Evaluating a Meta-Knowledge Annotation Scheme for Bio-Events

Raheel Nawaz[1] **Paul Thompson**[1,2] **Sophia Ananiadou**[1,2]
[1]School of Computer Science, University of Manchester, UK
[2]National Centre for Text Mining, University of Manchester, UK
E-mail: nawazr@cs.man.ac.uk, paul.thompson@manchester.ac.uk,
sophia.ananiadou@manchester.ac.uk

Abstract

The correct interpretation of biomedical texts by text mining systems requires the recognition of a range of types of high-level information (or *meta-knowledge*) about the text. Examples include expressions of negation and speculation, as well as pragmatic/rhetorical intent (e.g. whether the information expressed represents a hypothesis, generally accepted knowledge, new experimental knowledge, etc.) Although such types of information have previously been annotated at the text-span level (most commonly sentences), annotation at the level of the event is currently quite sparse. In this paper, we focus on the evaluation of the multi-dimensional annotation scheme that we have developed specifically for enriching bio-events with meta-knowledge information. Our annotation scheme is intended to be general enough to allow integration with different types of bio-event annotation, whilst being detailed enough to capture important subtleties in the nature of the meta-knowledge expressed in the text. To our knowledge, our scheme is unique within the field with regards to the diversity of meta-knowledge aspects annotated for each event, whilst the evaluation results have confirmed its feasibility and soundness.

1 Introduction

The ability to recognise high-level information (or meta-knowledge) relating to the interpretation of texts is an important task for text mining systems. There are several types of meta-knowledge that fall under this category. For example, the detection of expressions of speculation and negation is important across all domains, although the way in which these phenomena are expressed may be domain-specific. In scientific texts, it is also important to be able to determine other types of information, such as the author's rhetorical/pragmatic intent (de Waard et al., 2009). This would correspond to whether the information expressed represents a hypothesis, accepted knowledge, new experimental knowledge, etc.

The ability to distinguish between these different types of information can be important for tasks such as building and updating models of biological processes, like pathways (Oda et al., 2008), and curation of biological databases (Ashburner et al., 2000). Central to both of these tasks is the identification of *new knowledge* that can enhance these resources, e.g. to build upon an existing, but incomplete model of a biological process (Lisacek et al., 2005) or to ensure that the database is kept up to date. Any new knowledge added should be supported though evidence, which could include linking hypotheses with experimental findings. It is also important to take into account inconsistencies and contradictions reported in the literature.

The production of annotated corpora can help to train text mining systems to recognise types of meta-knowledge, such as the above. Although a number of such corpora have already been produced, different annotation schemes are required according to the exact domain under consideration, as well as the types of task that will be undertaken by the text mining system.

The work described in this paper is focused on the design and evaluation of the meta-knowledge annotation scheme described in Nawaz et al., (2010). The annotation scheme has been specifically designed to recognise a range of meta-knowledge types for events extracted from biomedical texts (henceforth *bio-events*). The aim is to facilitate the development of more useful systems in the context of various biomedical information extraction (IE) and textual inference (TI) tasks. Although the scheme has been designed

Proceedings of the Workshop on Negation and Speculation in Natural Language Processing, pages 69–77,
Uppsala, July 2010.

for application to existing bio-event corpora, it is intended to be applied to any type of bio-relation corpora, and can easily be tailored for other types of relations/events within the domain.

1.1 Bio-Event Representation of Text

Searching for relevant information in electronic documents is most commonly carried out by entering keywords into a search engine. However, such searches will normally return a huge number of documents, many of which will be irrelevant to the user's needs.

A more promising and efficient way of searching is over *events* that have been extracted from texts through the application of natural language processing methods. An event is a structured representation of a certain piece of information contained within the text, which is usually anchored to a particular word in the text (typically a verb or noun) that is central to the description of the event. Events are often represented by a template-like structure with slots that are filled by the event participants. Each event participant is also assigned a role within the event. These participants can be entities, concepts or even other events. This kind of event representation allows the information contained in a text to be represented as a collection of *nested* events.

