CrossMark

ORIGINAL PAPER

Enriching news events with meta-knowledge information

Paul Thompson[1] · **Raheel Nawaz**[1] ·
John McNaught[1] · **Sophia Ananiadou**[1]

Abstract   Given the vast amounts of data available in digitised textual form, it is important to provide mechanisms that allow users to extract nuggets of relevant information from the ever growing volumes of potentially important documents. Text mining techniques can help, through their ability to automatically extract relevant *event* descriptions, which link entities with situations described in the text. However, correct and complete interpretation of these event descriptions is not possible without considering additional contextual information often present within the surrounding text. This information, which we refer to as *meta-knowledge*, can include (but is not restricted to) the modality, subjectivity, source, polarity and specificity of the event. We have developed a meta-knowledge annotation scheme specifically tailored for news events, which includes six aspects of event interpretation. We have applied this annotation scheme to the ACE 2005 corpus, which contains 599 documents from various written and spoken news sources. We have also identified and annotated the words and phrases evoking the different types of meta-knowledge. Evaluation of the annotated corpus shows high levels of inter-annotator agreement for five meta-knowledge attributes, and moderate level of agreement for the sixth attribute. Detailed analysis of the annotated corpus has revealed further insights into the expression mechanisms of different types of meta-knowledge, their relative frequencies and mutual correlations.

✉ Sophia Ananiadou
Sophia.Ananiadou@manchester.ac.uk

[1]   National Centre for Text Mining, School of Computer Science, Manchester Institute of
Bioinformatics, University of Manchester, Manchester, UK

⚊ Springer

1 Introduction

The digital information era has made vast and continually growing amounts of data available in digital form. This potentially provides a very rich source of historical data for researchers. However, as the amount of data available grows, researchers face increasing difficulties in finding information that is of interest to their research questions. Simple keyword-based search systems are usually not adequate for this purpose, as researchers typically have to spend a lot of time trawling through volumes of mostly irrelevant data returned by their searches.

Text mining offers a solution to such problems, by automatically deriving rich semantic metadata about documents in a collection. This may include *named entities* (e.g., people, locations, organisations) and possibly more sophisticated information about how these entities are linked together in documents to describe *events* (e.g., attacks, arrests, deaths, births). For example, consider the following sentence:

(S1)   Oscar Pistorious *killed* his girlfriend in Pretoria last night.

The sentence describes a *death* event (indicated by the word *killed*), in which Oscar Pistorious is the agent/perpetrator and his girlfriend is the victim/subject of the event. The sentence also provides information about the timing (i.e., *last night*) and the location (i.e., *Pretoria*) of the event. This information can be systematically organised using an event representation scheme. For example, Fig. 1 shows the ACE 2005 (Walker et al. 2006) representation of the event.

Although the main focus of such annotation is on the identification of event participants, this alone is not sufficient for the correct and complete interpretation of these events. For example, the event might be described as something that has already occurred, or as something that is anticipated to occur in the future. It may be described as a definite occurrence, or there may be some degree of speculation about whether it actually happened or will happen. Furthermore, the event may correspond to the point of view of the author or that of a third party, and either party may express subjectivity or opinions towards the event. As an illustration of these subtle (but important) aspects of event interpretation, consider three more sentences (S2–S4):

(S2)   Mr Pistorious told the court that he deeply regrets *shooting* his girlfriend.
(S3)   According to unconfirmed reports, Oscar Pistorious may have *fatally shot* his girlfriend, Reeva Steenkamp, at his residence in Pretoria.
(S4)   Mrs Steenkamp said that she holds Oscar responsible for the tragic events that led to her daughter's *death*.

All three of the above sentences (S2–S4) are similar to S1 (and to each other), in that they all refer to the same event (i.e., the death of Reeva Steenkamp caused by Oscar Pistorious). However, the interpretation of the event is different in each sentence. S1 and S3 report the event as new or emerging information, while S2 and S4 mention it as already known or presupposed information. In S1, the information

Fig. 1 ACE 2005 representation of the event mentioned in sentence S1

TYPE: LIFE_DIE
TRIGGER: *killed*
AGENT: *Oscar Pistorious*
VICTIM: *his girlfriend*
PLACE: *Pretoria*
TIME: *last night*

source of the event is the author herself; in S2 and S4, the source is someone involved in the event; and in S3 the information has been attributed to unknown third-party sources. The occurrence of the event is mentioned speculatively in S3, while S1, S2 and S4 report it with apparent certainty. Finally, S2 and S4 contain indications of negative sentiments towards the event, while S1 and S3 do not contain any sentiment or opinion about the event.

These examples demonstrate that merely detecting the event participants and their respective roles in the event is not sufficient; instead, additional contextual information is required for correct/complete interpretation of the event. We refer to this type of contextual information as *meta-knowledge* (Nawaz et al. 2010b) pertaining to the event. However, it is important to note that the term *extra-propositional aspects of meaning* (Morante and Sporleder 2012) can also be used to refer to similar types of information.

The ability to automatically recognise meta-knowledge information has been shown to be important for various types of Natural Language Processing (NLP) applications, including information extraction, question answering, summarisation, essay analysis and opinion mining (Wiebe et al. 2004; Riloff et al. 2005; Stoyanov et al. 2005; Webber et al. 2012). Such meta-knowledge has also been shown to improve the sophistication of event extraction systems (Miwa et al. 2012b; Chen et al. 2009; Nawaz et al. 2013a), and can provide additional filtering criteria in semantic search systems (Hirohata et al. 2008).

Building on previous work aimed at enriching biomedical events with meta-knowledge information (Nawaz et al. 2010b, 2012b), this paper describes our work on carrying out a similar type of enrichment of events within a different domain, i.e., news stories. The content of such texts, together with the types of events annotated within them, are very different from those in scientifically and academically oriented articles. Accordingly, we have made substantial changes to the annotation scheme employed, to make it more suitable for application to events concerning news. For this purpose, we took the ACE 2005 corpus (Walker et al. 2006) as our starting point, and modified and updated the annotations based on our new annotation scheme. We chose the ACE 2005 corpus because it is a well-known resource, which already contains some meta-knowledge annotations.

Our main contributions are as follows:

- We have developed a new meta-knowledge annotation scheme tailored for news events, together with associated annotation guidelines. The annotation

scheme comprises six meta-knowledge attributes. In relation to the original ACE 2005 annotation scheme, we have added two new annotation attributes (i.e., SUBJECTIVITY and SOURCE-TYPE) and have refined one attribute (i.e., MODALITY) by adding two new values (i.e., *Speculated* and *Presupposed*) and further specifying the definition of the existing values (i.e., *Asserted* and *Other*). We have not changed the existing values for the remaining three attributes (i.e., POLARITY, GENERICITY and TENSE). However, we have refined the annotation guidelines to further clarify the distinction between the values of these attributes.

- We have annotated the entire ACE 2005 corpus according to the new annotation scheme.
- We have annotated cue phrases that provide evidence for the assignment of specific attribute values.

The newly added attributes are intended to facilitate the development and/or enhancement of various NLP applications in which the ability to compare/contrast opinions or viewpoints can be important, e.g., systems that take multiple perspectives into account when carrying out summarisation (Teufel and Moens 2000) or question answering (Wiebe et al. 2003).

Evaluation of the annotated corpus has shown high inter-annotator agreement for the majority of the added/modified categories, whilst analysis of the annotated attributes has revealed various interesting patterns and correlations.

The meta-knowledge annotations and guidelines may be downloaded from http://www.nactem.ac.uk/ace-mk. The annotations are licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International licence.

The remainder of this paper is organised as follows: Sect. 2 provides a brief introduction to event-based text mining, and further highlights the need for meta-knowledge annotation. Section 3 describes the proposed annotation scheme in detail. Section 4 describes the annotation process and evaluation. Section 5 provides a detailed discussion on the analysis of annotated attributes and values. Finally, Sect. 6 contains brief concluding remarks.

2 Background and motivation

Following on from the discussion above, this section provides a more detailed account of event-based text mining, describes the significance of meta-knowledge and its annotation at the event level, and concludes with a brief overview of the ACE 2005 corpus.

2.1 Event-based text mining

As briefly mentioned in Sect. 1, event representations aim to capture the information content of a given text by systematically linking together the entities (e.g., people, organisations, locations, etc.) with events (e.g., actions, relations,

situations and states) mentioned in the text (Sauri and Pustejovsky 2009). The entities constitute the "players" (or participants) in the event and, according to the type of event being described, are linked together in different ways, with each participant playing a specific semantic role in the description of the event. For example, the event representation in Fig. 1 assigns the semantic roles of AGENT and VICTIM to the entities *Oscar Pistorious* and *his girlfriend* respectively. The event itself is also usually assigned a semantic type from a pre-defined list or ontology. For example, following the ACE 2005 event representation scheme, the event in Fig. 1 has been assigned the semantic type DIE, which is a sub-type of LIFE. Finally, central to the description of the event is a word or phrase (called the event *trigger*) around which the event participants are arranged. These triggers typically correspond to either verbs (e.g., S1, S2 and S3) or nouns (e.g., S4).

The goal of event extraction systems is to automate the process of recognising events in unstructured text, and to create structured representations such as the above. These structures can be exploited by NLP systems in various ways, e.g., to assist in automatic summarisation (e.g., Liao et al. 2013) or to create semantically-based search systems (e.g., Miyao et al. 2006). Particularly in the biomedical domain, automatic event extraction has been shown to have a broad range of applications (Ananiadou et al. 2015).

