Are You Sure That This Happened? Assessing the Factuality Degree of Events in Text

Roser Saurí*
Barcelona Media - Innovation Center

James Pustejovsky**
Brandeis University

Identifying the veracity, or factuality*, of event mentions in text is fundamental for reasoning about eventualities in discourse. Inferences derived from events judged as not having happened, or as being only possible, are different from those derived from events evaluated as factual. Event factuality involves two separate levels of information. On the one hand, it deals with polarity, which distinguishes between positive and negative instantiations of events. On the other, it has to do with degrees of certainty (e.g., possible, probable), an information level generally subsumed under the category of epistemic modality. This article aims at contributing to a better understanding of how event factuality is articulated in natural language. For that purpose, we put forward a linguistic-oriented computational model which has at its core an algorithm articulating the effect of factuality relations across levels of syntactic embedding. As a proof of concept, this model has been implemented in De Facto, a factuality profiler for eventualities mentioned in text, and tested against a corpus built specifically for the task, yielding an* F_1 *of 0.70 (macro-averaging) and 0.80 (micro-averaging). These two measures mutually compensate for an over-emphasis present in the other (either on the lesser or greater populated categories), and can therefore be interpreted as the lower and upper bounds of the De Facto's performance.*

1. Introduction

When we talk about situations in the world, we often leave pieces of information vague or try to complete the story with approximations, either because we do not know all the details or we are not sure about what we know. With a lesser or greater degree, this vagueness is pervasive in all types of accounts, regardless of the topic and the degree of proximity of the speaker with the facts being reported: our last family gathering, what we read about the tsunami and its aftermath in Japan, our perspective on a particular topic, or how we feel today. Even in scientific discourse, findings tend to be expressed with degrees of cautiousness.

* Voice and Language Group, Barcelona Media - Innovation Center, Diagonal 177, 08018 Barcelona, Catalonia. E-mail: roser.sauri@barcelonamedia.org.
** Computer Science Department, Brandeis University, 415 South Street, Waltham MA 02454, USA. E-mail: jamesp@cs.brandeis.edu.

Submission received: 4 April 2011; revised submission received: 1 October 2011; accepted for publication: 30 November 2011.

© 2012 Association for Computational Linguistics

The linguistic mechanisms for coping with the vagueness and fuzziness in our knowledge are commonly referred to as **speculative language**. This involves different levels of grammatical manifestation, most significantly quantification over entities and events, modality, and hedging devices of a varied nature. We can be vague or approximate with the temporal and spatial references of situations in the world, when quantifying the frequency of usual events, assessing the number of participants involved, and describing or adscribing them into a class. We also qualify our statements with approximative language when giving an opinion, or when we are not certain about the degree of veracity of what we are telling.

The present article focuses on a particular kind of speculation in language, specifically, that concerning the factuality status of eventualities mentioned in discourse. Whenever we talk about situations, we express our degree of certainty about their factual status. We can characterize them as an unquestionable fact, or qualify them with some degree of uncertainty if we are not sure whether the situation holds, or will hold, in the world.

Identifying the factuality status of event mentions is fundamental for reasoning about eventualities in discourse. Inferences derived from events judged as not having happened, or as being only possible, are different from those derived from events evaluated as factual. Event factuality is also essential for any task involving temporal ordering, because the plotting of event mentions into a timeline requires different actions depending on their veracity. Karttunen and Zaenen (2005) discuss its relevance for information extraction, and in the area of textual entailment, factuality-related information (modality, intensional contexts, etc.) has been taken as a basic feature in some systems participating in the PASCAL RTE challenges (e.g., Hickl and Bensley 2007). The need for this type of information is also acknowledged in the annotation schemes of corpora devoted to event information, such as the ACE corpus for the Event and Relation recognition task (e.g., ACE 2008), or TimeBank, a corpus annotated with event and temporal information (Pustejovsky et al. 2006).

Significantly, in the past few years this level of information has been at the focus of much research within the NLP area dedicated to the biomedical domain. Distinguishing between what is reported as a fact versus a possibility in experiment reports or in patient health records is a crucial capability for any robust information extraction tool operating on that domain. This interest has resulted in the compilation of domain-specific corpora devoted particularly to that level of information, such as BioScope (Vincze et al. 2008), and others that include event factivity as a further attribute in the annotation of biomedical events, such as GENIA (Kim, Ohta, and Tsujii 2008). Furthermore, factuality-related information was the main focus in the CoNLL-2010 shared task on Learning to Detect Hedges and their Scope in Natural Language Text (Farkas et al. 2010), and the topic in a subtask of the BioNLP'09 and BioNLP'11 shared task editions on Event Extraction (Kim et al. 2009),[1] dedicated to predict whether the biological event is under negation or speculation.

The overall goal of this article is to contribute to a better understanding of this particular aspect of speculation. We analyze all the ingredients involved in computing the factuality nature of event mentions in text, and put forward a computational model based on that. As a proof of concept, the model is implemented into De Facto, a factuality profiler, and its performance tested against FactBank, a corpus annotated

1 For the 2011 edition, refer to: `https://sites.google.com/site/bionlpst/`.

with factuality information built specifically for the task and currently available to the community through the Linguistic Data Consortium (Saurí and Pustejovsky 2009a).

The article begins by defining event factuality and its place in speculative language (Section 2). The basic components for the model on event factuality are presented in Section 3, and the algorithm integrating these is introduced in Section 4. Section 5 reports on the experiment resulting from implementing the proposed model into De Facto, and Section 6 relates the present work to other research in the field.

2. Event Factuality and Speculative Language

2.1 Defining Event Factuality

Event factuality (or **factivity**) is understood here as the level of information expressing the factual nature of eventualities mentioned in text. That is, expressing whether they correspond to a fact in the world (Example (1a)), a possibility (Examples 1b, 1c), or a situation that does not hold (Example 1d), as is the case with the events denoted by the following underlined expressions:[2]

(1) a. Har-Shefi regretted <u>calling</u> the prime minister a traitor.

b. Results indicate that Pb2+ <u>may inhibit</u> neurite initiation.

c. Noah's flood <u>may have not been</u> as biblical in proportion as previously thought.

d. Albert Einstein <u>did not win</u> a Nobel prize for his theories of Relativity.

The fact that an eventuality is depicted as holding or not does not mean that this is the case in the world, but that this is how it is characterized by its informant. Similarly, it does not mean that this is the real knowledge that informant has (his true cognitive state regarding that event) but what he wants us to believe it is.

Event factuality rests upon distinctions along two different parameters: the notions of certainty (what is certain vs. what is only possible) and polarity (positive vs. negative). In some contexts, the factual status of events is presented with absolute certainty. Then, depending on the polarity, events are depicted as either situations that have taken or will take place in the world (here referred to as **facts**; see Example (1a)), or situations that do not hold in the world (here called **counterfacts**; see Example (1d)). In other contexts, events are qualified with different shades of uncertainty. Combining that with polarity, events are seen as **possibly factual** (Example (1b)) or **possibly counterfactual** (Example (1c)).[3]

2 In this article, the terms *event* and *eventuality* will be used in a very broad sense to refer to both processes and states, but also other abstract objects such as situations, propositions, facts, possibilities, and so on. Furthermore, events in the examples will be identified by marking only their verb, noun, or adjective head, together with their modal and negation particles when deemed necessary. This follows the convention assumed in TimeML, a specification language for representing event and time information (Pustejovsky et al. 2005).

3 The term *counterfactual* has a long tradition in philosophy of language and linguistics, where it refers to conditional (or if–then) statements expressing what would be the case if their antecedent was true, although it is not. For example: *If Gandhi had survived the fatal gun attack, he would have continued working for a better world.* Here, however, we extend its use to refer to negated events in general. One can argue that negated events are facts as well. For example, it is a fact that Gandhi did not survive the fatal gun attack. The term *counterfact* must be understood here as *negative fact*.

Factuality is expressed through a complex interaction of many different aspects of the overall linguistic expression. It involves explicit polarity and modality markers, but also lexical items, morphological elements, syntactic constructions, and discourse relations between clauses or sentences.

Polarity particles, which convey the positive or negative factuality of events, include elements of a varied nature: adverbs (*not, neither, never*), determiners (*no, non*), pronouns (*none, nobody*), and so forth. At another level, **modality particles** contribute different degrees of certainty. In English, they can be realized as verbal auxiliaries (*must, may*), adverbials (*probably, presumably*), and adjectives (*likely, possible*). All these categories display an equivalent gradation of modality (Givón 1993).

In many cases, the factuality of events is conveyed by what we refer to as **event-selecting predicates** (ESPs), that is, predicates (either verbs, nouns, or adjectives) that select for an argument denoting an event of some sort. ESPs are of interest here because they qualify the degree of factuality of their embedded event, which can be presented as a fact in the world (Example (2)), a counterfact (Example (3)), or a possibility (Example (4)). In these examples, the ESPs are in boldface and their embedded events are underlined.

(2) a. Some of the Panamanians **managed** [to <u>escape</u> with their weapons].

b. The defendant **knew** that [he <u>had been</u> in possession of narcotics].

(3) a. 1,200 voters **were prevented** from [<u>casting</u> ballots on election night].

b. The manager **avoided** [<u>returning</u> the phone calls].

(4) a. I **think** [they <u>voted</u> last weekend].

b. Hawking **speculated** that [most extraterrestrial life <u>would be similar</u> to microbes].

Absolute factuality is conveyed by ESPs belonging to classes fairly well studied in the literature, such as: implicative (Example (2a)) (Karttunen 1970); factive (Example (2b)) (Kiparsky and Kiparsky 1970); perception (e.g., **see** *a car explode*); aspectual (e.g., **finish** *reading*), and change-of-state predicates (e.g., **increase** *its exports*). Counterfactuality is brought about by other implicative predicates, like *avoid* and *prevent* (Example (3)) (Karttunen 1970), whereas predicates such as *think, speculate*, and *suspect* qualify their complements as not totally certain (Example (4)) (Hooper 1975; Bach and Harnish 1979; Dor 1995). The group of ESPs that leave the factuality of their event complement underspecified is also significant. The event is mentioned in discourse, but no information is provided concerning its factual status. Several predicate classes create this effect, for example: volition (e.g., *want, wish, hope*), commitment (*commit, offer, propose*), and inclination predicates (*willing, ready, eager, reluctant*), among others (cf. Asher 1993).

Other information at play is evidentiality (e.g., a *seen* event is presented with a factuality degree stronger than that of an event *reported* by someone else), and mood (e.g., indicative vs. subjunctive). Factuality information is also introduced by certain syntactic constructions involving subordination. In some cases, the embedded event is presupposed as a fact, as in non-restrictive relative clauses (Example (5a)) or participial clauses (Example (5b)). In others, like purpose clauses, the event is intensional and thus presented as underspecified (Example (5c)).

(5) a. Obama, [who took office in January], inherited a budget deficit of $1.3 trillion.

b. [Having revolutionized linguistics], Chomsky moved to political activism.

c. Stronach resigned as CEO of Magna [to seek a seat in Canada's Parliament].

Finally, a further means for conveying factuality information is available at the discourse level. Some events may first have their factual status characterized in one way, but then be presented differently in a subsequent sentence.

2.2 Notions Connected to Event Factuality

Event factuality results from the interaction between polarity and certainty. Here we review the connections of these two notions with other ones in the study of language.

Certainty. The axis of certainty is related to **epistemic modality**, a category dealing with the degree of certainty of situations in the world. Epistemic modality has been studied from both the logical and linguistic traditions. Within linguistics, authors from different traditions converge in analyzing modality as a subjective component of discourse (e.g., Lyons 1977; Chafe 1986; Palmer 1986; Kiefer 1987), a view that is adopted in the present analysis.[4] Traditionally, the study of epistemic modality in linguistics has been confined to modal auxiliaries (e.g., Palmer 1986), but more recently a wider view has been adopted which includes other parts of speech as well, such as epistemic adverbs, adjectives, nouns, and lexical verbs (e.g., Rizomilioti 2006).

In a more secondary way, the axis of certainty is also related to the system of **evidentiality**, concerned with the way in which information about situations in the world is acquired, such as directly experienced, witnessed, heard-about, inferred, and so on (van Valin and LaPolla 1997; Aikhenvald 2004). Different types of evidence have an effect on the way the factuality of an event is evaluated. For instance, something reported as seen can more easily be assessed as a fact than something reported as inferred.

Certainty touches as well on the notion of **epistemic stance**, developed from a more cognitivist perspective and which is defined as the pragmatic relation between speakers and their knowledge regarding the things they talk about (Biber and Finegan 1989; Mushin 2001). Similarly, within Systemic Functional Linguistics, the Appraisal Framework develops a taxonomy of the mechanisms employed for expressing subjective information such as attitude, its polarity, graduation, and so forth (Martin and White 2005).

Within NLP, most work on uncertainty and speculative information has been approached from a hedging-based perspective. The notion of **hedging** is initially defined by Lakoff (1973, page 471) as "words whose job is making things fuzzier or less fuzzy." In particular, he uses this term to analyze linguistic constructions that express degrees of the *is_a* relationship (e.g., *is a sort of, in essence/strickly speaking... is...*). Due to the fuzziness aspect of hedges, subsequent work extends the notion to include expressions for qualifying the degree of commitment of the writer with respect to what is asserted (Hyland [1996], among others). By this definition, hedging and event factuality seem to be overlapping concepts. They differ on the extent of the phenomena they

4 This is different, however, from most of the work within truth-conditional semantics, which conceives modality as independent from the speaker's perspective (e.g., Kratzer 1991).

each cover, however. First, hedging is confined only to partial degrees of uncertainty, whereas factuality includes also the levels of absolute certainty. Second, in addition to degrees of writer's commitment towards the veridicity of her statements, hedging (but not factuality) encompasses speculative expressions belonging to other scales, most significantly, expressions of usuality (to quantify the frequency of events: *often, barely, tends to*, etc.), expressions of category membership (i.e., *is_a* downgraders, such as *is a sort of*, presented by Lakoff [1973]), as well as lack of knowledge (e.g, *little is known*).

