- b. e1:*continue*        fp(e1) = { <s<sub>0</sub>, t<sub>0</sub>, FACT> }  
       e3:*allow*         fp(e3) = { <s<sub>0</sub>, t<sub>0</sub>, FACT> }  
       e4:*inspections*    fp(e4) = { <s<sub>0</sub>, t<sub>0</sub>, POSSIBLE\_FACT> }

## References

- Asher, N. (1993). *Reference to Abstract Objects in English*. Dordrecht, The Netherlands: Kluwer Academic Press.
- Bach, K. & Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. Cambridge, Massachusetts, USA: The MIT Press.
- Ballmer, T. & Brennenstuhl, W. (1981). *Speech Act Classification. A study in the Lexical Analysis of English Speech Activity verbs*. Berlin, Heidelberg, New York: Springer-Verlag.

- Barwise, J. & Perry, J. (1983). *Situations and Attitudes*. Cambridge, Massachusetts, USA: The MIT Press.
- Bergler, S. (1992). *Evidential Analysis or Reported Speech*. PhD thesis, Brandeis University.
- Chafe, W. (1986). Evidentiality in english conversation and academic writing. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The Linguistic Coding of Epistemology*. Norwood, New Jersey, USA: Ablex Publishing Corporation.
- de Haan, F. (1997). *The Interaction of Modality and Negation: a Typological Study*. New York, USA: Garland.
- Dor, D. (1995). *Representations, Attitudes and Factivity Evaluations. An Epistemically-based Analysis of lexical Selection*. PhD thesis, Stanford University.
- Eggins, S. (1994). *An Introduction to Systemic Functional Grammar*. London, England: Pinter Publishers.
- Ginzburg, J. & Sag, I. A. (2000). *Interrogative Investigations*. Stanford, California, USA: CSLI.
- Halliday, M. A. K. (1985). *An introduction to functional grammar*. London, England: Edward Arnold.
- Hooper, J. B. (1975). On assertive predicates. In J. Kimball (Ed.), *Syntax and semantics, IV* (pp. 91–124). New York, USA: Academic Press.
- Jackendoff, R. & Culicover, P. W. (2003). The semantic basis of control. *Language*, 79(3), 517–556.
- Karttunen, L. (1970). Implicative verbs. *Language*, 47, 340–358.
- Karttunen, L. (1971). Some observations on factivity. *Papers in Linguistics*, 4, 55–69.
- Karttunen, L. (1973). Presuppositions of compound sentences. *Linguistic Inquiry*, 4(2), 169–193.
- Kiparsky, P. & Kiparsky, C. (1970). Fact. In M. Bierwisch & K. E. Heidolph (Eds.), *Progress in Linguistics. A Collection of Papers* (pp. 143–173). The Hague, Paris: Mouton.
- Koenig, J.-P. & Davis, A. R. (2001). Sublexical modality and the structure of lexical semantics. *Linguistics and Philosophy*, 24, 71–124.
- Light, M., Qiu, X. Y., & Srinivasan, P. (2004). The language of Bioscience: Facts, speculations, and statements in between. In *BioLINK 2004: Linking Biological Literature, Ontologies, and Databases*, (pp. 17–24).

- Nairn, R., Condoravdi, C., & Karttunen, L. (2006). Computing relative polarity for textual inference. In *Inference in Computational Semantics, ICoS-5*.
- Palmer, F. R. (1986). *Mood and Modality*. Cambridge, England: Cambridge University Press.
- Peterson, P. L. (1997). *Fact Proposition Event*. Dordrecht, The Netherlands: Dordrecht, Kluwer Academic Publishers.
- Pustejovsky, J., Castano, J., Ingria, R., Saurí, R., Gaizauskas, R., Setzer, A., & Katz, G. (2003). Timeml: Robust specification of event and temporal expressions in text. In *IWCS-5, Fifth International Workshop on Computational Semantics*.
- Vendler, Z. (1967). *Linguistics in Philosophy*. Ithaca, New York, USA: Cornell University Press.
- Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2).
- Wierzbicka, A. (1987). *English Speech Act Verbs. A Semantic Dictionary*. Sydney, Australia: Academoc Press.