Manually annotated corpora of event representations facilitate the development of automatic event extraction systems. Several such corpora have been developed, often in the context of challenges aimed at pushing forward the state of the art in event extraction. These include the MUC (Grishman and Sundheim 1996) and ACE (Strassel et al. 2008) series (primarily newswire) and the BioNLP shared tasks (e.g., Nédellec et al. 2013) (biomedical text). These challenges have stimulated the development of a wide range of event extraction systems in each domain, e.g., (Aone and Ramos-Santacruz 2000; Ji and Grishman 2008; Miwa et al. 2012a; Bjorne and Salakoski 2013).

2.2 Significance of meta-knowledge

As discussed in Sect. 1, the mere recognition of event triggers and their participants is not sufficient for correct and complete event representation. As seen in the example sentences S1–S4, contextual meta-knowledge information is often present within the text, and must be considered to interpret the event correctly. Various types of meta-knowledge information have been demonstrated to be highly relevant in news articles. The expression of different sentiments and opinions in news articles has already been widely studied, e.g., (Bautin et al. 2008; Balahur et al. 2010), because news stories are rarely reported in a neutral way (Godbole et al. 2007). The identification of information source is also very important, given that as many as 90 % of news articles can contain direct or indirect reported speech (Bergler 2006). Additionally, attribution of information to a particular source could either be done in a positive way, to bolster a claim already made in the text, or otherwise to distance the author from the attributed material, implicitly lowering its credibility (Anick and Bergler 1992).

In the past few years, several corpora annotated with certain aspects of meta-knowledge have been created. However, each effort generally has a main focus, such as the identification of information about speculation/certainty, e.g., (Rubin et al. 2006; Rubin 2010), degree of factuality, e.g., FactBank (Sauri and Pustejovsky 2009), opinions, e.g., MPQA (Wiebe et al. 2005) or temporal information, e.g., TimeBank (Pustejovsky et al. 2003). There is often some level of overlap in the types of annotations in these different corpora, since the focussed information is usually supplemented with other information that is considered relevant to correct interpretation, such as polarity (positive or negative) and information source. In addition to the types of information annotated, these corpora vary in a number of other ways, including whether or not they annotate cue expressions that provide evidence for the categories assigned, and the granularity of the textual units annotated—these may be sentences, (sub-sentence) expressions or events. Related efforts in scientific domain (e.g., Wilbur et al. 2006; Nawaz et al. 2010a; Medlock and Briscoe 2007; Vincze et al. 2008; Light et al. 2004) identify some domain-specific features, although their annotation of features such as negation, speculation/certainly level and type of evidence/information source demonstrate the cross-domain importance of these types of information.

2.3 Meta-knowledge annotation of news events

It has been previously noted (Sauri and Pustejovsky 2009; Thompson et al. 2011a) that a given unit of text may contain a number of propositions or events, each of which may have a different interpretation, in terms of the types of meta-knowledge features introduced above. Since a single sentence may contain sentiments about multiple topics (Yi et al. 2003), the assignment of subjectivity values at the level of events can help to disentangle sentiments expressed towards different events in the sentence. Similarly, a sentence may contain some events which have already taken place and some events that are anticipated, feared, or speculated. For example, consider the following sentences (S5 and S6):

(S5) The Steenkamp family fears that Oscar Pistorious may not be found *guilty* of premeditated *murder* of Reeva Steenkamp.
(S6) Mr Roux said that he was relieved that Oscar was not found *guilty* of premeditated *murder*.

The above sentences contain the same event mentioned in sentences S1–S4. However, they also contain a second event, referring to the conviction of Oscar Pistorious for the crime of murder. The ACE 2005 event representation for S5 is shown in Fig. 2. The event representation for S6 would be similar, except that the value of the AGENT field in E1 and the DEFENDANT field in E2 would omit the surname Pistorious, and the VICTIM field in E1 would be empty.

The sentences S5 and S6 are similar, in that they both express the event E1 as presupposed (i.e., already known) information, and the event E2 is negated in both sentences. However, there are significant differences between the interpretation of event E2 in each sentence. In S5, E2 is presented as a speculation by a source involved in the event (i.e., *the Steenkamp family*). Moreover, the source has

ID:	E1	ID:	E2
TYPE:	LIFE_DIE	TYPE:	JUSTICE_CONVICT
TRIGGER:	*murder*	TRIGGER:	*guilty*
AGENT:	*Oscar Pistorious*	DEFENDENT:	*Oscar Pistorious*
VICTIM:	*Reeva Steenkamp*	CRIME:	*murder*

Fig. 2 ACE 2005 representation of the events mentioned in sentence S5

expressed negative sentiment towards the possible non-occurrence of this event (as denoted by the verb *fears*). However, in S6, the event E2 is presented as something that has already happened. Moreover, the source (i.e., *Mr Roux*) has expressed positive sentiment towards the event (according to his use of the verb *relieved*).

The above examples serve to illustrate the importance of identifying meta-knowledge at the event level. This importance has been demonstrated through the production of corpora containing one or more meta-knowledge features identified at the event level. Examples include Sauri and Pustejovsky (2009), Pustejovsky et al. (2003), Thompson et al. (2011b), and Walker et al. (2006). It has also been shown that meta-knowledge annotation at the event level can complement information annotated for coarser-grained units (Liakata et al. 2012). Such corpora could also form the basis for studying discourse structure at the event level, either by identifying discourse relations that hold between events, or by studying patterns of features that hold across sequences of events, in a similar way to the preliminary work carried out in (Nawaz et al. 2013c). Event-level discourse analysis could complement previous research into identifying discourse relations between coarser-grained units of text (e.g., Carlson et al. 2003; Marcu and Echihabi 2002; Prasad et al. 2008, 2011).

The utility of event-level meta-knowledge annotation has been demonstrated through the development of systems that have been trained to assign individual meta-knowledge attribute values to existing events (Nawaz et al. 2012a, 2013a, b) as well as fully integrated systems that are able to recognise events *and* multiple types of associated meta-knowledge (e.g., Ahn 2006; Miwa et al. 2012b). In terms of the performance of automatic meta-knowledge recognition, micro-averaged F-Scores generally range between around 70 and 98 %, according to the attribute being recognised.

Although, as mentioned above, there are already several corpora annotated with meta-knowledge features at the event level, these do not constitute ideal resources for training systems to assign fine-grained meta-knowledge attributes to complex event structures prevalent in news articles. For example, the GENIA-MK corpus (Thompson et al. 2011b) provides five types of meta-knowledge annotation for events occurring in biomedical abstracts. Whilst this annotation includes some domain-independent features, the large differences between the characteristics of scientific academic texts and news stories mean that even domain-independent

information is usually expressed in very different ways in the two text types. In contrast, the FactBank corpus (Sauri and Pustejovsky 2009) contains news stories. However, the types of event annotated do not have the same type of complex structure that was introduced above, i.e., event participants are not identified and characterised.

2.4 ACE 2005 corpus

We chose the ACE 2005 corpus (Walker et al. 2006) as our starting point for creating and implementing a meta-knowledge annotation scheme for news events. This was motivated by the following main reasons:

Size The ACE 2005 corpus comprises 599 news articles and contains annotations for 15,382 different entities and 5349 different events. The size of the corpus has already been shown to be sufficient to facilitate the training of a machine learning event extraction system with state-of-the-art performance (Miwa et al. 2014). A prototype, integrated system for extracting news events and associated meta-knowledge has been developed. Meta-knowledge in this system corresponds to the original attributes in the ACE 2005 corpus, as detailed below. The system has been used in the development of a semantic search system for the New York Times archive,[1] which allows search results to be refined based upon the presence of specific event types and meta-knowledge values (Thompson et al. 2013).

Event Normalisation All events in the corpus are grounded to one of the 33 designated event types, which fall under 8 different top-level categories that are frequently reported in news stories. These top-level categories are LIFE, MOVE-MENT, TRANSACTION, BUSINESS, CONFLICT, CONTACT, PERSONNEL and JUSTICE. For example, in the event representation of sentence S5 (shown in Fig. 2), event E1 has been assigned the event type DIE, which is a subtype of the event category LIFE, while event E2 has been assigned the CONVICT subtype of the category JUSTICE. For each event type, the ACE 2005 annotation scheme also specifies a potential set of semantic roles which can be instantiated by entities of specific types. For example, five sematic roles (AGENT, VICTIM, INSTRUMENT, TIME, and PLACE) are defined for the event type DIE, with type restrictions on each participant (e.g., the AGENT can only be an entity of type PERSON or ORGANISATION). The DIE event shown in Fig. 1 has four of these roles instantiated, while the DIE event in Fig. 2 only has two roles instantiated.

Owing to the fine-grained annotation, the normalisation of named entities and events, the specification of semantic roles for each event type, and the implicit restrictions on the types of entities participating in an event, the ACE 2005 corpus constitutes a highly suitable basis for developing semantically enhanced search and question answering systems. For example, such applications can potentially answer questions like, "Who was killed by Oscar Pistorious?", and "How/when/where did Reeva Steenkamp die?"

---

[1] A demo of this system can be found at: http://nactem.ac.uk/ISHER-NYT/. Please contact the authors for access.

Range The news articles have been taken from a variety of sources, including both written and spoken news. These include: broadcast news (BN), broadcast conversation (BC), conversational telephone speech (CTS), newswire (NW), Usenet newsgroups/discussion forums (UN) and weblogs (WL). Table 1 shows the distribution of these events across the six types of article sources.

Given such diversity of texts within the corpus, it provides a highly suitable test set for verification and validation of the proposed attributes and their respective categories in our annotation scheme.

Existing Meta-Knowledge Annotation The ACE 2005 corpus already includes some meta-knowledge attributes annotated at the level of events, in the form of attribute-value pairs. A brief description of each existing attribute is as follows:

POLARITY—This value is set to *Negative* if it is explicitly stated that the event did not take place. Otherwise the value is set to *Positive*. For example, referring back to sentence S5 and its event representation in Fig. 2, the polarity value for event E1 would be set to *Positive*, while the value for E2 would be *Negative*, as the word *not* explicitly negates the conviction event.