Polarity. The second axis configuring event factuality is the system of polarity, so called because it articulates the polar opposition between positive and negative contexts. Due to its recent adoption in the NLP area of sentiment analysis, the term polarity is often taken to express only the direction of an opinion. Here, we use the term in its original grammatical sense, that is, as conveying the distinction between affirmative and negative contexts (e.g., Horn 1989). Being more abstract, this definition encompasses the different facets of the positive/negative opposition, and not only the one that is relevant in opinion mining.

2.3 Key Elements in the Factuality System

Identifying event factuality in text poses challenges at different levels of analysis. We explore them in the current section.

A scale of factuality degrees. Concerning distinctions at the level of both polarity and certainty (or modality, as is more commonly referred to within linguistics), the factuality of events can be characterized as a double-axis scale. Figure 1 illustrates the system.

The axis of polarity defines a binary distinction (positive vs. negative), and the axis of modality conveys certainty as a continuous scale that ranges from truly certain to completely uncertain, passing through a whole spectrum of shades that languages accommodate in different ways, depending on the grammatical resources they have available. For example, assuming only a limited number of words in English, one can create the following distinctions: *improbable, slightly possible, possible, fairly possible, probable, very probable, most probable, most certain, certain*.

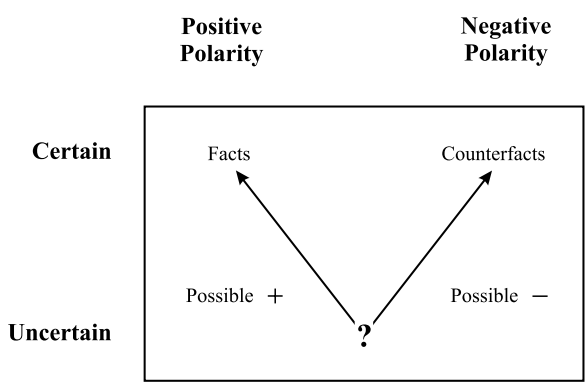

Figure 1
The double range of factuality.

This continuum poses a challenge in the setting of a model of factuality with potential cross-linguistic validity. Many linguists agree, however, that speakers are able to map areas of the modality axis into discrete values (Lyons 1977; Horn 1989; de Haan 1997). The goal is therefore identifying the factuality distinctions that reflect our linguistic intuitions as speakers, and that can also help define a set of sound and stable criteria for differentiating among them. The factual value of markers such as *possibly* and *probably* is fairly transparent. What, however, is the contribution of elements like *think*, *predict*, *suggest*, or *seem*?

Interactions among factuality markers. The factuality status of a given event cannot be determined from the strictly local modality and polarity operators scoping over that event alone; rather, if present, other non-local markers must be considered as well to obtain the adequate interpretation. Consider:

(6) a. Several EU member states will **continue** to **allow** passengers to **carry** duty-free drinks in hand luggage.

b. Several EU member states will **continue** to **refuse** to **allow** passengers to **carry** duty-free drinks in hand luggage.

c. Several EU member states **may refuse** to **allow** passengers to **carry** duty-free drinks in hand luggage.[5]

In all three examples above the event *carry* is directly embedded under the verb *allow*, but receives a different interpretation depending on the elements scoping over that. In Example (6a), where *allow* is embedded under the factive predicate *continue*, *carry* is characterized as a fact in the world. Example (6b), on the other hand, depicts it as a counterfact because of the effect of the predicate *refuse* scoping over *allow*, and finally, Example (6c) presents it as uncertain due to the modal auxiliary *may* qualifying *refuse*.[6]

Any treatment aiming at adequately handling the contents of sentences like these needs to incorporate the notion of scope in its model, but scope is not enough. As these data show, the factuality value of an event does not depend on the element immediately scoping over it. Neither does it rely on the meaning resulting from some sort of additive (or concatenative) operation among all the markers. In Example (6b), for example, two of the factuality markers that include the event *carry* in their scope (*continue* and *refuse*) typically mark contradictory information. The first one presupposes the factuality of the event it scopes over, and the second negates it. Which should be the resulting factuality value for *carry* if only scope information is used?

Factuality as a property qualifying events and not the whole sentence. Factuality is a property that qualifies the nature of events, hence operating at a level of units smaller than sentences. Frequently sentences express more than one event (or proposition), each of them qualified with a different degree of certainty. Consider Example (7),[7] where the main event *have an easier time* (e_3) is depicted as a possibility in the world, event *crossover*

5 The original sentence in this set is (6b) (`http://www.irishtimes.com/newspaper/ireland/2011/0502/1224295867753.html`). The other two have been adapted for the argument's sake.

6 The verb *allow* is generally used as a two-way implicative predicate, that is, as a predicate that holds a direct relation between its truth (or falsity) and that of its embedded event (Karttunen 1970).

7 Extracted from Rubin (2006, page 59).

voting being barred (e_2) is asserted as a fact, and event *crossover voting* (e_1) is uncertain—that is, the fact that it is barred does not mean that it does not take place.

(7) In future primaries, where crossover **voting**$_{e_1}$ is **barred**$_{e_2}$, Bush may well **have**$_{e_3}$ an easier time.

Facts and their sources. Certain event components, such as the temporal reference or the participants taking part in it, are inherent elements of any given event. For example, the visit to the zoo with Max in April, Ivet in August, and Arlet in December are three separate events, given the difference in participants and temporal location. By contrast, factuality is a matter of perspective. Different sources can have divergent views about the factuality of the very same event. Recognizing this is crucial for any task involving text entailment. Event e in Example (8), for instance (i.e., Ruby being the niece of the Egyptian president), will be inferred as a fact in the world if it cannot be qualified as having been asserted by a specific source, here Berlusconi (underlined).

(8) Berlusconi said that Ruby **was**$_e$ the niece of Egyptian President Hosni Mubarak.

By default, events mentioned in discourse always have an implicit source, namely, the author of the text. Additional sources are introduced in discourse by means of ESPs such as *say* or *pretend*:

(9) Nelles **said**$_{e_1}$ that Germany has been **pretending**$_{e_2}$ for long that nuclear power **is safe**$_{e_3}$.

In some cases, the different sources relevant for a given event may coincide with respect to its factual status, but in others they may be in disagreement. In Example (9), for instance, event e_3 (*nuclear power being safe*) is assessed as a fact according to Germany but as a counterfact according to Nelles, whereas the text author remains uncommitted.

The time variable. It is not only the case that two participants can present different views about the same event, but also that the same (or different) participant presents a diverging view at different points in time. Consider:

(10) a. *In mid-2001*, Colin Powell and Condoleezza Rice both publicly denied that Iraq **had weapons of mass destruction**.

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

A model of event factuality needs therefore to be sensitive to the distinctions in perspective brought about by sources and temporal references. Only under this assumption is it possible to account for the potential divergence of opinions on the factual status of events, as is common in news reports.

3. Towards a Model for Event Factuality

Having identified the main aspects involved in event factuality, we explore the interplay among these elements, and subsequently build a model that can explain these interactions. Based on the structure of linguistic expressions, this model will assume an event-

centered approach in order to tackle the factuality nature of each event independently of the others mentioned in the same sentence. Factuality distinctions are established at a fine-grained level, and multiple perspectives on the same event are accounted for by means of the notion of source as a participant introduced by predicates of report, knowledge, belief, and so on. We begin by introducing the notion of a *factuality profile* (Section 3.1), and then formalize the basic components that have a role in it, namely: *factuality values* (Section 3.2), *sources* (Section 3.3), and *factuality markers* (Section 3.4). The algorithm putting all these ingredients together will be presented in Section 4.

3.1 The Factuality Profile of Events

Whenever speakers talk about events, they qualify them with a degree of factuality. Here, we refer to this act of assigning a factuality value to a given event performed by a particular source at a specific point in time as a **factuality commitment act**. This involves four components:

- The event in focus, `e`.
- The factuality value assigned to that event, `f`, which touches on both polarity and epistemic modality distinctions as encoded in factuality markers.
- The source assigning the factuality value to that event, `s`.
- The time when the factuality value assignment takes place, `t`.

For instance, in Example (9) Germany is presented as defending that nuclear power is safe (event e_3). This corresponds to the factuality commitment act that assesses event e_3 as a fact in the world, performed by source *Germany* at an underspecified point in time t_1.

Given that events in discourse can be evaluated by more than one source and at several points in time, the factuality of each event can be characterized through more than one factuality commitment act. We define the set of factuality commitment acts associated to an event as its **factuality profile**. Formally, the factuality profile of a given event e, p_e, can be represented as follows:

$$p_e = \{\langle s, t, f\rangle \mid s \text{ is a relevant source for event } e \text{ \& } t \text{ is a point in time \& } f \text{ is the factuality value assigned by source } s \text{ to event } e \text{ at point in time } t\} \quad (1)$$

Using example (9) again, the factuality profile of event e_3 (nuclear power being safe) contains three factuality commitment acts: one by source *Germany*, who commits to the veradicity of the event, another for source *Nelles*, who disagrees, and finally another for the author, who keeps an underspecified position.

$$p_{e_3} = \{\langle germany, t_1, fact\rangle, \langle nelles, t_2, negative_fact\rangle, \langle author, t_3, underspecified\rangle\} \quad (2)$$

The model that will be presented here for determining the factuality profiles of events in text will disregard the temporal component and focus only on identifying relevant sources and factuality values.

3.2 How Certain Are You: Factuality Values

The values for characterizing event factuality must account for distinctions along both the polarity and the modality axes. Whereas polarity is a binary system with the values positive and negative, epistemic modality constitutes a continuum ranging from uncertain to absolutely certain. In order to obtain consistent annotation for informing and evaluating automatic systems, a discrete categorization of modality that effectively reflects the main distinctions applied in natural languages is desirable.

Within modal logic two operators are typically used to express modal contexts: necessity ($\Box$) and possibility ($\Diamond$). Most linguists, however, agree that this is inadequate to capture the richness of cross-linguistic data. It has generally been observed that, even though modality is a continuous system, a three-fold distinction is commonly adopted by speakers (e.g., Lyons 1977; Palmer 1986; Halliday and Matthiessen 2004). Horn (1989) analyzes modality and its interaction with polarity based on both linguistic tests and the logical relations holding at the basis of the Aristotelian Square of Opposition (in particular, the Law of Excluded Middle and the Law of Contradiction). In Horn's work, the system of epistemic modality is analyzed as a particular instantiation of scalar predication, that is, as a collection of predicates P_n such as $\langle P_j, P_{j-1}, ..., P_2, P_1 \rangle$, where P_n outranks (i.e., is stronger than) P_{n-1} on the relevant scale. The relations holding among predicates of the same scalar predication are manifested in syntactic contexts like the following (Horn 1972):

- Contexts with the possibility open that a higher value on the relevant scale obtains:
  - *(at least)* P_{n-1}*, if not (downright)* P_n.
  - P_{n-1}*, {or/ and possibly} even* P_n.
- Contexts by which a higher value in the scale is known to obtain:
  - P_{n-1}*, {indeed/ in fact/ and what is more}* P_n.
  - *not only* P_{n-1} *but* P_n.

This set of contexts allows him to conclude the existence of two independent epistemic scales that differ in quality (positive vs. negative polarity):[8]

(11) a. ⟨*certain, likely* (*probable*), *possible*⟩

b. ⟨*impossible, unlikely* (*improbable*), *uncertain*⟩

Based on Horn's distinctions, we divide the modality axis into the values *certain* (CT), *probable* (PR), and *possible* (PS), and the polarity axis into *positive* (+) and *negative* (−). Moreover, we add an *underspecified* value in both axes to account for cases of non-commitment of the source or in which the value is not known. A degree of factuality is then characterized as a pair ⟨MOD, *pol*⟩, containing a modality and a polarity value (e.g., ⟨CT,+⟩). For the sake of simplicity, these will be represented in the abbreviated form of: MOD*pol* (e.g., CT+). Table 1 presents the full set of factuality values.

8 The beauty of the system can be appreciated when mapped to the traditional Square of Opposition, used to account for the interaction between negation and quantifiers or modal operators (Horn [1989], following Aristotle). For a detailed account of that within the current framework, see Saurí and Pustejovsky (2009b).

Table 1
Factuality values.

	Positive	Negative	Underspecified
Certain	CT+ (factual)	CT− (counterfactual)	CTu (certain but unknown output)
Probable	PR+ (probable)	PR− (not probable)	[NA]
Possible	PS+ (possible)	PS− (not certain)	[NA]
Underspecified	[NA]	[NA]	Uu (unknown or uncommitted)

The table includes six fully committed (or specified) values (CT+, CT−, PR+, PR−, PS+, PS−), and two underspecified ones: the partially underspecified CTu, and the fully underspecified Uu. The use of the fully committed values should be clear from the paraphrases in the table, but uncommitted values deserve further explanation. The partially underspecified value CTu is for cases where the source has total certainty about the factual nature of the event but it does not commit to its polarity. This is the case of source *John* regarding event *e* in: *John knows whether Mary* $\underline{came_e}$. The fully underspecified value Uu, on the other hand, is used when any of the following situations applies:

(*i*) The source does not know the factual status of the event (e.g., *John does not know whether Mary* $\underline{came_e}$).

(*ii*) The source is not aware of the possibility of the event (e.g., *John does not know that Mary* $\underline{came_e}$).

(*iii*) The source does not overtly commit to the event (e.g., *John didn't say that Mary* $\underline{came_e}$).[9]

3.3 Who Said What: Factuality Sources

Sources are understood here as the cognitive individuals that hold a specific stance regarding the factuality status of events in text. They correspond to one of the following actor types:

Text author. Events mentioned in discourse always have a default source, which corresponds to the author of the text (speaker or writer).

Other sources. Contexts of report, belief, knowledge, inference, and so forth (created by predicates like *say, think*, *know*, *see*) introduce additional explicit sources, generally expressed by the logical subject of the predicate. Similarly, impersonal constructions (e.g., *it seems, it is clear, ...*) or passive constructions with no agentive argument (e.g., *it is expected*) introduce an implicit source which can be rephrased as *everybody* or *somebody*, among similar expressions. The factuality of the embedded event is assessed relative to this new (explicit or implicit) source, as well as to any source already present in the discourse, such as the text author.

In the current framework, these sources will be formally represented as: s_0 (author source), s_n for $n > 0$ (explicit source), and GEN (for implicit, generic source).

9 The value Uu could be seen as equivalent to others, such as PS− and PR−. Note, however, that in these two, but not in Uu, the source commits to a specific degree of uncertainty (possible or probable, respectively), as in *John said that Mary [may have not come]*, and *John said that [Mary has probably not come]*.

"Source" as a technical term. Although the term *source* is generally used as a synonym of *informant*, in the scope of the current work it is used in a very specific, technical sense. First, it not only refers to the typical informants, that is, those participants actively committing to the factuality of an event by means of a speech act or a writing event of some sort (e.g., *Mary says/claims/wrote...*), but also to those that are presented as holding (or being able to hold) a position about the factuality of that event—be it because they hold a mental attitude about the situation (*Mary knows/learned/thinks/suspects that...*), because they are the experiencers of a psychological reaction generated by the event in question (*Mary regrets/is sad that...*), or because they are presented as witnesses or perceivers of the situation (*Mary saw/heard that...*).

Second, the notion of source as used here includes participants that are presented as unaware of the relevant event as well. Consider:

(12) Galbraith is claiming that President Bush was unaware that there were two major sects of Islam just two months before the President ordered troops to invade Iraq.

A complete analysis of the facts, causes, and consequences regarding the war in Iraq needs to include the existence of two major sects of Islam, and what this means in terms of the potential stability of the area. But it should also include that President Bush did not know this piece of information beforehand, as claimed by the political actor Galbraith. Thus, the factuality analysis of the sentence must include President Bush as a source who at some point in time held an uncommitted factuality stance with regard to the existence of these two Islamic sects.

Nested sources. The status of the author is, however, different from that of the additional sources. The reader does not have direct access to the factual assessments made by these new sources, but only according to what the author asserts. Thus, we need to appeal to the notion of *nested source* as presented in Wiebe, Wilson, and Cardie (2005). That is, Nelles in Example (13) is not a licensed source of the factuality of event e_2, but Nelles according to the author, represented here as *nelles_author*.[10] Similarly, the source referred to as Germany corresponds to the chain: *germany_nelles_author*.

(13) Nelles **said**$_{e_1}$ that Germany has been **pretending**$_{e_2}$ for long that nuclear power **is safe**$_{e_3}$.

Source roles. We distinguish between two different source roles. Sources most immediately committed (or uncommitted, in the case of unaware sources) to the factuality status of an event perform the role of **cognizers** of that event. This is typically the case of sources introduced in contexts of report, witnessing, belief, and so forth. On the other hand, sources that present (or anchor) the factuality commitment of the cognizer towards an event are referred to as the **anchors**. The roles of cognizer and anchor are relative to each event. For instance, in Example (13) the cognizer of event e_2 (*Germany pretending*) is Nelles (according to the author, hence: *nelles_author*) and its anchor is the text author. On the other hand, the cognizer of event e_3 (*nuclear power being safe*) is

10 This is equivalent to the notation ⟨*author, nelles*⟩ in Wiebe's work. Here, we adopt a reversed representation of the nesting (i.e., the non-embedded source last) because it positions the most direct source of the event at the outmost layer, thus facilitating its reading.

Table 2
Polarity value given contextual polarity.

Marker value	Context polarity		
	+	**−**	**u**
+	+	−	u
−	−	+	u

Germany (based on what the author claims that Nelles says, thus: *germany_nelles_author*), and its anchor is Nelles (*nelles_author*).[11] Event e_1 (*Nelles saying*) is directly affirmed by the author, and so the distinction between cognizer and anchor at this level is irrelevant.

3.4 Expressing Factuality in Text: Factuality Markers

Event factuality is conveyed by means of explicit polarity and modality-denoting expressions of a wide variety. Section 2.1 gave a brief introduction to the main types (namely, polarity and modality particles, the ESPs and syntactic constructions), and Section 2.3 illustrated the natural interplay that takes place among them in the context of a sentence. In the current section we organize the factuality-relevant information present in lexical and syntactic structures so that it can be used by a model capable of accounting for the interaction of information across levels of embedding. The focus is on English data, but the information is easily applicable to other languages, such as those in the Romance and Germanic families.[12]

Here and in the following sections, we understand the notion of *context of a factuality marker* as the level of scope most immediately embedding it. For instance, the context of the polarity particle *never* in Example (14) (subsequent paragraph) is set by the main clause.

3.4.1 Polarity Particles. Polarity particles of negation (from the adverb *not* to pronouns like *nobody*) switch the original polarity of its context (cf. Polanyi and Zaenen 2006): If it is positive, the presence of a marker of negative polarity switches it to negative, and vice versa. Nothing changes if the original context is underspecified. For instance, in Example (14a) the context of the polarity particle *never* is positive, and so the resulting polarity for event *train* is negative, as opposed to what happens in Example (14b). In Example (14c) the contextual polarity is underspecified, and so is the factuality value for event *train*.

(14) a. It is the case that [$_{context:CT+}$ John *never* $\underline{\text{trains}}_e$]. (*train*$_e$: CT−)

b. It is **not** the case that [$_{context:CT-}$ John *never* $\underline{\text{trains}}_e$]. (*train*$_e$: CT+)

c. It is **unknown** whether [$_{context:Uu}$ John *never* $\underline{\text{trains}}_e$]. (*train*$_e$: Uu)

Table 2 models the interaction between contextual polarity (columns) and the polarity value contributed by a new marker (rows).

11 Therefore, a source performing the cognizer role for one event can be the anchor source of another.
12 As a matter of fact, we plan to port it to Catalan and Spanish in the near future.

Table 3
Modality value given contextual factuality.

	Contextual factuality											
	Polarity = +				Polarity = −				Polarity = u			
Marker	**CT**	**PR**	**PS**	**U**	**CT**	**PR**	**PS**	**U**	**CT**	**PR**	**PS**	**U**
CT	CT	PR	PS	Uu	PS	PR	PS	Uu	CT	PR	PS	Uu
PR	PR	PR	PS	Uu	PR	PR	PS	Uu	PR	PR	PS	Uu
PS	PS	PS	PS	Uu	CT	PR	PS	Uu	PS	PS	PS	Uu
U	Uu	Uu	Uu	Uu	Uu	Uu	Uu	Uu	Uu	Uu	Uu	Uu

3.4.2 Particles of Epistemic Modality. The following are some of the most common modality particles, paired with the factuality value that they express.[13]

(15)

Value		**Adverbs**	**Adjectives**	**Auxiliaries**
CT+	:	*certainly, necessarily*	*certain, necessary, sure*	
PR+	:	*apparently, probably*	*apparent, likely, probable*	
PS+	:	*possibly, presumably, seemingly*	*possible, presumed, hypothetical*	*can, could; may, might*
CT−	:		*impossible*	
PR−	:		*improbable, unlikely*	
PS−	:		*uncertain, unsure*	
Uu	:	*reportedly, supposedly*	*reported, supposed*	*must, should*

A modality particle, however, does not necessarily color the event it scopes over with its inherent modal value. The factuality value projected to that event depends on the interaction between the particle on the one hand, and the modality and polarity of its context, on the other. Consider:

(16) a. Koenig **denies** [$_{context:CT-}$ that Freidin *may* have left$_e$ the country]. (*left*$_e$: CT−)

b. Koenig **suspects** [$_{context:PR+}$ that Freidin *may* have left$_e$ the country]. (*left*$_e$: PS+)

In Example (16a), *may* is used in a context of negative polarity and absolute certainty (CT−) set by *deny*, whereas in Example (16b), it is used in a context of positive polarity and probable modality (PR+) set by *suspect*. As a result, in the first example, event *e* is presented as a counterfact according to Koenig (CT−), but as a possibility in the second (PS+).

Table 3 illustrates the interaction between the polarity and modality values from the context (columns) and the modal value contributed by the marker (rows).[14] Note that the resulting values do not specify polarity information, except for the contexts where contextual modality or polarity is underspecified (columns 4, 8, and 12, and last row), where the resulting polarity is *u* (underspecified). In all other cases, the polarity contributed by the marker will interact with that from the context as specified in Table 2. That is, positive contextual polarity will respect the original polarity denoted by the marker, whereas negative polarity will switch it. For instance, the marker *impossible*, which has an inherent value of CT−, in a negative context will express PS+ (e.g., *it is not*

13 Modal auxiliaries in English can express different types of modality (e.g., epistemic or deontic). Disambiguating among the possible interpretations of the same auxiliary is a goal beyond the scope of the current research.

14 It has been compiled by exploring corpus data as well as made-up examples. Combinations with mid values (probability) are highly unusual; the resulting values are only estimated.

impossible that...). The reader can use Table 3 to verify the interactions between *deny* and *may* in Example (16a) (corresponding to the value in column 5, row 3), and *suspect* and *may* in Example (16b) (column 3, row 2).

3.4.3 Event-Selecting Predicates (ESPs). As presented earlier, ESPs are predicates with an event-denoting argument (for instance, predicates of report, knowledge, belief, or volition). As part of their meaning, they qualify the factuality nature of that event. Here, we distinguish between two kinds of ESPs: those introducing a new source in discourse, referred to as **Source Introducing Predicates** (SIPs), and those that do not, called **Non-Source Introducing Predicates** (NSIPs).

Source Introducing Predicates (SIPs). The additional source they contribute tends to correspond to their logical subject. They typically belong to one of the following classes:

(a) Predicates of report; for example, *say, add, claim, write, publish*.

(b) Predicates of knowledge: *know, remember, learn, discover, forget, admit*.

(c) Predicates of belief and opinion: *think, consider, guess, predict, suggest*.

(d) Predicates of doubt: *doubt, wonder, ask*.

(e) Predicates of perception: *see, hear, feel*.

(f) Predicates expressing proof: *prove, show, support, explain*.

(g) Predicates expressing some kind of inferencing process: *infer, conclude, seem* (as in: *it seems that*).

(h) Predicates expressing some psychological reaction as a result of an event or situation taking place: *regret, be glad (that)*.

As part of their lexical semantics, SIPs express the factuality value that both the new source they introduce (that is, the cognizer) as well as the anchor, assign to their event-denoting complement. Compare the following examples built with two different SIPs: *know* and *say*. For each sentence, the columns *anchor* and *cognizer* display the factual values that these two sources assign to the embedded event *e* (underlined).

(17)				**anchor**	**cognizer**
	a. *The client*$_s$	**knew**	that his father <u>had been killed</u>$_e$.	CT+	CT+
	b. ...	**said**	...	Uu	CT+

By using the SIP *know* (Example (17a)), the anchor (here the text author) is positioning himself as agreeing with the client (the cognizer) in considering that his father had been killed. On the other hand, by using the SIP *say* (Example (17b)) the anchor remains uncommitted. Distinctions of this kind are fundamental for any task requiring perspective identification. SIPs can therefore be characterized and grouped according to the configuration in the factuality assignments performed by anchor and cognizer. Notice that none of the SIPs in the following list has the same factual configuration.

(18)		**anchor**	**cognizer**		**anchor**	**cognizer**
	say:	Uu	CT+	*know:*	CT+	CT+
	deny:	Uu	CT−	*forget:*	CT+	Uu
	suppose:	Uu	PR+	*pretend:*	CT−	CT+

Table 4
Lexicon fragment for SIPs. Entries: *know* and *say*.

		Contextual factuality								
		mod=CT			**mod**<CT			**mod**=U		
		+	**−**	**u**	**+**	**−**	**u**	**+**	**−**	**u**
know	(a)	CT+	CT+	CT+	CT+	CT+	CT+	CT+	CT+	CT+
	(c)	CT+	Uu	Uu	Uu	Uu	Uu	Uu	Uu	Uu
say	(a)	Uu	Uu	Uu	Uu	Uu	Uu	Uu	Uu	Uu
	(c)	CT+	Uu	Uu	Uu	Uu	Uu	Uu	Uu	Uu

Moreover, the factuality assessments made by anchor and cognizer will vary depending on the polarity and modality in the SIP context. Compare the factuality assignments for sentences *a* in the following examples with those for sentences *b*, where the SIP is in a context of negative polarity.

(19)		**Context**	**Example**	**anchor**	**cognizer**
	a.	CT+	He$_s$ **knows** his father <u>died</u>$_e$ of a heart attack.	CT+	CT+
	b.	CT−	He$_s$ **does not know** his father <u>died</u>$_e$...	CT+	Uu
	a′.	CT+	He$_s$ **said** his father <u>died</u>$_e$ of a heart attack.	Uu	CT+
	b′.	CT−	He$_s$ **did not say** his father <u>died</u>$_e$...	Uu	Uu

These data can be systematized into a lexicon for SIPs, with each entry specifying the factual value assigned to the embedded event by both the anchor and the cognizer, relative to the polarity and modality values of the SIP context. The structure of lexical entries is as shown in Table 4, where each predicate has the information distributed in two different rows: one for the anchor (*a*), and another for the cognizer (*c*). For instance, the factuality value of event *die* in Example (19a) can be found in the 1st column of the rows for *know*, whereas the value for *die* in Example (19b) is in the 2nd column of the same rows.

Non-source Introducing Predicates (NSIPs). For convenience, all ESPs that do not contribute any additional source in discourse are grouped under the term of NSIPs. These include a varied set of predicate classes, such as:

(a) Implicative and semi-implicative predicates: *fail, manage*, or *allow*.

(b) Predicates introducing a future event as their complement, like volition (*want*), commissive (*offer*), and command (*require*) predicates.[15]

15 These predicates are considered as introducing a new source in Wiebe, Wilson, and Cardie (2005). Here, however, they are treated as NSIPs due to semantic considerations. Whereas SIPs express the epistemic attitude of their (logical) subject concerning the degree of certainty of the embedded event, predicates like *want* or *offer* denote the role of their subjects as either having some degree of responsibility on the embedded event (e.g., *promise/offer to go; force somebody to go*), or being in a greater or lesser favorable state towards its accomplishment (e.g., *need/want to go*). In other words, they express distinctions within the space of deontic modality. Nothing precludes us from treating them as SIPs if preferred, however.