TENSE—The possible values for this attribute are: *Past*, *Present*, *Future* or *Unspecified*. These values are assigned according to the time that the event took place with respect to the textual anchor time (i.e., the time of broadcast or publication). *Unspecified* is assigned if it is not clear when the event took place or if it has taken place. For example, the value of E2 in S5 would be *Future*, while the value for E1 would be *Past*.

MODALITY—There are only two possible values for this attribute. The value is set to *Asserted* when the author or speaker makes reference to the event as though it were a real occurrence. In all other cases the value is set to *Other*. For example, the modality value for event E1 in S5 would be *Asserted*, while the value for E2 would be *Other*. This is because the death event (E1) is being described as something that has actually happened, but speculation is expressed towards the conviction event (E2).

GENERICITY—This attribute can also have two possible values. The value is set to *Specific* if the event is understood as a singular occurrence at a particular place and time, or a finite set of such occurrences; otherwise, the value is set to *Generic*. For example, the death events in sentences S1–S6 and the conviction events in S5 and S6 would all be assigned the value *Specific*, as they mention specific events. As an example of a *Generic* event, consider the death event mentioned in sentence S7:

(S7)   It is hoped that these measures will reduce the number of civilian *deaths*.

Although the above-mentioned attributes capture some aspects of event interpretation, they do not encode the subjective attitudes (pertaining to the event) that might have been expressed in the text. Similarly, the source of an event and its relative relationship to the event is not identified. Another limitation of the existing

Table 1 Distribution of annotated events across the six subparts of the ACE 2005 corpus

	BN	BC	CTS	NW	UN	WL	Overall
No of events	1184	914	468	1557	719	507	5349

meta-knowledge annotation is that the MODALITY attribute has been designed only to identify events that have actually taken place, and there is no way to distinguish events that have speculation expressed towards them. Moreover, no distinction is made between events being reported as "new" information and those describing "old/known" information. We also noticed that there were some inconsistencies in the original annotation of the above attributes. This is further discussed in Sect. 4. Finally, the existing meta-knowledge annotations do not include the corresponding evidence for the assignment of specific values, i.e., the words/phrases often present in the text that indicate a particular aspect of meta-knowledge regarding a specific event. Accordingly, we have aimed to improve the current meta-knowledge annotation in the ACE 2005 corpus, with the ultimate goal of facilitating the training of event extraction systems that are able to recognise rich meta-knowledge to a high degree of accuracy.

3 Annotation scheme

Our proposed scheme for enriching news events with meta-knowledge information consists of six attributes with a fixed set of values for each attribute. In comparison to the ACE 2005 annotation scheme, we have carried out the following:

- Added two new attributes (i.e., SUBJECTIVITY and SOURCE-TYPE).
- Refined one attribute (i.e., MODALITY) by adding two new values (i.e., *Speculated* and *Presupposed*) and further specifying the definition of the existing two values (i.e., *Asserted* and *Other*).
- Refined the annotation guidelines for the remaining three attributes (i.e., POLARITY, GENERICITY, and TENSE) to further clarify the distinction between the values of these attributes. We have re-annotated these three attributes, although we have not changed the original values.
- We have annotated the cue words/phrases that provide evidence for the assignment of particular attribute values, and linked them to the appropriate events.
- We have annotated named information sources and linked them to the appropriate events.

Figure 3 shows the updated annotation scheme. A brief description of each attribute is as provided below.

3.1 Source-type

This attribute aims to capture the source or origin of the information being expressed by the event. Our approach can be compared to various efforts to annotate information about attribution (e.g., Prasad et al. 2007; Pareti and Prodanof 2010; Pareti 2012a, b). All of these studies recognise the importance identifying details about the information source, and the latter efforts specifically aim to annotate the respective text spans that correspond to the *source* of the information, and to the *cue* (i.e., the word or phrase linking the source and information). In all of the above
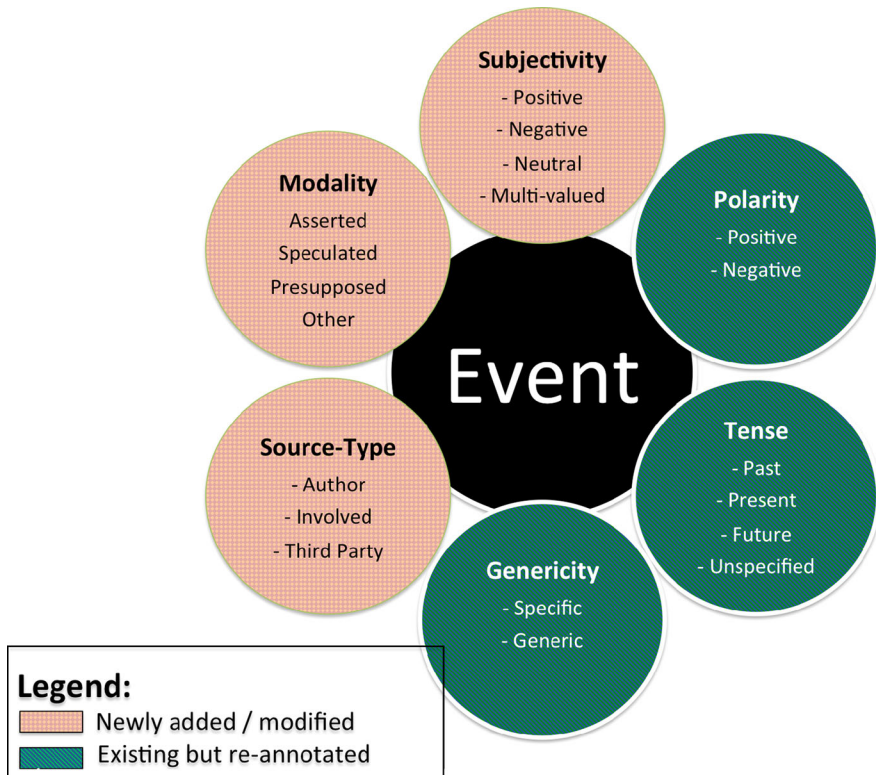
Fig. 3 Updated annotation scheme

efforts, an attribute is assigned to distinguish between different types of source, i.e., the writer, another specified agent, or an arbitrary, unspecified agent. In another study specifically targeted at news (Rubin 2010), a distinction is made between sources corresponding to direct participants and third-party experts. Taking inspiration from these previous studies, we distinguish between events that can be attributed to the correspondent/author, someone involved in the event, or some other third party. In case of third parties, we distinguish between named third party sources and unnamed third party sources (since unnamed sources are often considered less reliable than named sources). We annotate cues in all cases. Additionally, where the source is named, this is also annotated and linked to the event.

Brief descriptions of each value are as follows:

Author This value is assigned to events that are presented as information provided by the author, or as representing their own point of view. This is the default value, assigned to events unless there is any evidence for one of the other values. For example, the LIFE_DIE event reported in sentence S1 is being reported by the author (and there is no mention of any other source). Therefore, it would be assigned the *Author* value.

Involved This value indicates that the information expressed by the event is attributed to a specified source who is somehow involved or has close links to the actions described by the event. This may be an individual, group, government, political or terrorist organisation who is clearly involved in the event. This value is always determined through the presence of an explicit cue word or phrase, together with the name of the source. For example, consider sentences S2, S4, S5 and S6. In all four cases the source is named and is someone involved in the event.

Third-party This value indicates that the information expressed by the event can be attributed to a third party source that is not involved in the event. Third parties are always indicated by an explicit word or phrase. However, unlike involved sources, the description of third party sources can be vague sometimes, e.g., in sentence S2, the third party source is not named.

3.2 Subjectivity

Most news stories contain mentions of subjective opinions or attitudes towards the events being described. For example, an event that has already occurred can be praised, condoned or condemned. Similarly, a hypothetical or future event can be planned, proposed, wished for, or feared.

A broad range of different types of information can be grouped under the umbrella of "subjectivity". For example, taking inspiration from (Banfield 1982) and linking subjectivity to "private states" (Quirk 1985), Wiebe (1994) defines subjectivity analysis as the study of linguistic expressions of opinions, sentiments, emotions, evaluations, beliefs and speculations. Whilst the implicit subjectivity of events can depend upon complex interactions between explicit subjective expressions, advantages/disadvantages for particular event participants (Wiebe and Deng 2014; Deng et al. 2013) or emotions felt by them (Russo and Caselli 2013), the nature of news texts means that it is often difficult to distinguish between finely grained sub-categories of subjectivity (Balahur et al. 2010). As such, we decided to take a relatively simple approach to subjectivity annotation, which is focussed on identifying positive and negative sentiments that are expressed towards the event by the information source. In this respect, the information encoded through this attribute is comparable to the "attitude-type" annotation in the MPQA corpus (Wiebe et al. 2005). However, we also identify cases in which multiple types of subjectivity, both positive and negative, are specified in the context of an event, by multiple information sources. Given the complexity of the complete annotation task, which involves considering various other aspects of meta-knowledge, annotation of subjectivity information has been kept intentionally simple, and is restricted to identifying explicit expressions of subjectivity towards the event as a whole by the identified information source. Such subjectivity may be expressed either through an explicit cue, or through an event trigger that expresses strong subjectivity, such as *terrorism, genocide* or *massacre*.

Brief descriptions of each possible value are as follows:

Positive This value is assigned if the information source evaluates the event as good for themselves, for social groups with whose interests they identify, or for the wider community, whether or not they could be considered harmful to others. Such

events are often characterised by words indicating approval or anticipation, e.g., verbs like *want* and *urge*; adjectives like *good* and *positive*; nouns like *happy* and *excited*; and adverbs like *hopefully*, etc.