Table 5
Lexicon fragment for NSIPs. Entries: *manage* and *fail*.

		Contextual factuality											
		CT			PR			PS			U		
		+	**−**	**u**	**+**	**−**	**u**	**+**	**−**	**u**	**+**	**−**	**u**
manage	(a)	CT+	CT-	CTu	PR+	PR-	PRu	PS+	PS-	PSu	Uu	Uu	Uu
fail	(a)	CT-	CT+	CTu	PR-	PR+	PRu	PS-	PS+	PSu	Uu	Uu	Uu

(c) Change of state predicates: *increase*, *change*, or *improve*.

(d) Aspectual predicates: *begin*, *continue*, and *terminate*.

By contrast to SIPs, NSIPs express a unique factuality assignment, attributed to the anchor source. Table 5 illustrates this with the lexical entries for NSIPs *manage* and *fail*. We invite the reader to verify the factuality values of the embedded event as provided by the table, given different factuality contexts of the NSIP (*manage/didn't manage/may have managed to go*, etc.).

3.4.4 Syntactic Constructions. Factuality information can be also conveyed through syntactic constructions involving subordination. Here we focus only on three of these structures: restrictive relative clauses, participial clauses, and purpose clauses.[16]

Purpose clauses. The main event denoted by a purpose clause is intensional in nature. Thus, all its relevant sources will assess it as underspecified (Uu), as is the case of *seek* in the following example, where the "b" part shows the factual assessment:

(20) a. Stronach resigned as CEO of Magna [to seek$_e$ a seat in Canada's Parliament].

b. $f(e, s_0)$=Uu

Relative and participial clauses. Three different situations apply. We illustrate them focusing on relative clauses, but assume the same treatment for participial clauses as well. First, in generic contexts, the event denoted by the relative clause is presupposed as corresponding to a fact in the world (CT+), regardless of the modality and polarity of the event in the main clause. In the following sentence, for example, the main event e_1 is characterized as a counterfact (CT−) but the event *working* in the relative clause is presented as a fact (CT+).

(21) a. After World War II, industrial companies could not **fire**$_{e_1}$ the women [$_{relative_cl.}$ that had been ***working***$_{e_2}$ in their plants during the war period].

b. $f(e_1, s_0)$=CT− $f(e_2, s_0)$=CT+

16 Our decision is motivated by practical reasons. These are the only constructions recognized by the dependency parser on which De Facto, the implementation of our model, relies.

Second, in quoted contexts the anchor remains uncommitted with respect to the event in the relative clause:

(22) a. "[$_{quoted}$ After World War II, industrial companies could not **fire**$_{e_2}$ the women [$_{rel_cl.}$ that had been ***working***$_{e_3}$ in their plants during the war period]]," **argued**$_{e_1}$ Prof. Poe$_{s_1}$.

b. **anchor** : $f(e_3, author)$=Uu **cognizer** : $f(e_3, prof.poe_author)$=CT+

Third, in reported speech and attitudinal contexts, both the cognizer and the anchor commit to the event in the relative clause as a fact (CT+).[17]

(23) a. Prof. Poe$_{s_1}$ **thinks/said**$_{e_1}$ [$_{attit./rep.}$ that after World War II, industrial companies could not ***fire***$_{e_2}$ the women [$_{rel_cl.}$ that had been ***working***$_{e_3}$ in their plants during the war period]].

b. **anchor** : $f(e_3, author)$=CT+ **cognizer** : $f(e_3, prof.poe_author)$=CT+

The last two interpretations have for long been a matter of discussion in the literature. Here, we embrace the analyses defended by Geurts (1998) and Glanzberg (2003), among others. As will be shown in Section 5.4, this area turned out to be a source of both disagreement among annotators and error from our system.

4. Computing the Factuality Profiles of Events

The current section puts forward an algorithm for a factuality profiler, that is, a tool for computing the factuality profiles of events in text. As such, it integrates all the components presented so far: the scalar system of factuality degrees, an organized view of factuality informants, as well as the structuring of the linguistic devices employed by speakers to convey distinctions of factuality. The details of the system presented here are further elaborated in Saurí (2008).

4.1 Computational Approach

The core procedure of the factuality profiler applies top–down, traversing a dependency tree. Two reasons motivate a top–down approach. The first one is of empirical nature. As seen, syntactic subordination is directly involved in the factual characterization of events (mainly through ESPs), and due to the recursive character of natural language, the factuality of a given event may depend on non-local information located several levels higher in the tree (cf. the set of sentences in Example (6)).

The second reason for a top–down approach is methodological. We conceive the factuality profiler as a neutral and naive decoder; neutral in that it takes all sources as equally reliable; and naive, because it assumes that sources are trustworthy, based on the Gricean maxim of quality. That is, our model assumes that the information presented in the text is true, without questioning anyone's view or adopting a particular side.[18] In our model, the naive decoder assumption is applied by initiating the tree top of each

17 Technically speaking, the presupposition is blocked at the quoted level in Example (22), whereas it is projected up to the embedding level in Example (23).

18 We can then consider a later postprocessing using different weights in order to favor one source as more reliable than another.

Evaluation level				Factuality profile	Marker contribution	F_n: context values	S_n: relevant sources
l_0	Step 1:					CT+	s_0
	Step 2:		may		mod:PS	PS+	s_0
	Step 3:		not		pol:−	PS−	s_0
	Step 4:	Mia_{s_m}	be_aware_{e_1}	$p_{e_1} = \{<\text{PS}-, s_0>\}$	be_aware (lex.entry)		
l_1	Step 5:					CT+ Uu	(a) s_0 (c) $s_m_s_0$
	Step 6:	Joe_{s_m}	knows_{e_2}	$p_{e_2}=\{<\text{CT}+, s_0>, <\text{Uu}, s_m_s_0>\}$			
			...				

Figure 2
Computing event factuality in *Mia may not be aware that Joe knows (Paul is the father)*.

sentence with a default factuality value of CT+; that is, all sentences are assumed to be true according to their author. This initial value will be potentially modified by the factuality markers available at subsequent levels of the tree. Consider the sentence:

(24) Mia may not **be aware**$_{e_1}$ that Joe **knows**$_{e_2}$ Paul **is**$_{e_3}$ the father.

Figure 2 exemplifies the initial steps of the procedure computing the factuality profiles of its events (the full-fledged algorithm will be presented in Section 4.3, after introducing the relevant technical details).

The computation proceeds as follows. At the top level of the sentence, there is only one source involved, namely, the author of the text (s_0). She is the one uttering the sentence, and thus the one assessing the factuality of the event placed at its top level (i.e., Mia not being aware of something, e_1). By the naive decoder assumption, the factuality at the top level is set to CT+ (Step 1 in Figure 2).

As the algorithm proceeds down the tree, this value is updated to PS+ by the modal auxiliary *may* (Step 2) and to PS− by the polarity marker *not* (Step 3).[19] This is the factuality value available when the parser reaches event e_1 (*be aware*), which is consequently characterized as PS− according to source s_0, the text author. In other words, the factuality profile of event e_1, $p_{be_aware_{e_1}}$, is the set of factuality values relative to the relevant sources at its level: $p_{be_aware_{e_1}}=\{\langle \text{PS}-, s_0\rangle\}$ (Step 4). In the figure, this is indicated by the dotted line.

The computation continues. Being a SIP, the predicate *be aware* contributes a new source in the situation. In addition to the author (s_0), now there is also the source Mia ($s_m_s_0$). Mia is the cognizer of event e_2 (she is in an "unaware" epistemic stance concerning Joe's knowledge), whereas the author is the source anchoring that epistemic stance. Determining these roles is crucial, because now we can appeal to the lexical information in Table 4 in order to set the perspective of each of these sources. In accordance with the information there, the anchor of an epistemic state introduced by the SIP *be aware* (which behaves like the SIP *know*) in a context of factuality PS− is characterized with a factuality

19 For convenience, the contribution of the marker is signaled with mod if it affects the modality value, and pol if it impacts the polarity. Some lexical elements (e.g., the complementizer *that*) are left off the representation when not relevant for the computation.

stance of certainty (CT+), whereas the cognizer, being unaware, remains uncommitted (Uu) (Step 5). Because there are no other factuality markers affecting these values, when the parser reaches event e_2 (Joe knowing something) these are the factuality assignments constituting the factuality profile of that event: $p_{know_{e_2}}=\{\langle \text{CT+},s_0\rangle,\langle \text{Uu},s_m\text{-}s_0\rangle\}$ (Step 6).

Thus, the factuality of every event corresponds to the factuality information available at its context, as computed from the interaction of the different factuality markers scoping over it. SIPs are crucial inflection points throughout this computation, given that they reset the evaluation situation by introducing additional sources and characterizing the factuality perspective these take. Computationally, this is modeled by means of the concept of *evaluation level*. Every time a new source is incorporated in the discourse by means of a SIP, a new evaluation level is created. The next section details the technical specificities of this notion.

4.2 Evaluation Levels

Consider each sentence, S, as consisting of one or more **evaluation levels**, l. By default, sentences have a root evaluation level, l_0. Sentences with SIPs have more, corresponding to the levels of embedding created by these predicates. For example, a sentence with two SIPs, in boldface in Example (25b), has three evaluation levels. We identify each evaluation level by its embedding depth, expressed in the bracket subindices.[20]

(25) a. [$_{l_0}$ Paul is the father].

b. [$_{l_0}$ Mia may not be **aware** that [$_{l_1}$ Joe **knows** [$_{l_2}$ Paul **is** the father]].

Each evaluation level l_n has:

A set S_n of relevant sources. At the root level l_0, S_0 contains only one relevant source, s_0, corresponding to the author of the text. At each higher level $l_{n>0}$, a new source is introduced by the SIP triggering it.

A set E_n of events (one or more), the factuality of which is evaluated relative to each relevant source $s \in S_n$.

A set F_n of contextual factuality values. At the beginning of each new level, one or more factuality values are set (cf. the value CT+ applying the naive decoder assumption at the top level). These values are relative to the relevant sources in S_n, because each source may assess the same event differently.

The task of event identification can be carried out by already existing event recognizers. The next sections define the operations for identifying the set of relevant sources S_n and the factuality values these assign to each event in any evaluation level l_n.

4.2.1 Identifying Relevant Sources and Their Roles. The process for identifying the set of relevant sources S_n at each evaluation level l_n can be defined inductively.

20 Note that, because evaluation levels are only triggered by SIPs, a sentence can contain several levels of syntactic embedding and yet only one evaluation level, corresponding to the top one, l_0. The following example contains three embedded clauses (signaled with curly brackets) but only one evaluation level.

[$_{l_0}$ {*After four years there*}, *Freidin managed* {*to return to the country* {*where she was originally from*}}].

Definition 1
Relevant Sources

i. The set of relevant sources at level l_0 contains only a (non-nested) source, which corresponds to the text author: $S_0 = \{s_0\}$.

ii. The set of relevant sources at level l_n, where $n > 0$, is:
$S_n = S_{n-1} \cup \{s_n\text{-}z \mid s_n \text{ is the new source introduced at level } l_n \ \& \ z \in S_{n-1}\}$

Clause (*i*) needs no additional comment. Clause (*ii*) states that the set of relevant sources S_n at level l_n contains (a) the set of relevant sources at the previous level l_{n-1}, that is, S_{n-1} (this is expressed as the first part of the union); and (b) the set of all source chains composed of the new source s_n introduced at that level by the corresponding SIP, and a relevant source from the preceding level, $z \in S_{n-1}$ (second part of the union).

We use the sentence *Mia is not aware that Joe knows Paul is the father* to illustrate the set of relevant sources S_n identified at each level l_n by the previous definition:

(26)

[$_{l_0}$Mia$_{s_m}$ is not	**aware**$_{e_1}$ that	[$_{l_1}$Joe$_{s_j}$	**knows**$_{e_2}$	[$_{l_2}$Paul	**is**$_{e_3}$ the father]]]
	s_0		s_0		s_0
			$s_m\text{-}s_0$		$s_m\text{-}s_0$
					$s_j\text{-}s_0$
					$s_j\text{-}s_m\text{-}s_0$

Definition 1 seems to return an excessive number of sources at level l_2. In particular, the source chains $s_j\text{-}s_0$ and $s_j\text{-}s_m\text{-}s_0$ appear to be redundant, because both of them refer to the same person, Joe. Notwithstanding, the analysis is adequate if we want to account for Joe's epistemic stance relative to the other sources involved in the situation. Source expressions $s_j\text{-}s_0$ and $s_j\text{-}s_m\text{-}s_0$ represent in fact two different perspectives. Expression $s_j\text{-}s_m\text{-}s_0$ includes a reference to Mia, that is, it presents Joe's epistemic stance according to Mia, based on what the author says. On the other hand, expression $s_j\text{-}s_0$ refers to Joe's perspective only according to the author.

As asserted in the sentence, Mia is clueless about Joe's knowledge concerning Paul's paternity, whereas according to the author, Joe knows the fact. Strictly speaking, then, the event Paul being the father (e_3) is evaluated by $s_j\text{-}s_0$ as a fact in the world (CT+), but will be presented with an uncommitted value (Uu) from the perspective expressed by $s_j\text{-}s_m\text{-}s_0$.

The next step now is determining the roles for each of these sources. In Section 3.4 on factuality markers, we saw that this distinction is crucial for identifying the factuality stance of each involved source. The mechanism for finding the anchors A_n and cognizers C_n at each evaluation level l_n can be stated as follows:

Definition 2
Source Roles

i. At level l_0: $A_0 = \{s_0\}$ and $C_0 = \{s_0\}$.

ii. At level l_n, for $n > 0$:
$A_n = \{s \mid s \in S_{n-1} \ \& \ f(e_{n-1}, s) \neq Uu\}$ and
$C_n = \{s_n\text{-}s_a \mid s_n \text{ is the new source introduced at level } l_n \ \& \ s_a \in A_n\}$.

Clause (*i*) defines the sets of anchors and cognizers at the evaluation level l_0, which contains only the relevant source s_0 (the text author). At this level, the distinction

between anchor and cognizer is irrelevant, and so we arbitrarily establish s_0 as performing both roles.

Clause (*ii*) defines anchors and cognizers for higher evaluation levels, $l_{n>0}$. In particular, *anchors* are defined as those sources from the previous evaluation level, $s \in S_{n-1}$, that are not uncommited (Uu) towards the factuality of e_{n-1}, which is the SIP event embedding l_n (in the definition, the notation $f(e,s)$ expresses the factuality assessment made by source s over event e). Returning to Example (26), this restriction prevents selecting source Mary ($s_m_s_0$) as the anchor of event e_3, because she is presented as having an uncommitted perspective (she doesn't know) on event e_2. Given that more than one source in a level can commit to the same event, an event can have more than one anchor, hence the notion of anchor set.

Last, clause (*ii*) defines *cognizers* as those sources composed of the new source introduced at level l_n, s_n, nested relative to any anchor source at that level, $s_a \in A_n$. Computationally the notion of cognizer is therefore dependent on that of anchor, and given that more than one anchor is possible at each level, the cognizer role can be performed by several source chains as well. All other sources not satisfying the definition of either anchor or cognizer are assigned the role of *none*, expressed as (_).

We apply Definition 2 to the earlier sentence, as well as to a second one, structurally identical but with different SIPs setting each evaluation level:

(27)

$[_{l_0}$Mia$_{s_m}$ is not	**aware**$_{e_1}$ that	$[_{l_1}$Joe$_{s_j}$	**knows**$_{e_2}$	$[_{l_2}$Paul	**is**$_{e_3}$ the father]]]
	(a,c) s_0		(a) s_0		(a) s_0
			(c) $s_m_s_0$		(_) $s_m_s_0$
					(c) $s_j_s_0$
					(_) $s_j_s_m_s_0$

(28)

$[_{l_0}$Mia$_{s_m}$	**says**$_{e_1}$ that	$[_{l_1}$Joe$_{s_j}$	**told**$_{e_2}$ her	$[_{l_2}$Paul	**is**$_{e_3}$ the father]]]
	(a,c) s_0		(a) s_0		(_) s_0
			(c) $s_m_s_0$		(a) $s_m_s_0$
					(_) $s_j_s_0$
					(c) $s_j_s_m_s_0$

The roles for sources at level l_0 and l_1 are the same in both sentences: The role assignment at level l_0 is trivial, while at level l_1, Mia is the cognizer of event e_2 (Joe telling/knowing something) because she is the one cognitively aware, or unaware, of the fact that Joe is telling/knows something. Nonetheless, source roles are different at level l_2. In Example (27) Mia cannot be the anchor of Joe's epistemic stance because she is presented as unaware of that (Uu). Instead, the source anchoring Joe's epistemic stance concerning event e_3 is the author of the sentence, that is, s_0 (as opposed to $s_m_s_0$). Because of this, in Example (27) the cognizer role is performed by the source chain $s_j_s_0$, whereas in Example (28) is performed by $s_j_s_m_s_0$.

4.2.2 Identifying Contextual Factuality Values. In order to compute the factuality values assigned by the relevant sources to the events at each level, we start by associating a contextual factuality value f to each relevant source $s \in S_n$ every time a new level l_n is opened. We represent this mapping as $\langle f, s \rangle$, and subsequently define the set of contextual factuality values at level l_n as: $F_n = \{\langle f, s \rangle | f \text{ is a factuality value } \& \ s \in S_n\}$. The set of contextual factuality values F_n can be obtained as follows.

Definition 3
Contextual Factuality Values

i. At level l_0: $F_n = \{\langle \text{CT+}, s_0 \rangle\}$

ii. At level l_n, for $n>0$: $F_n = \{\ \langle f, s \rangle \mid s \in S_n \ \&\ f = Lex(e_{n-1}, c_{e_{n-1}}, r_s)\}$

Clause (*i*) sets the contextual factuality for evaluation level l_0. By default, at level l_0 the set F_n contains only the value CT+ relative to the text author: $\langle \text{CT+}, s_0 \rangle$. This applies the naive decoder assumption.

In clause (*ii*), the contextual factuality value f associated to each source s is determined by function *Lex*, which performs a lookup into the SIPs lexical base (Table 4) given the following parameters:

$\boldsymbol{r_s}$**:** The role performed by the source $s \in S_n$ (anchor, cognizer, or none).

$\boldsymbol{e_{n-1}}$**:** The SIP in the previous evaluation level l_{n-1} that is embedding the current level, l_n. The information in its lexical entry will provide the contextual factuality values for the relevant sources at the current evaluation level (cf. Table 4).

$\boldsymbol{c_{e_{n-1}}}$**:** The *committed* factuality value that was assigned to SIP e_{n-1} in the previous level l_{n-1}. All factuality values, except for the fully underspecified Uu, are considered *committed values*. For instance, in Example (29), the factuality value to be used for setting the contextual factuality values for level l_2 is CT+, the only committed value assigned to event *knows* (e_1) in level l_1.

(29)

$[_{l_0}$Mia$_{s_m}$ is not	**aware**$_{e_1}$ that	$[_{l_1}$ Joe$_{s_j}$	**knows**$_{e_2}$	$[_{l_2}$ Paul	**is**$_{e_3}$ the father]]]
	(a,c) $f(e_1, s_0)$=CT−		(a) $f(e_2, s_0)$=CT+ (c) $f(e_2, s_m\text{-}s_0)$=Uu		(a) $f(e_3, s_0)$=CT+ (_) $f(e_3, s_m\text{-}s_0)$=Uu (c) $f(e_3, s_j\text{-}s_0)$=CT+ (_) $f(e_3, s_j\text{-}s_m\text{-}s_0)$=Uu

We illustrate how clause (*ii*) works with the operation of setting the contextual factuality values when opening the evaluation level l_1 in Example (29). The SIP embedding this level (corresponding to parameter e_{n-1} in function *Lex*) is *be aware*, which receives the committed factuality value of CT− (parameter $c_{e_{n-1}}$). Furthermore, at level l_1 there are two relevant sources, s_0 and $s_m\text{-}s_0$, the first one performing the role of anchor, and the second the role of cognizer (parameters r_s). With all that information at hand, the contextual factuality values for level l_1 will be obtained by means of a dictionary lookup performed by function $Lex(e_{n-1}, c_{e_{n-1}}, r_s)$. Using the lexical information in Table 4, this operation can establish the following contextual factuality values for the anchor and cognizer sources at the new level l_1:

a. $Lex(be_aware, \text{CT}-, anchor_{s_0}) = \text{CT+}$

b. $Lex(be_aware, \text{CT}-, cognizer_{s_m\text{-}s_0}) = \text{Uu}$ (3)

If the role is *none*, there is no need to perform the lexical look-up. The contextual factuality value will be set to underspecified (Uu).

4.3 Algorithm

The factuality profiler algorithm is provided in Algorithm 1, which further develops that presented in Saurí and Pustejovsky (2007) by incorporating syntactic constructions.

Algorithm 1 De Facto: the Factuality Profiler.

```
1:  n ← 0
2:  set level l_n
3:  for all i in TREE do
4:      #PART 1: CHECK FOR SYNTACTIC MARKER
5:      if i is head of relative, participle, or purpose clause then
6:          update contextual factuality, F_n
7:      end if
8:      #PART 2: CHECK FOR EVENT
9:      if i is an event then
10:         obtain the factuality profile of i, p_i
11:     end if
12:     #PART 3: CHECK FOR LEXICAL MARKER
13:     if i is a SIP then
14:         n ← n + 1
15:         set level l_n
16:     else if i is another type of marker then
17:         update contextual factuality, F_n
18:     end if
19: end for
```

Its core procedure (lines 3–19) consists of three main components. Part 1 implements the effect of syntactic-based factuality markers (specifically, relative, participle, and purpose clauses), Part 2 is in charge of assigning the factuality value to every found event, and Part 3 implements the effect of lexical markers on the contextual factuality values.

Part 3 (checking whether the node found is a lexical marker of any sort and subsequently updating the contextual factuality values) needs to be performed after Part 2 (obtaining the factuality profile of any found event) due to the double nature of ESPs, which are both event-denoting expressions and, at the same time, lexical markers. As markers, they affect the contextual factuality of their embedded events. Hence, their factuality profile (Part 2) needs to be obtained before they update the context values (Part 3). This is illustrated in Figure 2. When the algorithm index i is at node *be_aware*, it must first obtain the factuality profile of that event (Step 4) before updating the contextual factuality according to the semantics of the verb *be aware* (Step 5). By contrast, Part 1 needs to be run before evaluating the factuality of the event given that it implements the effect of syntactic constructions imposing a specific factuality value to its main event.

The functionality of the algorithm splits into three main components, which are in charge of: (i) setting each new evaluation level l_n; (ii) updating the set of contextual factuality values, F_n, every time a new marker is found; and (iii) obtaining the factuality profile of events. We discuss them in what follows.

***(i)* Set Level** l_n (lines 1–2 and 14–15). This function is called every time a new level is opened, be it at the top of the tree (lines 1–2) or when a SIP is found (lines 14–15). It executes the following steps:

1. ***Identify the set of relevant sources at the current level,*** S_n**.** This procedure is carried out applying Definition 1.

Algorithm 2 Syntactic Markers in De Facto.

```
1: if i is head of a participle or a relative clause then
2:    if i is in a quoted area then
3:       anchor=Uu ; cognizer=CT+     #APPLY PLUG
4:    else
5:       anchor=CT+ ; cognizer=CT+  #PROJECT PRESUPPOSITION
6:    end if
7: else if i is head of a purpose clause then
8:    anchor=Uu ; cognizer=Uu        #APPLY UNCOMMITMENT
9: end if
```

2. ***For each*** $s \in S_n$***, identify its role (anchor, cognizer, or none).*** Computed applying Definition 2.
3. ***Set the contextual factuality values,*** F_n***.*** This is performed applying Definition 3, based on lexicon look-up.

(*ii*) Update the contextual factuality, F_n (lines 5–6 and 16–17). The update may be triggered by either a syntactic or a lexical marker. Lexical markers that are appropriate here are polarity particles, modality particles, or NSIPs.[21] Any time one of them is found in l_n, the profiler updates the contextual factuality values $v \in F_n$ according to the information it conveys (lines 16–17). Syntactic constructions, on the other hand, reset the contextual factuality values according to Algorithm 2, which articulates the linguistic analysis concerning participle, relative, and purpose clauses, as presented in Section 3.4.

(*iii*) Obtain the factuality profile of e**,** P_e (lines 9–10). Applied when an event is found. Due to the on-the-fly updating of the contextual factuality values in F_n whenever a new level is set (i) or a new marker is found (ii), the event profile is in fact already computed. The factuality profile for event e_n, p_{e_n}, corresponds to the set of contextual factuality values F_n available at that point.

5. Experiments and Evaluation

5.1 Implementation

The modeling of the factuality profiler put forward here has been implemented and evaluated against a corpus annotated for that purpose. The resulting tool, called De Facto, integrates the algorithm in the previous section, along with the linguistic resources with lexical and syntactic information structured as presented in Section 3.4, and articulated around the scalar definition of factuality values developed in Section 3.2. The approach is therefore entirely symbolic, involving lexical look-up while top–down traversing the dependency tree of each sentence. The lexical resources informing De Facto include those listed here. They will be made available to the community in the near future.

21 Recall that SIPs affect the contextual factuality as they set a new evaluation level.

Table 6
Distribution of ESPs in De Facto.

Part of Speech	SIPs	NSIPs	Total
Verbs	204	189	393
Nouns	58	107	165
Adjectives	27	61	88
Total	**289**	**357**	**646**

Polarity particles: A total of 11 negation particles distributed among adverbs (such as *not, neither*), determiners (*no, non*), and pronouns (*none, nobody*), together with the table on contextual polarity interactions (Table 2).

Modality particles: The set of 31 particles presented in Example (15), each accompanied with their default modality interpretation, as well as their interaction table (Table 3).

ESPs: The lexical entries for a total of 646 ESPs, distributed as shown in Table 6. Lexical entries structure their factuality information as illustrated in Tables 4 and 5 (for SIPs and NSIPs, respectively). The information in each lexical entry was compiled manually in a data-driven fashion by exploring its use in our corpora of reference, TimeBank and the American National Corpus (Slate and NYTimes fragments).[22]

De Facto takes as input a document (or a set of them) and returns the factuality profiles of each event. Input documents have been tokenized, POS-tagged, and parsed into dependency trees with the Stanford Parser (version 1.6; de Marneffe, MacCartney, and Manning 2006). In the current implementation, De Facto does not incorporate any component for recognizing events nor identifying source mentions in text. This information was generated from manual annotation and fed to the tool. The chaining of different source mentions into relevant sources is computed automatically, however, by means of Definition 1.

As output, De Facto returns the factuality profile of each event in the input text. Example (31) shows the factuality profiles for the events in (30).

(30) $\underline{\text{Analysts}}_{s_1}$ **said**$_{e_1}$ the $\underline{\text{government}}_{s_2}$ **knew**$_{e_2}$ a peaceful solution **was**$_{e_3}$ in reach.

(31)
$$\begin{array}{lll}
e_1\text{: } \textit{said} & \text{fp}(e_1) = \{\ \langle s_0, \text{CT+}\rangle\ \} & \text{(anchor)} \\
e_2\text{: } \textit{knew} & \text{fp}(e_2) = \{\ \langle s_0, \text{Uu}\rangle, & \text{(anchor)} \\
 & \quad\ \langle s_{1_}s_0, \text{CT+}\rangle\ \} & \text{(cognizer)} \\
e_3\text{: } \textit{was} & \text{fp}(e_3) = \{\ \langle s_0, \text{Uu}\rangle, & \\
 & \quad\ \langle s_{1_}s_0, \text{CT+}\rangle, & \text{(anchor)} \\
 & \quad\ \langle s_{2_}s_0, \text{Uu}\rangle, & \\
 & \quad\ \langle s_{2_}s_{1_}s_0, \text{CT+}\rangle\ \} & \text{(cognizer)}
\end{array}$$

5.2 Development and Evaluation Corpus

For developing and evaluating De Facto, we compiled FactBank, a corpus annotated with information concerning the factuality of events (Saurí and Pustejovsky 2009a).