Negative This value applies when an event is evaluated as bad or harmful from the perspective of the source. Such events are often characterised by words indicating disapproval, apprehension, or fear, e.g., verbs like *worry*, *fear*; adjectives like *bad* and *negative*; nouns like *sad* and *afraid*; and adverbs like *unfortunately*, etc. Sometimes the event trigger itself also plays the role of a negative subjectivity cue, e.g., words like *genocide*, *holocaust*, *massacre*, *ambush*, etc.

Multi-valued Occasionally, two or more sources express opposite (i.e., positive and negative) sentiments about the same event. This value is used to identify such instances.

Neutral This is the default value for events with no explicit subjectivity information specified.

Referring back to sentence S5, the conviction event E2 (Fig. 2) would be assigned the *Negative* subjectivity value and the word *feared* would be annotated as the subjectivity cue, since this word denotes the stance of the information source, i.e., *the Steenkamp family*. However, the similar event in S6 would be assigned the *Positive* value and the word *relieved* would be marked as the corresponding cue, according to the sentiment expressed by Mr. Roux, who is the information source in this sentence. As example of *Multi-valued* subjectivity, consider the sentence S8 (below), where two different information sources refer to the same event, but with opposing sentiments.

(S8)   While President Obama was congratulating the nation, Al-Qaida issued a statement, vowing to avenge Osama's *death*.

3.3 Modality

As discussed in Sect. 2.4, this attribute already existed in the ACE 2005 corpus. However, the original aim of this attribute was only to distinguish between events that have actually taken place (i.e., *Asserted* events) and those that are planned, anticipated or feared (i.e., *Other* events). We have refined the values of this attribute to further distinguish between speculated and certain events, and between events describing new and presumed information. This has resulted in the addition of two new values (i.e., *Presupposed* and *Speculated*), and the redefinition of the existing values (i.e., *Asserted* and *Other*). A brief description of each value is as follows:

Asserted This value is assigned to definite events, i.e., situations where something has actually happened or is happening. However, in contrast to the original ACE 2005 annotation scheme, we have added the additional constraint that this value is only to be assigned to events that assert new information into the discourse.

Presupposed This is a new value, assigned to definite events that describe situations that are assumed to be already known by the listener/reader, or have been previously mentioned within the discourse. This is a relatively broad definition. For example, in comparison to the classes of *information status* (Prince 1992), it covers both *hearer-old* and *discourse-old* events. Likewise, compared to the *givenness*

hierarchy (Gundel et al. 1993), our definition of *Presupposed* includes four statuses (*in focus*, *activated*, *familiar*, and *uniquely identifiable*). We have introduced this value since, according to the fast moving nature of news events, it is important to be able to identify the "newest" part of an on-going news story.

Speculated This value is used to identify events for which there is some explicitly expressed uncertainty regarding their occurrence. Although related corpora make a greater number of distinctions with regard to certainty levels, e.g., Rubin (2007) distinguishes 5 different levels, it was found that annotators could only reach slight levels of agreement (0.15 κ) on such a detailed scale (Rubin 2010), hence our decision to use a more simple distinction.

Other This is the default value for events that do not fit into any of the above categories.

Referring back to the sentences S1–S4, the MODALITY value assigned to the LIFE_DIE event in S1 would be *Asserted*, as it describes an event that has actually taken place and is being reported as new information. Even though the LIFE_DIE events in S2 and S4 describe definite occurrences, they are not being presented as new information. Therefore, they will be assigned the *Presupposed* value. Finally, the LIFE_DIE event in S3 is presented as a speculation; therefore it will be assigned the *Speculated* value.

3.4 Polarity, genericity, and tense

Although we have not changed the existing values for these three attributes, we had noticed some apparent annotation inconsistencies in the ACE 2005 corpus. Therefore, we decided to re-annotate these attributes and produced extended guidelines to facilitate this. This is further discussed in the following section.

4 Annotation process and evaluation

This section contains brief discussions on the annotation of existing attributes, the annotation of meta-knowledge cues, an overview of the annotation process, and the evaluation of the annotations produced.

4.1 Annotation of existing attributes

Whilst the original ACE annotation guidelines included only very brief information about how to annotate the existing attributes, we have produced a new set of guidelines, covering both existing and new attributes. These guidelines include more detailed explanations for each attribute and its possible values, along with examples. We have included expanded explanations for the existing attributes, as we found that the very brief original guidelines had sometimes led to inconsistent annotations in the original corpus. For example, for the TENSE attribute, the *Unspecified* value was sometimes assigned whenever the event trigger was not a

tensed verb, e.g., words like *death* or *war,* even when the textual context of the event made clear the time of the event with respect to the textual anchor time.

In order to address the problem of existing inconsistent annotations, we decided that the task undertaken as part of the current work should include not only the annotation of the new or changed attributes, but also the review and possible update of the values of the unchanged attributes. By expanding the guidelines for these attributes, we aimed to foster a more common understanding amongst annotators of when to assign the most appropriate value, and hence to increase the consistency of the annotations. For example, we updated the guidelines to ensure that the value of the TENSE attribute reflects the time of the event according to the textual context. Additionally, by creating a full set of guidelines for all attributes, the same scheme can straightforwardly be applied to other corpora in the future.

4.2 Annotation of cue phrases

As previously mentioned, cue phrases can be helpful in identifying and character-ising meta-knowledge features of text spans and/or events. Several previous studies have found that such cues can be important in the interpretation of various aspects of academic texts, e.g., 85 % of speculated statements in biology articles have been found to be conveyed through the presence of particular cue words and phrases (Hyland 1996). Other studies have found that further types of discourse-related information can also be expressed through specific cues (e.g., Rizomilioti 2006; Thompson et al. 2008). Based on these findings, we previously enriched a corpus of events in biomedical text with information about their interpretation, including the identification of cue words and phrases (Thompson et al. 2011b). Subsequent training of a system that could automatically recognise events and their interpre-tation found that the presence of such cues improves the accuracy of predictions made about meta-knowledge information (Miwa et al. 2012b).

Based on the above findings, we decided to identify cues in the ACE 2005 corpus as part of the annotation effort. The aim is both to improve the quality of results obtained from machine learning, as well as providing a means to carry out an analysis of the type of language used to convey the various types of meta-knowledge information. Annotators were asked to identify any words or phrases in the same sentence as the event that provide evidence for the assignment of a specific value for one of the meta-knowledge attributes, to label them accordingly (e.g., *Modality-Cue, Subjectivity-Cue,* etc.) and to link these cues to the appropriate event. So, for example, in sentence S5, the word *may* would be annotated as a *Modality-Cue,* and linked to the event with the trigger *guilty,* as evidence for the assignment of the *Speculated* modality value. Similarly, in S6, *said* would be annotated as a *SourceType-Cue* and linked to the event with the trigger *guilty.*

Based on previous work (Thompson et al. 2011b; Vincze et al. 2008), we decided that, as a general rule, the span of the cue annotation should be the minimum unit of text which can be used to determine the correct value for the given annotation attribute. If the length of the cue is more than a single word, then the cue phrase must be a continuous span of text. This maintains consistency with the rest of the
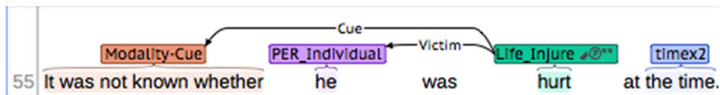
Fig. 4 Annotated sentence in brat

annotations in the ACE 2005 corpus, since all original annotations constitute continuous spans.

4.3 Annotation process

Based on the above observations about the original guidelines and existing annotation in the ACE 2005 corpus, we decided that the annotation process should consist of the steps detailed below. These were carried out for all 5349 events in the complete ACE 2005 corpus:

1. Reviewing and possibly updating the values of existing meta-knowledge attributes (i.e., POLARITY, TENSE, MODALITY and GENERICITY),
2. Assigning values for the new SUBJECTIVITY and SOURCE-TYPE attributes, as well as identifying the named information source in the text, if present, and linking it to the appropriate event.
3. Identifying and annotating cue words/phrases that provide evidence for the assignment of particular values to each of the six attributes, if such cues are readily identifiable in the text, and linking them to the appropriate event.

The annotation was carried out with the aid of the brat annotation tool.[2] This was chosen for a number of reasons. Firstly, it is very simple to use. Secondly, it provides support to display the complex event structures that are annotated in the ACE 2005 corpus. Finally, it is web-based and requires no installation, meaning that annotators can straightforwardly complete their tasks in any location where they have Internet access.

Figure 4 shows a simple example of an annotated sentence from the ACE 2005 corpus in brat. The original ACE annotation identified the LIFE_INJURE event, with the trigger *hurt*, and the *Victim* role in the event being played by the *PER_Individual* entity *he*. Using brat, it is straightforward to annotate new text spans by dragging the mouse over the span and then choosing a category from a pop-up menu. In Fig. 4, as part of the new annotation effort, the span *It is not known whether* has been annotated and assigned the category *Modality-Cue,* since it provides evidence for the assignment of the *Speculated* Modality value. The event and the cue are then linked by dragging the mouse between them.

The values of the meta-knowledge attributes are assigned by clicking on the event trigger. This brings up a pop-up window, with drop-down menus that allow appropriate values for each attribute to be assigned (Fig. 5).

Fig. 5 Meta-knowledge attribute annotation in brat

4.4 Corpus evaluation

During its development phase, the annotation scheme was tested and refined through an iterative process, in which two annotators with computational linguistics expertise annotated a common set of documents, and then compared and discussed the results. This process was particularly useful in highlighting the need to re-annotate the existing attributes in the ACE 2005 corpus.