22 Documented, respectively, at: `http://www.timeml.org/site/timebank/timebank.html`, and `http://americannationalcorpus.org`.

Table 7
Confusion matrix: Gold standard (rows) vs. De Facto output (columns).

	CT+	CT–	Ctu	PR+	PR–	PS+	PS–	Uu	NA	Total
CT+	1,131	0	0	0	0	2	0	84	59	1,276
CT−	13	33	0	0	0	0	0	1	4	51
CTu	1	0	0	0	0	0	0	0	0	1
PR+	12	0	0	8	0	0	0	3	2	25
PR−	0	0	0	0	0	0	0	0	0	0
PS+	7	0	0	0	0	22	0	2	2	33
PS−	0	0	0	0	0	0	2	0	0	2
Uu	226	4	1	2	0	17	0	532	22	804
Total	1,390	37	1	10	0	41	2	622	89	2,192

FactBank consists of 208 documents, which include all those in TimeBank (Pustejovsky et al. 2006) and a subset of those in the AQUAINT TimeML Corpus.[23] The TimeBank part was used for developing De Facto and its associated linguistic resources, and the AQUAINT TimeML part was set as the gold standard for evaluating its performance. TimeBank contains 183 documents (amounting to 88% of the documents in FactBank) and 7,935 events (83.6% of the events), and the AQUAINT part has 25 documents (12%) and 1,553 events (16.4%).

Overall, FactBank contains a total of 9,488 events. Given that each event can have more than one relevant source, FactBank has a total of 13,506 event/source pairs manually annotated with the set of factuality distinctions introduced in Table 1. The annotation has applied a battery of discriminatory tests grounded on the linguistic and logical relations at the core of Horn's analysis (refer to Section 3.2). The inter-annotation agreement from that exercise is $\kappa = 0.81$ (over 30% of events in the corpus). In terms of pairwise F_1-score (that is, taking one of the annotators as the gold standard), the agreement between annotators yielded: CT+: 0.93, CT−: 0.83, PR+: 0.57, PR−: 0.46, PS+: 0.56, PS−: 0.75, and Uu: 0.88. Overall, these results are highly satisfying considering the difficulty of the task and thus validate the approach on the annotation. See further details in Saurí and Pustejovsky (2009b).

5.3 Performance

The confusion matrix resulting from mapping the subset of FactBank used as gold standard against De Facto output is shown in Table 7. The total number at the bottom-right corner corresponds to the number of event/source pairs in the gold standard, that is, the number of instances to be classified with a factuality value. Classes PRu and PSu are not shown because they have no instance in the gold standard.

Instances classified in the NA column correspond to event/source pairs for which De Facto did not return a factuality judgment. An analysis of this pointed to errors in the dependency trees as the possible cause of this behavior. In other words, they seemed to be pairs involving sources mentioned in subordinated clauses that had not been parsed properly and, as a consequence, De Facto could not pair with their corresponding events. Because subordination structures are fundamental in De Facto's algorithm, we

23 `http://www.timeml.org/site/timebank/timebank.html`.

decided to evaluate the system on two different versions of the gold standard: a first one with the dependency trees originally returned by the parser (corresponding to the data in Table 7), and a second one where dependency errors on subordination had been manually corrected. In total, we corrected an estimated 2% (at the lowest bound) of the dependencies involving subordination structures.

Table 8 shows the results from running De Facto against both versions of the gold standard. De Facto's performance is evaluated in terms of precision and recall (P&R) and their harmonic mean, F_1 score. We considered only those categories for which there exist more than 10 instances classified as such in the gold standard; that is: CT+, CT−, PR+, PS+, Uu. Furthermore, P&R for the whole corpus is obtained by applying the measures of macro- and micro-averaging (last two columns in the table). Macro-averaging averages the result obtained in each class, and micro-averaging applies over the set of instances, regardless of class distribution. The first measure gives equal weight to each class and hence over-emphasizes the performance of the less populated ones, and the second one over-emphasizes the performance of the largest classes because it assigns equal weight to each instance. Given the uneven class distribution in our gold standard, we take the combination of both measures as indicative of the lower and upper bounds of the result.

As can be seen from Table 8, the corrected version of the gold standard attains much higher recall than the original one (especially for the classes CT−, PR+ and Uu). The reason for that is the absence of event/source pairs tagged as NA by our system (as opposed to what was appreciated in the confusion matrix on Table 7). In the corrected version, De Facto was able to follow the dependency tree, appropriately pair all the events with their sources, and return a factuality value for each pair.

The results obtained in all the categories for the corrected version of the gold standard are equivalent to or higher than those in the original one, except for the very particular case of PR+ precision. The fact that increasing the quality of the parsing results in better performance of the system validates the linguistic model in De Facto.

The results for CT−, PR+, and PS+ must be interpreted cautiously, given the sparsity of data in these classes. Nevertheless, the high precision achieved for CT− is encouraging, especially considering that polarity here is not only determined locally but by means of subordinating predicates as well. Similarly, the distinction between the two modal degrees PR and PS seems pertinent and possible to determine by the system. No instance was misclassified between the two, as shown in the confusion matrix (Table 7).

Table 8
P&R for each relevant category and for the whole corpus (macro- and micro-average).

	CT+	**CT–**	**PR+**	**PS+**	**Uu**	**Macro-A**	**Micro-A**
			Original parses				
Precision	0.81	**0.89**	0.80	**0.54**	0.86	0.78	0.82
Recall	**0.89**	0.65	**0.32**	0.67	0.66	0.64	0.79
F_1	**0.85**	0.75	**0.46**	0.59	0.75	0.70	0.80
			Corrected parses				
Precision	0.86	**0.90**	0.73	**0.56**	0.86	0.78	0.85
Recall	**0.92**	0.75	**0.44**	0.67	0.77	0.71	0.85
F-1	**0.89**	0.82	**0.55**	0.61	0.81	0.74	0.85

Evaluating De Facto's performance on both versions of the gold standard provides a look into two different aspects of the system. Whereas the original version shows its impact on a standard NLP pipeline, the corrected version puts the proposed algorithm to test by exposing it to complex sentences with several levels of embedding. In order to assess De Facto's results regarding these two aspects, we generated a baseline from a supervised learning approach, by means of support vector machines (SVM). We followed Prabhakaran, Rambow, and Diab (2010), which is state-of-the-art on automatic tagging of committed belief (cf. Diab et al. 2009b), a notion equivalent to modality and which distinguishes between certain vs. uncertain events. The classification that they propose is less fine-grained than ours (certain vs. probable vs. possible), but the information supporting the distinctions is exactly the same, and therefore we adopted the features employed in their best classifier (listed from 1 to 12 in the following example). In addition, we added feature 13 given that our classifier was not aiming at identifying event mentions in the text (contrary to Prabhakaran, Rambow, and Diab's model), and features 14 and 15 to cope with distinctions along the axis of polarity (not addressed by that system).

(32)

1. isNumeric	Word is Alphabet or Numeric?
2. POS	Word's POS tag
3. verbType	Modal, auxiliary or regular (*nil* if not a verb)
4. whichModalAmI	If I am a modal, what am I? (*nil* if not a modal)
5. amVBwithDaughterTo	Am I a VB (base verb) with a daughter *to*?
6. haveDaughterPerfect	Do I have a *have* form daughter? (only for verbs)
7. haveDaughterShould	Do I have a *should* daughter? (only for verbs)
8. haveDaughterWh	Do I have a daughter which is: *where, when, while, who, why*?
9. haveReportingAncestor	Am I an event with an ancestor whose lemma is: *believe, accuse, insist, seem, tell, say, find, conclude, claim, trust, think, suspect, doubt, suppose*?
10. parentPOS	What is my parent POS tag?
11. whichAuxIsMyDaughter	If my daughter is an auxiliary, what is it? (*nil* if not an auxiliary)
12. whichModalIsMyDaughter	If my daughter is a modal, what is it? (*nil* if not a modal)
13. amEvent	Am I an event?
14. whichPolarAmI	If I am a polar marker, am I a conjunction (*nor*), a pronoun (*none*) or other?
15. whichPolarIsMyDaughter	If my daughter is a polar particle, what type is it?
16. amSource	Am I a source?
17. whichSIPtypeAreMyAncest.	If I am a source, what SIP type are my ancestors? (based on the SIP classification in Section 3.4.3)
18. whichDepRelWithMyParent	If I am a source, what is my dependency relation with my parent?
19. whichSIPtypeAmI	If I am a SIP, which type am I?

Prabhakaran, Rambow, and Diab's work assesses the committed belief of only the author source, but in our case an event can receive several factuality values from different sources. Hence, we decided to generate two different models: the *author level model*, in which the factuality of events is assessed relative to the author of the text (i.e., at the level of source s_0), and the *top source level model*, in which event factuality is assigned according to the source with a higher level of nesting in the set of relevant sources for that event (e.g., s_m-s_j-s_0). Thus, features 16–19 were added to convey information on the top-level sources as well.

Following Prabhakaran, Rambow, and Diab's work, we trained our SVM classifiers using YAMCHA (Kudo and Matsumoto 2000) and used the same parameters applied to their best classifier: context width of 2 (i.e., the feature vector of any token includes the two tokens before and after), and the *One versus all* method for multiclass classification on a quadratic kernel with a c value of 0.5. For evaluation, we performed a 10-fold cross-validation.

Table 9 shows the results (F_1 measure) of the two SVM classifiers (author and top source levels, as well as their average) running on both the original and the corrected versions of the gold standard. For a more meaningful comparison with our system, we also computed De Facto's performance on these two source levels. The results are shown in Table 10, where we also added, as a reference point, the figures obtained from evaluating De Facto on all source levels (corresponding to the F_1 rows in Table 8).

Furthermore, we assessed whether De Facto's improvement over the baseline is statistically significant applying a one-sample two-tailed t-test over the results for every category at each source level. We applied the one-sample version of the t-test because De Facto's performance results do not conform a distribution, because they were obtained from running the system once over the evaluation subcorpus. In the test, the sample data corresponds to the results from the 10 runs of the SVM classifier, whereas the De Facto's value is taken as the expected (or null) hypothesis. For the top and author levels, the degree of freedom is $df = 9$ (from 10 runs $-$ 1), while for their average it is $df = 19$ (10 + 10 runs $-$ 1).

As seen in Table 9, there is no significant difference between the baseline generated from the original and the corrected versions of the corpus, which is explained by the fact that the SVM models are based on fairly local linguistic features and use very little information on subordination structures. What is in fact most noticeable in the baselines is the difference between the results on the author and the top source levels for the less populated classes (CT−, PR+, and PS+). The top source level reaches much higher results, which could be explained by the greater use of dependency-based features providing information on the top source (features 15–18). This hypothesis underlines the role of deep linguistic features for identifying the factuality of event mentions in text.

By contrast to the baseline, De Facto shows a significant improvement when running on the corrected version of the gold standard, which proves the adequacy of its model to the linguistic information it targets. The downside of that is potentially too much dependency on high quality linguistic data in order to obtain acceptable performance degrees. Nevertheless, the results in the two tables show that De Facto is performing equal or better than the SVM classifiers when fed with original (not corrected) output from a standard NLP pipeline, especially in the case of less populated classes, which happen to be the ones with marked polarity and modality values, that is, which feature negative polarity, or probable and possible modality values.

Table 9
Baseline performance (F_1 measures).

	CT+	CT–	PR+	PS+	Uu	Macro-A	Micro-A
				Original parses			
Author	0.88	0.53	0.07	0.29	0.75	0.53	0.83
Top sources	0.92	0.69	0.51	0.50	0.57	0.66	0.86
Average	0.90	0.61	0.29	0.39	0.66	0.59	0.84
				Corrected parses			
Author	0.88	0.54	0.07	0.27	0.77	0.53	0.83
Top sources	0.92	0.67	0.50	0.50	0.51	0.64	0.85
Average	0.90	0.61	0.28	0.38	0.64	0.58	0.84

Table 10
De Facto performance (F_1 measures).

	CT+	CT–	PR+	PS+	Uu	Macro-A	Micro-A
				Original parses			
All sources	0.85	0.75	0.46	0.59	0.75	0.70	0.80
Author	0.88	0.88 ***	0.67 ***	0.33	0.78	0.73 ***	0.84 *
Top sources	0.90	0.79 *	0.33 *	0.66 **	0.58	0.67	0.84
Average	0.89	0.84 ***	0.50 **	0.50 *	0.68	0.70 ***	0.84
				Corrected parses			
All sources	0.89	0.82	0.55	0.61	0.81	0.74	0.85
Author	0.90	0.91 ***	0.67 ***	0.35	0.84 **	0.75 ***	0.88 *
Top sources	0.93	0.85 **	0.53	0.67 **	0.65 *	0.74 *	0.88
Average	0.92	0.88 ***	0.60 ***	0.51 *	0.75 *	0.75 ***	0.88 **

* $p \leq 0.05$
** $p \leq 0.01$
*** $p \leq 0.001$

The low performance of the SVM models is due to the small sample of these classes in the corpus, and so it can be expected that with more training data the classifiers will learn to perform better, a fact that makes them dependent on the availability of significantly larger annotated corpora. De Facto, on the other hand, is grounded on the linguistic expression that articulates factuality distinctions in natural language, and therefore does not depend as much on corpus size but on a good modeling of the interaction among the relevant linguistic structures. In this sense, the results shown here are quite promising regarding the capabilities of our system, even though it suffers from some limitations, as will be seen next.

5.4 Error Analysis

We analyzed the errors returned by De Facto when run on the manually corrected version of the corpus. With this choice, we wanted to avoid error from the parser and hence obtain a more precise assessment on the adequacy of our computational model. This version of the corpus has 320 event/source pairs wrongly classified (14.6% on the total number of pairs), whereas the original version has 464 pairs (21.2%).