Given the labour-intensive nature of the annotation process, the majority of the annotation effort was carried out by only one of the two annotators mentioned above. However, in order to evaluate the quality and consistency of the annotation, approximately one-fifth of the corpus (1000 events, roughly balanced amongst the six portions of the corpus) was also annotated by the second annotator. This has allowed us to calculate inter-annotator agreement scores. Following this, a consolidated version of the double-annotated part of the corpus was created, by discussing and reaching a consensus on any disagreements that occurred. Table 2 shows the agreement rates achieved between the two annotators.

Table 2 shows that there are variations in agreement, according to the attribute being annotated. In terms of the interpretations of Kappa provided in (Viera and Garrett 2005), the agreement achieved for the GENERICITY and POLARITY attributes is "almost perfect", for TENSE, MODALITY and SUBJECTIVITY, agreement is "substantial" and for SOURCE-TYPE, the agreement level is considered "moderate". Therefore, the levels of agreement achieved can be considered acceptable in all cases.

It is perhaps unsurprising that the attributes that achieve the highest levels of agreement are the ones that were already present in the ACE 2005 corpus, since the task for these attributes was mainly to review the existing values according to the updated guidelines. However, it should also be noted that although two new values were added to the MODALITY attribute, and the definitions of existing values were changed, "substantial" agreement was still achieved. Although the agreement for the SUBJECTIVITY is about 0.15 lower than for MODALITY, this is still considered to be "substantial" agreement. We consider this to be an encouraging result, given the complexity of the task, i.e., the potential subtlety of the ways in which positive or negative subjectivity can be expressed, and the variety of the types of cues that can be used. The wide range of vocabulary used in subjective expressions has been confirmed by other efforts that have annotated this type of

Table 2 Agreement rates for annotated discourse attributes

Attribute	Agreement (Kappa)
Genericity	0.871
Polarity	0.869
Tense	0.805
Modality	0.804
Subjectivity	0.659
Source-type	0.530

information, e.g., (Wiebe et al. 2005; Kessler et al. 2010). The fact that these studies report similar levels of agreement to ours, in terms of the identification of subjective expressions and/or their linking to target expressions, serves to emphasise the complexity of tasks that involve subjectivity identification.

We have also calculated agreement for cue phrase identification. Since certain meta-knowledge attributes (e.g., TENSE and GENERICITY) rarely have associated cue phrases, we report average agreement on the choice of appropriate cue phases over all attributes, in cases where annotators agree on the value of the corresponding attribute. It can be problematic to calculate Kappa when comparing choices of annotated text spans, given that chance agreement can be very small. Thus we have calculated cue phrase annotation agreement in terms of *positive specific agreement* (Hripcsak and Rothschild 2005), which approximates the proportion of positive cases that were agreed upon. The agreement rates are reported in Table 3, in terms of both exact matches (i.e., where the cue spans annotated by both annotators have to match exactly) and relaxed matches (i.e., where it is sufficient for there to be some level of overlap between the spans chosen by each annotator).

As shown in Table 3, there is a high degree of consensus between the annotators about which cue phrases to annotate. We found that disagreements may occur if there are multiple possible cues for a given dimension in a sentence. The relatively small difference in agreement rates between exact and relaxed spans illustrates that sufficient guidance was given to annotators regarding the extent of text to mark up as a cue.

4.5 Annotation challenges and resolution

As the above results show, the main annotation challenges were encountered for the SUBJECTIVITY and SOURCE-TYPE attributes. The majority (71 %) of SUBJECTIVITY disagreements in the double-annotated part of the corpus involved discrepancies between the *Negative* and *Neutral* values. Further investigation and

Table 3 Agreement rates for cue phrases

Agreement criterion	Agreement (positive specific agreement)
Exact match	0.895
Relaxed match	0.948

discussion of these revealed that in most of these cases, one or other of the annotators had failed to notice the negative subjectivity. In the consolidated corpus, most of these cases were thus agreed upon as instances of negative subjectivity. To give some idea of the complexity of identifying subjectivity cues, 324 unique negative subjectivity cues and 179 unique positive subjectivity cues were annotated in the whole corpus. On average, each negative subjectivity cue is associated with 1.84 events, and each positive subjectivity cue is associated with 1.78 events. This demonstrates that there are few "typical" ways of expressing positive or negative subjectivity, which makes the annotation task more difficult.

The most commonly occurring negative subjectivity cue, *terrorism* (which also functions as an event trigger) appears only 18 times in the entire corpus. In comparison, for the *Speculation* value of the MODALITY attribute, each unique cue is, on average, used almost three times more frequently than positive or negative subjectivity cues. Furthermore, the most commonly occurring cue for *Speculation* (i.e., *if*) occurs 87 times in the corpus, i.e., around five times more frequently than *terrorism.*

For the SOURCE-TYPE attribute, which has the lowest levels of agreement, 158 of the 173 disagreements (91 %) were found to be cases where one of the annotators had assigned the *Author* value, while the other annotator had assigned either the *Involved* or *Third Party* value. An examination of these disagreements showed that they were mostly annotation errors, in which one of the annotators had missed the fact that the information was explicitly stated as having come from a source other than the author. Such information was frequently missed when a short phrase such as *X said* was placed at the end of the sentence and far removed from the actual event. The nature of this type of error meant that nearly all occurrences could be agreed upon and corrected in the consolidated version of the corpus. However, it is worth noting that there were very few instances (15 in total) where the two annotators disagreed on whether to assign *Involved* or *Third Party* to events with a *Source* other than *Author.*

5 Annotation analysis

In this section, we present a discussion and analysis of the complete, updated ACE 2005 corpus annotation, considering each of the six annotated attributes separately. In each case, we consider statistics from the corpus as a whole, and also its subparts, i.e., BN (Broadcast News), BC (Broadcast Conversation), CTS (Conversational Telephone Speech), NW (Newswire), UN (Usenet Newsgroups/Discussion Forums) and WL (Weblogs).

5.1 Modality

Almost half of the events (around 47 %) correspond to the newly introduced values (i.e., *Speculated* or *Presupposed*). This provides strong evidence that our decision to include these categories was well-motivated, since these types of information occur frequently, but were not distinguished in the original version of the ACE 2005 corpus.

Table 4 shows detailed corpus statistics for the values assigned to the MODALITY attribute, both in the corpus as a whole and in the individual parts of the corpus. Overall, just over half of all events belong to the *Asserted* category. However, in the various sub-parts of the corpus, the proportions of *Asserted* events vary quite considerably. The highest percentages are found in the two types of news reports (i.e., *NW* and *BN*), with 56.4 and 60 % of events being *Asserted,* respectively. This is perhaps unsurprising, given that the purpose of these reports is to provide new information about events that are happening in the world. In the other sections of the corpus, which are generally concerned with discussing news stories rather than reporting on them, the general trend appears to be that the less formal the setting, the lower the number of asserted events. For example, in the *BC* portion of the corpus, which contains transcripts of conversations from CNN, 47.4 % of events are *Asserted*. This becomes even lower in the more informal settings of telephone conversations and discussion groups. The percentage is higher in weblogs, since these generally provide an overview of a particular topic.

In more informal interactions, the proportion of *Speculated* events becomes higher than the average over the complete corpus, since the focus is on discussing, interpreting and speculating about current affairs. Indeed, the percentage of *Speculated* events rises as high as 46.6 % in the UN texts, where there are around 10 % more *Speculated* events than *Asserted* events. This is in contrast to news reports (BN and NW), where speculation levels are very much lower, and are less than twice as numerous as *Asserted* events. It is interesting to note, however, that even these proportions of speculated events are still considerably higher than in scientific academic texts. In (Thompson et al. 2011b), it was found that only 8.1 % of events in abstracts of biomedical articles showed any degree of uncertainty. However, academic abstracts are a very different type of text, where authors mostly want to try to present their most certain results, in order to convince the reader of the validity of their work. News reports, on the other hand, aim to present the most relevant and up-to-date details about a particular story. This may include some less reliable, unverified information or rumours, possibly coming from multiple sources. It is important that such information is explicitly flagged as being uncertain, in order to retain credibility in the case that any of the information reported is later contradicted, when new details about the story are obtained.

In the corpus as a whole, just over one-sixth of the events are *Presupposed*, with proportions in the sub-parts ranging between about 14 and 23 %. In news reports, the reader/listener's attention is held by ensuring that the majority of the report asserts new details. In a smaller number of cases, events that are already known

Table 4 Statistics for the modality attribute (total counts and percentages)

Value	Overall	BN	BC	CTS	NW	UN	WL
Asserted	2753 (51.5)	710 (60.0)	433 (47.4)	202 (43.2)	878 (56.4)	257 (35.7)	273 (53.8)
Speculated	1569 (29.3)	289 (24.4)	308 (33.7)	155 (33.1)	326 (20.1)	335 (46.6)	156 (30.8)
Presupposed	941 (17.6)	172 (14.5)	162 (17.7)	107 (22.9)	326 (20.1)	101 (14.1)	73 (14.4)
Other	86 (1.6)	13 (1.1)	11 (1.2)	4 (0.9)	27 (1.7)	26 (3.6)	5 (1.0)

Table 5 Cues for speculated modality

Cue	Count	Cue	Count	Cue	Count
If	87	May	31	What	18
Would	72	Can	22	Believe	17
Could	60	Whether	21	Likely	17

about may be mentioned, in order to provide updates, or to provide context or background information about the news stories. In parts of the corpus concerned with discussions of news stories, the introduction of previously known information can also be important, as a stimulus for subsequent discussion, interpretations and evaluations of news stories.

In terms of specific cues that have been annotated, only cues for the *Speculated* category appear with any regularity. The most frequently annotated cues are shown in Table 5.