Most disagreements between De Facto's output and the gold standard are due to limitations in our system (84.4%), which mainly classify into insufficient coverage of factuality markers, either lexical or syntactic, and structural and lexical ambiguity. Other disagreements are due to some inaccuracy in the gold standard annotation (7.5%), or to an incorrect analysis from the dependency parser which escaped our manual correction (8.1%). Table 11 shows the error type distribution, distinguishing between lexical and syntactic error when relevant.

Insufficient coverage. There are a number of syntactic constructions crucially involved in determining the factuality nature of events and which, nevertheless, have not been

accounted for here, most commonly: copulative phrases, cleft structures (e.g., *But it's not tonight we're worried$_e$ about*), and conditional constructions (of the form *if... then...*, and equivalent). This amounts to 32.5% of the total error. De Facto also suffers from gaps at the lexical level, even though in a much lesser degree (1.9%). It lacks, for example, ESPs such as *conspiracy* (as in: *a conspiracy to commit murder*) or *easy* (e.g., *it is easier to do it*).

Ambiguity. De Facto does not cope with lexical polysemy of any type (18.1% of the total error). For example, the modal auxiliary *would* is employed in embedded contexts to express future (and hence CT+, which is how De Facto models this tense), but there are certain constructions in which it expresses some degree of uncertainty. A further interesting case involves ambiguity regarding the temporal reference of events. De Facto assumes that aspectual predicates of termination (e.g., *stop, finish*) qualify their embedded event as a fact (that is, it is a fact that they took place in the world), whereas the gold standard treats them as counterfactual (the event does not hold anymore).

At the syntactic level, there are cases of truly ambiguous constructions, such as relative and participial clauses, as well as event-denoting nouns, when embedded under contexts of report, propositional attitudes, or uncertainty (28.1% of the total error). Some of these ambiguities have long been discussed in the linguistics literature, and happened to be a source of remarkable disagreement among the FactBank annotators as well (cf. Saurí and Pustejovsky 2009b). The high error rate in this area seemed to suggest that the approach assumed in De Facto for these constructions (following Geurts [1998] and Glanzberg [2003]; see Section 3.4.4) was not completely adequate. Thus, we experimented running De Facto without the part of the algorithm dealing with them (Algorithm 2, lines 1–9). The results, however, are inconclusive. Although there is a slight improvement of 1 or 2 points over the F_1 of categories PS+ (from 0.59/0.61 to 0.61/0.63 when running on the original/corrected parses), and Uu (from 0.75/0.81 to 0.77/0.82 on the original/corrected parses), there is a decrease in other categories, such as PR+ (from 0.46 to 0.43, original parses) and CT− (from 0.82 to 0.80, corrected parses).

Overall, the main limitations observed here are shared with other work also approaching tasks of sub-sentential interpretation by means of linguistically heavy and resource-intensive models, such as Moilanen and Pulman (2007) or Neviarouskaya, Prendinger, and Ishizuka (2009), which address sentiment analysis based on the principle of compositionality. Moilanen, Pulman, and Zhang (2010) successfully explore the feasibility of combining this approach with a machine learning-based classifier.

Table 11
Error classification.

	Error source	%	% Lexical	% Syntactic
De Facto limitations	Insufficient coverage	34.4	1.9	32.5
	Ambiguity	46.2	18.1	28.1
	Other	3.8	–	–
	Subtotal	**84.4**	**20**	**60.6**
Other error sources	Gold standard	7.5	–	–
	Wrong dependency trees	8.1	–	–
	Subtotal	**15.6**	–	–

6. Related Work

The last decade has seen a growing interest on speculative language and its treatment within NLP. This has crystallized into research from a variety of perspectives, including general but also domain-specific (mainly biomedical), and reflects not only in the building of processing systems, but also in the area of corpus creation, where most of the conception and structuring of factuality-related information takes place, thus providing the support for more applied investigations.

6.1 Factuality Information in Corpora

In some corpora, factuality-related information is annotated as information complementary to the main phenomenon they target. It is, for instance, contemplated in different versions of the ACE corpus for the Event and Relation recognition task (see, e.g., ACE 2008), in the Penn Discourse TreeBank (Prasad et al. 2007), and in TimeBank (Pustejovsky et al. 2006). In other corpora, factuality information becomes the epicenter of their annotations. For example, Rubin (2007, 2010) is concerned with the notion of certainty, the Language Understanding Annotation Corpus (Diab et al. 2009a) focuses on the author's committed belief towards what is reported (a notion comparable to the modality axis in event factuality), and the small knowledge-intensive corpus by Henriksson and Velupillai (2010) targets degrees of certainty.

In the bioNLP area, factuality and related information is lately becoming a notable area of research and has led to the creation of remarkable corpus resources. The BioScope corpus (Vincze et al. 2008) contains more than 20,000 sentences annotated with speculative and negative key words and their scope. Based on this experience, Dalianis and Skeppstedt (2010) compiled a corpus of Swedish electronic health records with speculation and negative cues marked up, together with the values resulting from their interaction. The corpus presented in Wilbur, Rzhetsky, and Shatkay (2006) tags the polarity and certainty degree of clauses, along with other dimensions. The GENIA Event corpus (Kim, Ohta, and Tsujii 2008) contains 1,000 abstracts with biological events annotated with polarity and degrees of certainty, in addition to other information such as the lexical cues leading to these values (Ohta, Kim, and Tsuji 2007). Such an approach is followed by the currently on-going large scale annotation effort (Nawaz, Thompson, and Ananiadou 2010), with an event-centered annotation that includes polarity, degrees of certainty, and sources.

6.2 Systems for Identifying Factuality and Related Information

Systems devoted to identifying factuality-related information can be generally classified into two groups: (a) those prioritizing the identification of linguistic structure (that is, speculative cues and their scope); and (b) those focusing on the factuality values that result from these cues and their interaction. The first approach mostly revolves around the BioScope corpus, which has become a good catalyzer for research on this topic in the biomedicine domain. Part of it was used for the CoNLL-2010 shared task on Learning To Detect Hedges and their scope in Natural Language Text (Farkas et al. 2010). Moreover, it is at the basis of explorations on hedging and negation cues scope identification, such as Morante and Daelemans (2009a, 2009b), which apply a supervised sequence labeling approach, or Özgür and Radev (2009) and Velldal, Ovrelid, and Oepen (2010), which

combine supervised learning techniques with rule-based systems exploiting syntactic patterns.

Identifying modality and polarity cues and their scope is certainly a key aspect for determining the degree of factuality of events, but not sufficient if the values resulting from these cues and their interactions are not provided. Complementary to this perspective, the second approach to factuality-related information puts the emphasis on identifying speculative degrees (along the lines assumed in this article). Pioneering work within this view is Light, Qiu, and Srinivasan (2004), a paper exploring the use of speculative language in sentences from Medline abstracts. It experiments with a hand-crafted list of hedge cues as well as a supervised SVM in order to classify sentences as either certain, high, or low speculative. Drawing on this, Medlock and Briscoe (2007) address the classification of sentences into speculative or non-speculative as a weakly supervised machine learning task and perform experiments with SVMs, achieving a precision-recall breakeven point of 0.76. This line of research is further explored by Szarvas (2008). On the other hand, Shatkay et al. (2008) use the corpus developed by Wilbur, Rzhetsky, and Shatkay (2006) to explore machine learning classifiers for tagging data along the five dimensions in which it is marked up, including polarity and degrees of certainty. It is a challenging task in that it involves simultaneous multi-dimensional classification and, in some dimensions also, multi-label tagging. They experiment with SVMs and Maximum Entropy classifiers, and report very good results (macro-averaged F_1 of 0.71 for degrees of certainty and 0.97 for polarity).

Resourcing to rich linguistic information. As argued throughout the article, subordination structures play a crucial role in determining the factuality values of events as well as their relevant sources, but most of the work presented so far addresses the problem of event factuality identification by means of classifiers fed with linguistic features that are not fully sensitive to sentences' structural depth and the complex interactions among their constituents. Previous work using subordination syntax to model factuality is the tool for identifying polarity and modality using lexical information and subordinating contexts by Saurí, Verhagen, and Pustejovsky (2006). Similarly, Kilicoglu and Bergler (2008) use the data from Medlock and Briscoe (2007) to show the effectiveness of lexically centered syntactic patterns for distinguishing between speculative and non-speculative sentences.

These systems are, however, limited in that they neither account for the effect of multiple embeddings, nor distinguish between different sources. To our knowledge, the first system in which factuality-related information is computed applying top–down on a dependency tree, and hence potentially overcoming these limitations, is Nairn, Condoravdi, and Karttunen (2006), who model the percolation of the polarity feature down the syntactic structure. A somewhat comparable perspective is adopted in the work on sentiment analysis addressing the problem from a compositional perspective. For example, in Moilanen and Pulman (2007) and Moilanen, Pulman, and Zhang (2010) the well-known semantics principle of compositionality is applied for sentiment polarity classification at the (sub)sentence level, and in Neviarouskaya, Prendinger, and Ishizuka (2009), for recognizing emotions such as anger, guilt, or joy. All these cases involve the use of deep parsing and rich lexicons in a way very similar to the model presented here for event factuality. The main difference with respect to our approach, however, is that De Facto applies top–down, whereas these systems follow a bottom–up processing of the data, as determined by the principle of compositionality. Such difference is not trivial. A top–down approach allows to keep track and compute the

nesting of the different sources involved in the factuality assessment, a computation that does not follow naturally from processing bottom–up.

Factuality information according to their sources. A common feature in all the approaches mentioned so far is the lack of awareness of the role of information sources. The fundamental role of source participants is already acknowledged in previous work on opinion and perspective (most significantly, Wiebe, Wilson, and Cardie [2005]). Concerning factuality-related information, the work incorporating the parameter of sources in the computation is pretty recent. It is acknowledged in Diab et al. (2009b) and Prabhakaran, Rambow, and Diab (2010), who nevertheless explore only the feasibility of identifying the committed beliefs of the text author, as annotated in the Language Understanding Annotation Corpus (Diab et al. 2009a), by means of SVM classifiers, in the first case with basic linguistic features whereas in the second one incorporating dependency-based features, reaching a maximum overall F_1 of 53.97 and 64.0, respectively. The distinction of event factuality depending on sources is also present in the corpus presented by Nawaz, Thompson, and Ananiadou (2010), who differentiate between *current* (i.e., the author) or *other*. Nevertheless, no system has yet been built based on these data.

Factuality distinctions in the different systems. Determining the factuality value has generally been approached as a classification problem, but there is no agreement in the literature on what the classes should be. In assuming a three-fold distinction of values along the certainty axis (certain, probable, possible), our model takes a middle path between other proposals in the NLP literature that only differentiate between certain and uncertain (e.g., Medlock and Briscoe [2007] and its subsequent work, or Diab et al. [2009b]) and approaches that distinguish among four (e.g., Henriksson and Velupillai 2010) or even five degrees (Rubin 2007, 2010). As a matter of fact, our linguistic-based distinctions are shared with the approach in Wilbur, Rzhetsky, and Shatkay (2006), the GENIA corpus (Kim, Ohta, and Tsujii 2008) and, in particular, that in Nawaz, Thompson, and Ananiadou (2010).

7. Final Considerations

Knowing the factuality status of event mentions in discourse is important for any NLP task involving some degree of text understanding, but its identification presents challenges at different levels of analysis. First, we conceive event factuality as a continuum, but a discrete scale appears to be a better approach for its automatic identification. Second, the way language expresses the factuality of situations is complex because it involves multiple contributing and interrelating factors. And finally, the factuality of an event is always relative to the author but often involves other sources as well.

In this article, we put forward a computational model of event factuality with the aim of contributing to a better understanding of this level of speculation in language. The model is based on the grammatical structuring of factuality in languages such as English, and addresses the three aforementioned challenges. Specifically, it rests upon a three-fold distinction of the factuality scale, it acknowledges the possibility of different sources (with potentially contradictory views), and it is strongly grounded on the information provided by linguistic operators (including polarity and modality

particles, predicates of different types, and subordination constructions) together with their cross-level interactions.

The model has been implemented into De Facto, a tool that takes dependency trees as input and returns the factuality profiles of events in text. To the best of our knowledge, it is the only system capable of identifying event factuality degrees paired to all the relevant sources for each event. In order to better assess its results, we built a baseline with SVMs following the state of the art in the area. We run De Facto on two versions of the dependency parses: one with the dependency trees originally returned by the parser, and another where dependency errors in subordination constructions had been manually corrected. De Facto's performance increases significantly when run on the second one, thus proving that event factuality as modeled in our work is linguistically well-grounded. De Facto is not completely dependent on high-quality linguistic data, however. Its performance even when run on the original dependency trees is notably better than the baseline regarding the classes that are harder to identify, namely, those involving negative polarity or some degree of uncertainty, therefore showing the adequacy of De Facto as a component in a standard NLP pipeline as well.

De Facto has been implemented for English, and so the set of linguistic resources informing it are specific to this language. Porting it to other close languages, however, such as Romance or Germanic ones, is a feasible task. The conceptual distinctions of certainty and polarity are shared across these languages, as well as the main linguistic structures encoding factuality information and which are handled by De Facto, including specific lexical types (e.g., reporting, presuppositional, or implicative predicates of different kinds) and syntactic constructions (different structures of evidentiality such as hearsay, perception or inference, conditional structures, etc.). Hence, the porting to other languages would mainly involve a mapping of lexical entries.

Furthermore, given that most of these linguistic expressions are not domain-specific but belong to the general structure of any given language, it seems plausible to believe that the model can be applied to data from other domains, such as biomedicine, without the burden of having to compile large amounts of annotated corpus for every new area of knowledge. At most, it would involve enriching the set of hedging markers for each domain. More support is needed, however, in order to confirm this claim.

On the other hand, such a highly linguistically based approach has its drawbacks as well, because it suffers from limitations regarding its linguistic coverage (mainly syntactic constructions), and its incapability to deal with ambiguity in natural language. These are problems commonly shared with other work approaching tasks of sub-sentential interpretation by means of linguistically heavy and resource intensive models.