The high occurrence of words such as *if*, *would* and *whether* provides evidence that many speculated events occur within hypothetical contexts. Other events may occur in the context of questions (indicated by *what*), while modal auxiliaries such as *could, may* and *can,* together with related adverbs such as *likely,* show that there are also instances where the speculation relates to a degree of uncertainty about the truth of the event. Verbs that denote personal opinions, such as *believe* and *think,* tend to be more prevalent in the more informal text types, with the more formal or impersonal modal auxiliaries occurring with higher frequencies in news reports.

5.2 Subjectivity

As shown in Table 6, some sort of subjectivity is expressed for almost 1 in 5 events in the overall corpus. An interesting finding is that events are almost twice as likely to occur with negative subjectivity (11 % of all events) as with positive subjectivity (6 % of events). These proportions remain fairly stable in the different parts of the corpus, although events with negative subjectivity rise as high as 19 % in the WL section. Since weblogs usually represent personal takes on particular subjects, these are naturally more likely to contain more subjectivity than other text types, which may occasionally turn into "rants". The general trends shown in the results for subjectivity, however, provide evidence to support the age-old hypothesis that "bad news sells better than good news". Indeed, in a survey of news preferences, it was found that peoples' favourite subjects are war, weather, disaster, money and crime.[3]

We also observed that words with very negative connotations are often used instead of more neutral words, in order to help "sensationalise" a story. Examples can be seen in Table 7, which shows the most commonly annotated cues for positive and negative subjectivity. It should be noted that some of the most common negative subjectivity cues (e.g., *terrorism* and *genocide*) also act as the triggers of the corresponding events.

---

[3] http://www.guardian.co.uk/media/greenslade/2007/sep/04/thegoodnewsaboutbadnewsi.

Table 6 Statistics for the Subjectivity attribute (total counts and percentages)

Value	Overall	BN	BC	CTS	NW	UN	WL
Positive	319 (6.0)	56 (4.7)	59 (6.4)	27 (5.8)	82 (5.3)	52 (7.2)	43 (8.4)
Negative	591 (11.1)	84 (7.1)	115 (12.6)	44 (9.4)	166 (10.1)	86 (12.0)	86 (19.0)
Multi-valued	19 (0.3)	2 (0.1)	1 (0.1)	0 (0.0)	7 (0.4)	4 (0.5)	5 (1.0)
Neutral	4420 (82.6)	1042 (88.0)	739 (80.1)	397 (84.8)	1302 (83.6)	577 (80.2)	363 (71.6)

Table 7 Cues for positive and negative subjectivity

Negative				Positive			
Cue	Count	Cue	Count	Cue	Count	Cue	Count
Terrorism	18	Genocide	7	Support	10	Like	5
Not	14	Don't	7	Want	8	Interested	5
Against	12	No	7	Good	6	Trying	5
Opposed	9	Deadly	7	Hope	6	Wanted	5
Anti	8	Threat	6	Wants	6	Should	5

So, for example, *terrorism* or *terrorist attacks* will be used instead of the more neutral *attacks,* and *genocide* will be used instead of *killing.* Another way of intensifying the negative sentiments invoked by the mention of an event is to use strongly negative adjectives and adverbs, such as *deadly.* Examining the most commonly occurring negative subjectivity cues specifically for news reports reveals more of these, such as *fierce, bloody* and *horribly.* A further method is to use verbs with negative connotations as a means of reporting what people have said, the most commonly occurring examples including *threaten*, *condemn*, *warn* and *deny.*

The *multi-valued* subjectivity category, i.e., cases where events are reported with conflicting subjectivity values ascribed to the event by two (or more) different sources, is used very rarely, constituting less than 0.5 per cent of all events in the corpus. Nevertheless, the recognition of such cases may still be worthwhile, since they would be of interest to researchers looking for contradictory and opposing opinions.

To further investigate the expression of positive and negative subjectivity towards events, we analysed the correlations of these values with different *Modality* values. Tables 8 and 9 show the proportions of events with different *Modality* values that have been assigned the *Positive* and *Negative* subjectivity values, respectively. Looking at the tables, it can be observed that both *Positive* and *Negative* speculation are generally specified with reasonable frequency for *Speculated* events. That is to say, different types of opinions towards non-factual events are fairly easy to find. In contrast, *Positive* subjectivity is relatively rare amongst events with other *Modality* values. For instance, it is uncommon to find positive attitudes towards *Asserted* and *Presupposed* events, i.e., definite events that are known to be happening or to have happened. However, the figures in Table 9 illustrate that it is usually several times more likely for *Asserted* and *Presupposed* events to be marked with *Negative* than *Positive* subjectivity.

Table 8 Distribution of positive subjectivity events amongst different modalities

Value	Overall	BN	BC	CTS	NW	UN	WL
Asserted	66 (2.4)	16 (2.2)	11 (2.5)	4 (2.0)	12 (1.3)	14 (5.4)	9 (3.2)
Speculated	212 (13.5)	34 (11.8)	41 (13.3)	20 (12.9)	52 (15.9)	33 (9.9)	32 (20.5)
Presupposed	35 (3.7)	4 (2.3)	6 (3.7)	3 (2.8)	16 (4.9)	4 (3.9)	2 (2.7)
Other	6 (7.0)	2 (15.4)	1 (9.1)	0 (0.0)	2 (7.4)	1 (3.8)	0 (0.0)

Table 9 Distribution of negative subjectivity events amongst different modalities

Value	Overall	BN	BC	CTS	NW	UN	WL
Asserted	191 (6.9)	31 (4.4)	41 (9.5)	20 (9.9)	35 (4.0)	23 (8.9)	41 (15.0)
Speculated	225 (14.3)	33 (11.4)	53 (17.2)	11 (7.1)	51 (15.6)	46 (13.7)	31 (19.8)
Presupposed	169 (18.0)	19 (11.0)	20 (12.3)	13 (12.1)	76 (23.3)	17 (16.8)	24 (32.9)
Other	6 (7.0)	1 (7.7)	1 (9.1)	0 (0.0)	4 (14.8)	0 (0.0)	0 (0.0)

The percentages of events with *Negative* subjectivity are highest for *Presupposed* events in *NW,* where almost a quarter of such events have negative subjectivity expressed towards them, and *WL*, where the proportion rises to almost one-third. In *NW*, this could be due to the sensationalist nature of news stories, as explained above. In *WL*, writers are likely to express their own strong opinions. Interestingly, the proportion of *Presupposed* events with *Negative* subjectivity in the other type of news reports, i.e., *BN*, is less than half that in *NW*. Indeed, in general, there seems to be a lesser tendency to express negative subjectivity on *Presupposed* events in speech than in writing.

5.3 Source-type

In most cases (over 82 %), events are reported directly by the author or speaker, without mentioning a specific source, as shown in Table 10. Of the remaining events, those that represent information provided by people directly involved in the events in question are around twice as likely than information provided by uninvolved third parties. This pattern does seem logical—the most detailed, relevant and interesting information can usually be obtained from people directly involved in an event. However, such people may introduce some biased information into the discourse. Therefore, it is often a good idea to balance such details with information provided by experts or those people without direct involvement in the event.

Looking at the individual parts of the corpus reveals that the explicit identification of information source is particularly prevalent in newswire text, where events attributed to a particular source other than the author account for about 35 % of all events. The ratio of *Involved* to *Third Party* events remains about the same as the average over the complete corpus (i.e., about 2:1). Whilst a similar ratio holds for the other part of the corpus that constitutes news reports (i.e., *BN*), the

Table 10 Statistics for the source-type attribute (total counts and percentages)

Value	Overall	BN	BC	CTS	NW	UN	WL
Author	4401 (82.3)	1034 (87.3)	815 (89.2)	462 (98.7)	1007 (64.7)	669 (93.0)	414 (81.2)
Third party	329 (6.2)	66 (5.6)	32 (3.5)	5 (1.1)	169 (10.1)	16 (2.2)	41 (8.1)
Involved	619 (11.6)	84 (7.1)	67 (7.3)	1 (0.2)	381 (24.5)	34 (4.7)	52 (10.3)

absolute proportions of events with a SOURCE-TYPE other than *Author* are much lower in BN than for newswire, constituting only about 13 % of all events in this portion of the corpus. That is to say, events attributed to non-author sources are only about one-third as numerous as in newswire texts. Thus, there seems to be a noticeable divergence in the norms of how news is reported in speech or in writing.

The proportions of events with a non-author source are much lower in the parts of the corpus that contain discussions. Whilst in BC (which is from the CNN channel), the proportion is not much lower than for broadcast news (around 10 %), this falls to about 7 % in discussion groups, and only 1.3 % in conversational telephone speech. In contrast, in WL, the proportion is quite high (about 18 % of all events), with roughly equal numbers of *Involved* and *Third Party* events. This may be due to weblogs covering a topic in detail, and from multiple points of view.

5.4 Polarity

The results in Table 11 show that just under 4 % of events in the corpus are explicitly negated. There are very few variations amongst the different text types in the corpus (mostly ±0.5 % difference from this average). This small percentage is probably due to the fact that the purpose of the various texts and transcripts that make up the corpus is to report on and discuss things that have happened, rather than things that have not happened. The highest percentage of negated events by a small margin occurs in WL (4.7 % of events in this part of the corpus), possibly because their purpose is often to discuss a topic in detail, which may involve introducing negative as well as positive information. In comparison, approximately 50 % more events are negated in biomedical abstracts than in the ACE 2005 corpus (6.1 % in total) (Thompson et al. 2011b). One reason for this is that in biomedical text, it can sometimes be the case that a negative result can be more significant than a positive one (Knight 2003).