All in all, De Facto can provide valuable information for different NLP tasks. For example, it can be of great help in systems dedicated to identifying facts or tracking rumors on news reports, detecting degrees of uncertainty in medical records, or recognizing the different sources involved in reported situations. Similarly, event factuality information can contribute, together with other semantic layers (e.g., dependency relations, semantic role labeling, or event and entity coreference), to the challenging task of identifying textual entailment relations. In addition, any machine learning efforts towards event factuality identification can both train over De Facto's output, as well as benefit from the lexical types and syntactic features it uses when considering options for machine learning algorithm choice and feature engineering decisions. In other words, we believe that the linguistically motivated model we propose here can, in addition to provide actual information on natural language text, help us understand the phenomenon of event factuality and complement data-driven approaches commonly used in the field.

Acknowledgments

We are very grateful to Carlos Rodríguez-Penagos, Bernat Saurí, Jordi Atserias, Guillem Massó, Andreas Kaltenbrunner, Toni Badia, Sabine Bergler, and Marc Verhagen for their valuable comments and helpful discussions. We also want to thank our anonymous reviewers for helping make this a much better piece of work. All errors and mistakes are responsibility of the authors. This work was supported by an EU Marie Curie grant to R. Saurí, PIRG04-GA-2008-239414.

References

ACE, 2008. *Automatic Content Extraction. English Annotation Guidelines for Relations*. Linguistic Data Consortium, version 6.0–2008.01.07 edition. Available at `http://www.ldc.upenn.edu/Projects/ACE/`.

Aikhenvald, Alexandra Y. 2004. *Evidentiality*. Oxford University Press, Oxford.

Asher, Nicholas. 1993. *Reference to Abstract Objects in English*. Kluwer Academic Press, Dordrecht.

Bach, Kent and Robert M. Harnish. 1979. *Linguistic Communication and Speech Acts*. The MIT Press, Cambridge, MA.

Biber, Douglas and Edward Finegan. 1989. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text*, 9(1):93–124.

Chafe, Wallace. 1986. Evidentiality in English conversation and academic writing. In W. Chafe and J. Nichols, editors, *Evidentiality: The Linguistic Coding of Epistemology*. Ablex Publishing Corporation, Norwood, NJ.

Dalianis, Hercules and Maria Skeppstedt. 2010. Creating and evaluating a consensus for negated and speculative words in a Swedish clinical corpus. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 5–13, Uppsala, Sweden.

de Haan, Ferdinand. 1997. *The Interaction of Modality and Negation: a Typological Study*. Garland, New York.

de Marneffe, Marie-Catherine, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of LREC 2006*, pages 449–454, Genoa, Italy.

Diab, Mona, Bonnie Dorr, Lori Levin, Teruko Mitamura, Rebecca Passonneau, Owen Rambow, and Lance Ramshaw. 2009a. *Language Understanding Annotation Corpus*. Linguistic Data Consortium, Philadelphia, PA. LDC2009T10.

Diab, Mona T., Lori Levin, Teruko Mitamura, Owen Rambow, Vinodkumar Prabhakaran, and Weiwei Guo. 2009b. Committed belief annotation and tagging. In *Proceedings of the Third Linguistic Annotation Workshop, ACL-IJNLP'09*, pages 68–73, Suntec, Singapore.

Dor, Daniel. 1995. *Representations, Attitudes and Factivity Evaluations. An Epistemically-based Analysis of Lexical Selection*. Ph.D. thesis, Stanford University.

Farkas, Richárd, Veronika Vincze, György Móra, János Csirik, and György Szarvas. 2010. The CoNLL-2010 shared task: Learning to detect hedges and their scope in natural language text. In *Proceedings of the 14th CoNLL Conference – Shared Task*, pages 1–12, Uppsala, Sweden.

Geurts, Bart. 1998. Presuppositions and anaphors in attitude contexts. *Linguistics and Philosophy*, 21:545–601.

Givón, Talmy. 1993. *English Grammar. A Function-Based Introduction*. John Benjamins, Amsterdam.

Glanzberg, Michael. 2003. Felicity and presupposition triggers. In *University of Michigan Workshop in Philosophy and Linguistics*, Michigan.

Halliday, M. A. K. and Christian M. I. M. Matthiessen. 2004. *An Introduction to Functional Grammar*. Hodder Arnold, London.

Henriksson, Aron and Sumithra Velupillai. 2010. Levels of certainty in knowledge-intensive corpora: An initial annotation study. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 41–45, Uppsala, Sweden.

Hickl, Andrew and Jeremy Bensley. 2007. A discourse commitment-based framework for recognizing textual entailment. In *Proceedings of the Workshop on Textual Entailment and Paraphrasing*, pages 171–176, Prague.

Hooper, Joan B. 1975. On assertive predicates. In J. Kimball, editor, *Syntax and Semantics, IV*. Academic Press, New York, pages 91–124.

Horn, Laurence R. 1972. *On the Semantic Properties of Logical Operators in English*. Ph.D. thesis, UCLA. Distributed by the Indiana University Linguistics Club, 1976.

Horn, Laurence R. 1989. *A Natural History of Negation*. University of Chicago Press, Chicago, IL.

Hyland, Ken. 1996. Writing without conviction? Hedging in science research articles. *Applied Linguistics*, 14(4):433–454.

Karttunen, Lauri. 1970. Implicative verbs. *Language*, 47:340–358.
Karttunen, Lauri and Annie Zaenen. 2005. Veridicity. In G. Katz, J. Pustejovsky, and F. Schilder, editors, *Dagstuhl Seminar Proceedings*, Schloss Dagstuhl, link: `http://drops.dagstuhl.de/opus/volltexte/2005/314/pdf/05151.KarttunenLauri.Paper.314.pdf`.
Kiefer, Ferenc. 1987. On defining modality. *Folia Linguistica*, XXI:67–94.
Kilicoglu, Halil and Sabine Bergler. 2008. Recognizing speculative language in biomedical research articles: A linguistically motivated perspective. *BMC Bioinformatics*, 9(Suppl 11):S10.
Kim, Jin-Dong, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun'ichi Tsujii. 2009. Overview of BioNLP'09 shared task on event extraction. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task*, pages 1–9. Boulder, Colorado, USA.
Kim, Jin-Dong, Tomoko Ohta, and Jun'ichi Tsujii. 2008. Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics*, 9(1):10.
Kiparsky, Paul and Carol Kiparsky. 1970. Fact. In M. Bierwisch and K. E. Heidolph, editors, *Progress in Linguistics. A Collection of Papers*. Mouton, The Hague, Paris, pages 143–173.
Kratzer, Angelika. 1991. Modality. In A. van Stechow and D. Wunderlich, editors, *Semantik: Ein internationales Handbuch der zeitgenoessischen Forschung*. Walter de Gruyter, Berlin, pages 639–650.
Kudo, Taku and Yuji Matsumoto. 2000. Use of support vector learning for chunk identification. In *Proceedings of CoNLL-2000 and LLL-2000*, pages 142–144, Lisbon, Portugal.
Lakoff, George. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2(4):458–508.
Light, Marc, Xin Ying Qiu, and Padmini Srinivasan. 2004. The language of bioscience: Facts, speculations, and statements in between. In *BioLINK 2004: Linking Biological Literature, Ontologies, and Databases*, pages 17–24, Boston, Massachusetts, USA.
Lyons, John. 1977. *Semantics*. Cambridge University Press, Cambridge.
Martin, James R. and Peter R. R. White. 2005. *Language of Evaluation: Appraisal in English*. London and New York: Palgrave Macmillan.
Medlock, Ben and Ted Briscoe. 2007. Weakly supervised learning for hedge classification in scientific literature. In *Proceedings of the 45th ACL*, pages 992–999, Prague, Czech Republic.
Moilanen, Karo and Stephen Pulman. 2007. Sentiment composition. In *Proceedings of the RANLP*, pages 27–29, Borovets, Bulgaria.
Moilanen, Karo, Stephen Pulman, and Yue Zhang. 2010. Packed feelings and ordered sentiments: Sentiment parsing with quasi-compositional polarity sequencing and compression. In *Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2010)*, pages 36–43, Alacant, Spain.
Morante, Roser and Walter Daelemans. 2009a. Learning the scope of hedge cues in biomedical texts. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pages 28–36, Boulder, Colorado, USA.
Morante, Roser and Walter Daelemans. 2009b. A metalearning approach to processing the scope of negation. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*, pages 21–29, Boulder, Colorado, USA.
Mushin, Ilana. 2001. *Evidentiality and Epistemological Stance*. John Benjamin, Philadelphia, PA.
Nairn, Rowan, Cleo Condoravdi, and Lauri Karttunen. 2006. Computing relative polarity for textual inference. In *Inference in Computational Semantics, ICoS-5*, pages 67–76, Buxton, England.
Nawaz, Raheel, Paul Thompson, and Sophia Ananiadou. 2010. Evaluating a meta-knowledge annotation scheme for bio-events. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 69–77, Uppsala, Sweden.
Neviarouskaya, Alena, Helmut Prendinger, and Mitsuru Ishizuka. 2009. Compositionality principle in recognition of fine-grained emotions from text. In *Proceedings of the 3rd International ICWSM Conference*, pages 278–281, San Jose, California, USA.
Ohta, Tomoko, Jin-Dong Kim, and Jun'ichi Tsuji. 2007. *Guidelines for event annotation*. University of Tokyo, link: `http://www-tsujii.is.s.u-tokyo.ac.jp/~genia/release/Genia_event_annotation_guidelines.pdf`.

Özgür, Arzucan and Dragomir Radev. 2009. Detecting speculations and their scopes in scientific text. In *Proceedings of the 2009 EMNLP Conference*, pages 1398–1407, Suntec, Singapore.

Palmer, Frank R. 1986. *Mood and Modality*. Cambridge University Press, Cambridge.

Polanyi, Livia and Annie Zaenen. 2006. Contextual valence shifters. In W. B. Croft, J. Shanahan, Y. Qu, and J. Wiebe, editors, *Computing Attitude and Affect in Text: Theory and Applications*, volume 20. Springer-Verlag, New York, pages 1–10.

Prabhakaran, Vinodkumar, Owen Rambow, and Mona Diab. 2010. Automatic committed belief tagging. In *CoLing 2010. Poster Volume*, pages 1014–1022, Beijing, China.

Prasad, Rashmi, Nikhil Dinesh, Alan Lee, Aravind Joshi, and Bonnie Webber. 2007. Attribution and its annotation in the Penn Discourse Treebank. *Traitement Automatique des Langues*, 47(2):43–64.

Pustejovsky, James, Bob Knippen, Jessica Littman, and Roser Saurí. 2005. Temporal and event information in natural language text. *Language Resources and Evaluation*, 39(2):123–164.

Pustejovsky, James, Marc Verhagen, Roser Saurí, Jessica Littman, Robert Gaizauskas, Graham Katz, Inderjeet Mani, Robert Knippen, and Andrea Setzer. 2006. *TimeBank 1.2*. Linguistic Data Consortium, Philadelphia, PA. LDC2006T08.

Rizomilioti, Vassiliki. 2006. Exploring epistemic modality in academic discourse using corpora. In E. Arnó Macià, A. Soler Cervera, and C. Rueda Ramos, editors, *Information Technology in Languages for Specific Purposes*, volume 7. Springer, Berlin, pages 53–71.

Rubin, Victoria L. 2006. *Identifying Certainty in Texts*. Ph.D. thesis, Syracuse University.

Rubin, Victoria L. 2007. Stating with certainty or stating with doubt: Intercoder reliability results for manual annotation of epistemically modalized statements. In *Proceedings of the NAACL-HLT 2007*, pages 141–144, Rochester, NY.

Rubin, Victoria L. 2010. Epistemic modality: From uncertainty to certainty in the context of information seeking as interactions with texts. *Information Processing and Management*, 46:533–540.

Saurí, Roser. 2008. *A Factuality Profiler for Eventualities in Text*. Ph.D. thesis, Brandeis University.

Saurí, Roser and James Pustejovsky. 2007. Determining modality and factuality for text entailment. In *Proceedings of the First IEEE International Conference on Semantic Computing*, pages 509–516, Irvine, California, USA.

Saurí, Roser and James Pustejovsky. 2009a. *FactBank 1.0*. Linguistic Data Consortium, Philadelphia, PA. LDC2009T23.

Saurí, Roser and James Pustejovsky. 2009b. FactBank. A corpus annotated with event factuality. *Language Resources and Evaluation*, 43:227–268.

Saurí, Roser, Marc Verhagen, and James Pustejovsky. 2006. SlinkET: A partial modal parser for events. In *Proceedings of LREC 2006*, pages 1332–1337, Genoa, Italy.

Shatkay, Hagit, Fengxia Pang, Andrey Rzhetsky, and W. John Wilbur. 2008. Multi-dimensional classification of biomedical text: Toward automated, practical provision of high-utility text to diverse users. *Bioinformatics*, 24:2086–2093.

Szarvas, Gyorgy. 2008. Hedge classification in biomedical texts with a weakly supervised selection of keywords. In *ACL 08: HLT*, pages 281–289, Columbus, Ohio, USA.

van Valin, Robert D. and Randy J. LaPolla. 1997. *Syntax. Structure, Meaning and Function*. Cambridge University Press, Cambridge.

Velldal, Erik, Lilja Ovrelid, and Stephan Oepen. 2010. Resolving speculation: Maxent cue classification and dependency-based scope rules. In *Proceedings of the 14th CoNLL: Shared Task*, pages 48–55, Uppsala, Sweden.

Vincze, Veronika, György Szarvas, Richárd Farkas, György Móra, and János Csirik. 2008. The BioScope corpus: Biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9(Suppl 11):S9.

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

Wilbur, W. John, Andrey Rzhetsky, and Hagit Shatkay. 2006. New directions in biomedical text annotation: definitions, guidelines and corpus construction. *BMC Bioinformatics*, 7(1):356–365.