Table 11 Statistics for the polarity attribute (total counts and percentages)

Value	Overall	BN	BC	CTS	NW	UN	WL
Positive	5143 (96.1)	1140 (96.3)	879 (92.2)	452 (96.6)	1494 (96.0)	695 (96.6)	483 (95.3)
Negative	206 (3.9)	44 (3.7)	35 (3.8)	16 (3.4)	63 (4.0)	24 (3.3)	24 (4.7)

Table 12 Distribution of negated events amongst different modalities

Value	Overall	BN	BC	CTS	NW	UN	WL
Asserted	104 (3.8)	25 (3.5)	21 (4.8)	8 (3.9)	33 (3.8)	6 (2.3)	11 (4.0)
Speculated	95 (6.1)	18 (6.2)	14 (4.5)	7 (4.5)	27 (8.3)	16 (4.8)	13 (8.3)
Presupposed	3 (0.3)	1 (0.6)	0 (0.0)	0 (0.0)	2 (0.6)	0 (0.0)	0 (0.0)
Other	4 (4.7)	0 (0.0)	0 (0.0)	1 (25.0)	1 (3.7)	2 (7.7)	0 (0.0)

Table 13 Statistics for the Genericity attribute (total counts and percentages)

Value	Overall	BN	BC	CTS	NW	UN	WL
Generic	1028 (19.2)	137 (11.6)	245 (26.8)	110 (23.5)	163 (10.5)	245 (34.1)	128 (25.2)
Specific	4321 (80.8)	1047 (88.4)	669 (73.2)	358 (76.5)	1394 (89.5)	474 (65.9)	379 (74.8)

We also analysed how negated events are distributed amongst events with differing *Modality* values (Table 12). There are few major differences amongst the different portions of the corpus, with negated events generally around twice as likely to occur on *Speculated* than *Asserted* events. This is consistent with what was stated above, that in terms of definite events, it is much more common to state things that have happened, than things that did not happen. For a similar reason, negated *Presupposed* events are almost non-existent.

5.5 Genericity

As shown in Table 13, around four fifths of events in the corpus describe specific occurrences, whilst the remaining fifth describe generic situations. However, within the specific sections of the corpus, there are quite large variations in the distributions. The largest proportions of *Specific* events (almost 90 %) are to be found in the two types of news reports, whose main purpose is to provide information about specific events that have occurred in the recent past. In contrast, text types that contain more discussion are likely to contain general topics as well as specific events. This helps to explain why, in the remaining parts of the corpus, the proportion of *Generic* events is over 20 % in all cases, rising as high as one-third of all events in the UN corpus portion.

We observed that, on the basis of our detailed annotation guidelines, we were able to identify almost 200 more *Specific* events than were annotated in the original ACE 2005 corpus. This finding supports our decision to re-annotate the GENERICITY attribute.

5.6 Tense

Table 14 shows that over half of the events in the corpus are explicitly marked as having taken place in the past, with the highest proportions (around 60 %) in the two types of news reports and WL, whose articles are specifically focussed on reporting and summarising past events. The lowest percentage of past events is to be

Table 14 Statistics for the Tense attribute in the corpus (total counts and percentages)

Value	Overall	BN	BC	CTS	NW	UN	WL
Past	2996 (56.0)	710 (60.0)	429 (47.0)	236 (50.4)	965 (61.2)	347 (48.2)	309 (60.1)
Present	528 (9.9)	140 (11.8)	148 (16.1)	62 (13.2)	125 (8.0)	25 (3.5)	28 (5.5)
Future	865 (16.1)	183 (15.4)	152 (16.6)	63 (13.5)	291 (18.7)	102 (14.2)	74 (14.6)
Unspecified	960 (17.9)	151 (12.7)	185 (20.2)	107 (22.8)	176 (11.3)	245 (34.1)	96 (18.1)

found in the BC part of the corpus. It is also in this part of the corpus that the highest proportion of *Present* events is to be found. Indeed, it appears to be a general trend that *Present* events are more prominent in spoken communication than in written communication. This may be due to the fact that in "live" discussion situations, there is more of a tendency to talk about situations that are currently on-going, whilst in written discussion tends to consider things that have already happened. This is supported by the figures for the UN and WL parts of the corpus, which show that on-going events are mentioned very infrequently (around 5 % of events or less).

For events with *Unspecified* tense, there is quite a large amount of variation in the different parts of the corpus, ranging from 12.7 % in the BN portion to 34.1 % in UN. The proportions of *Unspecified* events correlate closely with the proportions of *Generic* events, which seems reasonable: discussions about generally occurring or habitual events are much less likely to be associated with tense information.

It is important to note that the overall number of *Unspecified* events in the updated corpus is almost half of that in the original ACE 2005 corpus. Therefore, the re-annotation of the values of the TENSE attribute was worthwhile.

6 Conclusion

In this paper, we have discussed how meta-knowledge information has a significant impact on the interpretation of events. Therefore, the automatic recognition of such information is important to allow the development of sophisticated and accurate NLP systems. We took the ACE 2005 corpus as our starting point, whose annotation scheme identifies events and encodes some basic aspects of event interpretation. We subsequently extended this scheme to encode a number of other aspects of meta-knowledge, by considering both domain-independent and domain-relevant features of news-related text. We created new annotation guidelines and enriched all 5349 events in the ACE 2005 corpus according to this scheme.

Our annotation effort has not only added new meta-knowledge attributes to the events, but has also identified textual evidence for their assignment (i.e., cues), which has previously been shown to be important for the automated recognition of meta-knowledge information. We verified the soundness and robustness of the scheme through double-annotation of a portion of the corpus and subsequent calculation of inter-annotator agreement, which ranged from 0.530 to 0.871 κ, according to attribute. Subsequent discussion and investigation of the attributes with

lower levels of agreement showed that the majority of discrepancies corresponded to systematic errors that were straightforward to correct.

We performed an analysis of the corpus, both as a whole and by considering the parts collected from different data sources separately. This analysis revealed a number of interesting differences in the meta-knowledge features of events, according both to the formality of the setting (e.g., formal news reports versus more informal discussions of news stories) and to whether the material is written or spoken.

As further work, we are developing a machine learning system that makes use of the enriched meta-knowledge information and associated cues to predict richer information relating to the interpretation of events. This will be used in the development an enhanced version of our semantic search system over news archives.

References

Ahn, D. (2006). The stages of event extraction. In *Proceedings of the workshop on annotating and reasoning about time and events* (pp. 1–8).

Ananiadou, S., Thompson, P., Nawaz, R., McNaught, J., & Kell, D. B. (2015). Event-based text mining for biology and functional genomics. *Briefings in Functional Genomics, 14*(3), 213–230. doi: 10.1093/bfgp/elu015.

Anick, P., & Bergler, S. (1992). Lexical structures for linguistic inference. *Lexical semantics and knowledge representation* (pp. 121–135). New York: Springer.

Aone, C., & Ramos-Santacruz, M. (2000). REES: A large-scale relation and event extraction system. In *Proceedings of the sixth conference on applied natural language processing* (pp. 76–83).

Balahur, A., Steinberger, R., Kabadjov, M. A., Zavarella, V., Van Der Goot, E., Halkia, M., et al. (2010). Sentiment analysis in the news. In *Proceedings of the 7th language resources and evaluation conference* (pp. 2216–2220).

Banfield, A. (1982). *Unspeakable sentences: Narration and representation in the language of fiction.* Abingdon: Routledge.

Bautin, M., Vijayarenu, L., & Skiena, S. (2008). International sentiment analysis for news and blogs. In *Proceedings of the international conference on weblogs and social media* (pp. 19–26).

Bergler, S. (2006). Conveying attitude with reported speech. *Computing attitude and affect in text: Theory and applications* (pp. 11–22). New York: Springer.

Bjorne, J., & Salakoski, T. (2013). TEES 2.1: Automated annotation scheme learning in the BioNLP 2013 Shared Task. In *Proceedings of the BioNLP shared task 2013 workshop* (pp. 16–25).

Carlson, L., Marcu, D., & Okurowski, M. E. (2003). Building a discourse-tagged corpus in the framework of rhetorical structure theory. *Current and new directions in discourse and dialogue* (pp. 85–112). New York: Springer.

Chen, Z., Ji, H., & Haralick, R. (2009). A pairwise event coreference model, feature impact and evaluation for event coreference resolution. In *Proceedings of the workshop on events in emerging text types* (pp. 17–22).

Deng, L., Choi, Y., & Wiebe, J. (2013). Benefactive/malefactive event and writer attitude annotation. In *Proceedings of ACL* (pp. 120–125).

Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-scale sentiment analysis for news and blogs. In *Proceedings of the international conference on weblogs and social media*.

Grishman, R., & Sundheim, B. (1996). Message understanding conference-6: A brief history. In *Proceedings of the 16th international conference on computational linguistics* (*COLING'96*) (pp. 466–471).

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language Resources and Evaluation, 69*, 274–307.

Hirohata, K., Okazaki, N., Ananiadou, S., & Ishizuka, M. (2008). Identifying sections in scientific abstracts using conditional random fields. In *Proceedings of the 3rd international joint conference on natural language processing* (pp. 381–388).

Hripcsak, G., & Rothschild, A. S. (2005). Agreement, the f-measure, and reliability in information retrieval. *Journal of the Americal Medical Informatics Association, 12*(3), 296–298.

Hyland, K. (1996). Talking to the academy: Forms of hedging in science research articles. *Written Communication, 13*(2), 251–281.

Ji, H., & Grishman, R. (2008). Refining event extraction through cross-document inference. In *Proceedings of ACL* (pp. 254–262).

Kessler, J. S., Eckert, M., Clark, L., & Nicolov, N. (2010). The ICWSM 2010 JDPA sentiment corpus for the automotive domain. In *International AAAI conference on weblogs and social media data challenge workshop*.

Knight, J. (2003). Negative results: Null and void. *Nature, 422*(6932), 554–555.

Liakata, M., Thompson, P., de Waard, A., Nawaz, R., Maat, H. P., & Ananiadou, S. (2012). A three-way perspective on scientific discourse annotation for knowledge extraction. In *Proceedings of the workshop on detecting structure in scholorly discourse (DSSD)* (pp. 37–46).

Liao, T., Liu, Z., & Wang, X. (2013). Research and implementation on event-based method for automatic summarization. In *Proceedings of the eighth international conference on bio-inspired computing: Theories and applications (BIC-TA)* (pp. 103–111).

Light, M., Qiu, X. Y., & Srinivasan, P. (2004). The language of bioscience: Facts, speculations, and statements in between. In *Proceedings of the BioLink 2004 workshop at HLT/NAACL* (pp. 17–24).

Marcu, D., & Echihabi, A. (2002). An unsupervised approach to recognizing discourse relations. In *Proceedings of the 40th annual meeting on association for computational linguistics* (pp. 368–375).

Medlock, B., & Briscoe, T. (2007). Weakly supervised learning for hedge classification in scientific literature. In *Proceedings of ACL* (pp. 992–999).

Miwa, M., Thompson, P., & Ananiadou, S. (2012a). Boosting automatic event extraction from the literature using domain adaptation and coreference resolution. *Bioinformatics, 28*(13), 1759–1765. doi:10.1093/bioinformatics/bts237.

Miwa, M., Thompson, P., Korkontzelos, I., & Ananiadou, S. (2014). Comparable study of event extraction in newswire and biomedical domains. In *Proceedings of COLING* (pp. 2270–2279).

Miwa, M., Thompson, P., McNaught, J., Kell, D. B., & Ananiadou, S. (2012b). Extracting semantically enriched events from biomedical literature. *BMC Bioinformatics, 13*(1), 108.

Miyao, Y., Ohta, T., Masuda, K., Tsuruoka, Y., Yoshida, K., Ninomiya, T., et al. (2006). Semantic retrieval for the accurate identification of relational concepts in massive textbases. In *Proceedings of ACL* (pp. 1017–1024).

Morante, R., & Sporleder, C. (Eds.). (2012). *Proceedings of the workshop on extra-propositional aspects of meaning in computational linguistics*: Association for Computational Linguistics.

Nawaz, R., Thompson, P., & Ananiadou, S. (2010a). Evaluating a meta-knowledge annotation scheme for bio-events. *Proceedings of the workshop on negation and speculation in natural language processing (NeSp-NLP 2010), ACL 2010* (pp. 69–77). Sweden: Uppsala.

Nawaz, R., Thompson, P., & Ananiadou, S. (2012a). Identification of manner in bio-events. In *Proceedings of the eighth international conference on language resources and evaluation (LREC 2012)* (pp. 3505–3510).

Nawaz, R., Thompson, P., & Ananiadou, S. (2012b). Meta-knowledge annotation at the event level: Comparison between abstracts and full papers. In *Proceedings of the third LREC workshop on building and evaluating resources for biomedical text mining (BioTxtM 2012)* (pp. 24–21).

Nawaz, R., Thompson, P., & Ananiadou, S. (2013a). Negated bio-events: Analysis and identification. *BMC Bioinformatics, 14*, 14.

Nawaz, R., Thompson, P., & Ananiadou, S. (2013b). Something old, something new: Identifying knowledge source in bio-events. *International Journal of Computational Linguistics and Applications, 4*(1), 129–144.

Nawaz, R., Thompson, P., & Ananiadou, S. (2013c). Towards event-based discourse analysis of biomedical text. *International Journal of Computational Linguistics and Applications, 4*(2), 101–120.

Nawaz, R., Thompson, P., McNaught, J., & Ananiadou, S. (2010b). Meta-knowledge annotation of bio-events. In *Proceedings of the 7th international conference on language resources and evaluation (LREC-2010),* 17–23 May (pp. 2498–2507).

Nédellec, C., Bossy, R., Kim, J.-D., Kim, J.-J., Ohta, T., Pyysalo, S., et al. (2013). Overview of BioNLP shared task 2013. In *BioNLP Shared Task 2013 Workshop in ACL 2013 Sofia* (pp. 1–7). 9 August 2013.

Pareti, S. (2012a). A database of attribution relations. In *Proceedings of LREC* (pp. 3213–3217).

Pareti, S. (2012b). The independent encoding of attribution relations. In *Proceedings of the eighth joint ACL-ISO workshop on interoperable semantic annotation (ISA-8)*.

Pareti, S., & Prodanof, I. (2010). Annotating attribution relations: Towards an Italian discourse treebank. In *Proceedings of LREC* (pp. 3566–3571).

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

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. K., et al. (2008). The Penn discourse treebank 2.0. In *Proceedings of LREC* (pp. 2961–2968).

Prasad, R., McRoy, S., Frid, N., Joshi, A., & Yu, H. (2011). The biomedical discourse relation bank. *BMC Bioinformatics, 12*, 188.

Prince, E. F. (1992). The ZPG letter: Subjects, definiteness, and information-status. In W. C. Mann, & S. A. Thompson (Eds.), *Discourse description: Diverse analyses of a fund raising text* (pp. 295–325). Amsterdam: John Benjamins.

Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., et al. (2003). The TimeBank corpus. In *Proceedings of corpus linguistics* (pp. 647–656).

Quirk, R. (1985). *A comprehensive grammar of the english language*. Harlow: Longman Publishing House.

Riloff, E., Wiebe, J., & Phillips, W. (2005). Exploiting subjectivity classification to improve information extraction. In *Proceedings of the national conference on artificial intelligence* (pp. 1106–1111).

Rizomilioti, V. (2006). Exploring epistemic modality in academic discourse using corpora. In E. Arnó Macià, A. Soler Cervera, & C. Rueda Ramos (Eds.), *Information technology in languages for specific purposes* (pp. 53–71). New York: Springer.

Rubin, V. L. (2007). Stating with certainty or stating with doubt: Intercoder reliability results for manual annotation of epistemically modalized statements. In *Proceedings of NAACL-HLT* (pp. 141–144).

Rubin, V. L. (2010). Epistemic modality: From uncertainty to certainty in the context of information seeking as interactions with texts. *Information Processing and Management, 46*(5), 533–540.

Rubin, V., Liddy, E., & Kando, N. (2006). *Certainty identification in texts: Categorization model and manual tagging results* (pp. 61–76). Computing attitude and affect in text: Theory and applications.

Russo, I., & Caselli, T. (2013). changeable polarity of verbs through emotions' attribution in crowdsourcing experiments. In *Proceedings of the first international workshop on emotion and sentiment in social and expressive media: Approaches and perspectives from AI (ESSEM 2013)* (pp. 131–139).

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

Stoyanov, V., Cardie, C., & Wiebe, J. (2005). Multi-perspective question answering using the OpQA corpus. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 923–930).

Strassel, S., Przybocki, M. A., Peterson, K., Song, Z., & Maeda, K. (2008). Linguistic resources and evaluation techniques for evaluation of cross-document automatic content extraction. In *Proceedings of the 6th language resources and evaluation conference* (pp. 2706–2709).

Teufel, S., & Moens, M. (2000). What's yours and what's mine: Determining intellectual attribution in scientific text. In *Proceedings of the 2000 joint SIGDAT conference on empirical methods in natural language processing and very large corpora* (pp. 9–17).

Thompson, P., McNaught, J., Montemagni, S., Calzolari, N., Del Gratta, R., Lee, V., et al. (2011a). The BioLexicon: A large-scale terminological resource for biomedical text mining. *BMC Bioinformatics, 12*(1), 397.

Thompson, P., Nawaz, R., Korkontzelos, I., Black, W., McNaught, J., & Ananiadou, S. (2013). News search using discourse analytics. In *Proceedings of the digital heritage international congress* (pp. 597–604).

Thompson, P., Nawaz, R., McNaught, J., & Ananiadou, S. (2011b). Enriching a biomedical event corpus with meta-knowledge annotation. *BMC Bioinformatics, 12*(393), 1.

Thompson, P., Venturi, G., McNaught, J., Montemagni, S., & Ananiadou, S. (2008). Categorising modality in biomedical texts. *Proceedings of the LREC 2008 workshop on building and evaluating resources for biomedical text mining* (pp. 27–34). Morocco: Marrakech.

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine, 37*(5), 360–363.

Vincze, V., Szarvas, G., Farkas, R., Mora, G., & Csirik, J. (2008). The BioScope corpus: Biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics, 9*(Suppl. 11), S9.

Walker, C., Strassel, S., Medero, J., & Maeda, K. (2006). *ACE 2005 multilingual training corpus*. Philadelphia: Linguistic Data Consortium.

Webber, B., Egg, M., & Kordoni, V. (2012). Discourse structure and language technology. *Natural Language Engineering, 18*(4), 437–490.

Wiebe, J. (1994). Tracking point of view in narrative. *Computational Linguistics, 20*(2), 233–287.

Wiebe, J., Breck, E., Buckley, C., Cardie, C., Davis, P., Fraser, B., et al. (2003). Recognizing and organizing opinions expressed in the World Press. In *Proceedings of the AAAI spring symposium on new directions in question answering* (pp. 12–19).

Wiebe, J., & Deng, L. (2014). A conceptual framework for inferring implicatures. In *Proceedings of the 5th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 154–159).

Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M. (2004). Learning subjective language. *Computational Linguistics, 30*(3), 277–308.

Wiebe, J., Wilson, T., & Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation, 39*(2–3), 165–210.

Wilbur, W. J., Rzhetsky, A., & Shatkay, H. (2006). New directions in biomedical text annotations: Definitions, guidelines and corpus construction. *BMC Bioinformatics, 7*, 356.

Yi, J., Nasukawa, T., Bunescu, R., & Niblack, W. (2003). Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In *Proceedings of the third IEEE international conference on data mining* (pp. 427–434).