A *bio-event* is an event specialised for the biomedical domain. Kim et al. (2008) define a bio-event as a dynamic bio-relation involving one or more participants. These participants can be bio-entities or (other) bio-events, and are each assigned a semantic role/slot like *theme* and *cause* etc. Each bio-event is typically assigned a type/class from a chosen bio-event taxonomy/ontology, e.g., the GENIA Event Ontology (Kim et al., 2008). Similarly, the bio-entities are also assigned types/classes from a chosen bio-term taxonomy/ontology, e.g., the Gene Ontology (Ashburner et al., 2000).

As an example, consider the simple sentence shown in Figure 1**.**

> *The results suggest that the narL gene product activates the nitrate reductase operon.*

Figure 1. A Simple Sentence from a Biomedical Abstract

This sentence contains a single bio-event, anchored to the verb *activates*. Figure 2 shows a typical structured representation of this bio-event.

The fact that the verb is anchored to the verb *activates* allows the event-type of *positive regulation* to be assigned. The event has two slots, i.e. *theme* and *cause* whose labels help to characterise the contribution that the slot filler makes towards the meaning of the event. In this case, the slots are filled by the subject and object of the verb *activates*, both of which correspond to different types of bio-entities (i.e. *operon* and *protein).*

EVENT-TRIGGER: *activates*
EVENT-TYPE: positive_regulation
THEME: *nitrate reductase operon:* operon
CAUSE: *narL gene product*: protein

Figure 2. Typical Structured Representation of the Bio-Event mentioned in Figure 1

IE systems trained to extract bio-events from texts allow users to formulate semantic queries over the extracted events. Such queries can specify semantic restrictions on the events in terms of event types, semantic role labels and named entity types etc. (Miyao et al., 2006), in addition to particular keywords. For example, it would be possible to search only for those texts containing bio-events of type *negative_regulation* where the cause is an entity of type *protein*. Such queries provide a great deal more descriptive power than traditional keyword searches over unstructured documents. Biomedical corpora that have been manually annotated with event level information (e.g., Pyysalo et al., 2007; Kim et al., 2008; Thompson et al., 2009) facilitate the training of systems such as those described above.

Whilst event-based querying has advantages for efficient searching, the extracted events have little practical use if they are not accompanied by meta-knowledge information to aid in their interpretation.

1.2 Existing Meta-knowledge Annotation

Various corpora of biomedical literature (abstracts and/or full papers) have been produced that feature some degree of meta-knowledge annotation. These corpora vary in both the richness of the annotation added, and the type/size of the units at which the meta-knowledge annotation has been performed. Taking the unit of annotation into account, we can distinguish between annotations that apply to continuous text-spans, and annotations that have been performed at the event level.

Text-Span Annotation: Such annotations have mostly been carried out at the sentence level. They normally concentrate on a single aspect (or

dimension) of meta-knowledge, normally either speculation/certainty level, (e.g., Light et al., 2004; Medlock & Briscoe, 2007; Vincze et al., 2008) or general information content/rhetorical intent, e.g., *background, methods*, *results*, *insights*. This latter type of annotation has been attempted both on abstracts, (e.g., McKnight & Srinivasan, 2003; Ruch et al., 2007) and full papers, (e.g. Teufel et al., 1999; Langer et al., 2004; Mizuta & Collier, 2004), with the number of distinct annotation categories varying between 4 and 14.

Despite the availability of these corpora, annotation at the sentence level can often be too granular. In terms of information content, a sentence may describe, for example, both an experimental method and its results. The situation becomes more complicated if a sentence contains an expression of speculation. If this is only marked at the sentence level, there may be confusion about which part(s) of the sentence are affected by the speculative expression.

Certain corpora and associated systems have attempted to address these issues. The BioScope corpus (Vincze et al., 2008) annotates the scopes of negative and speculative keywords, whilst Morante & Daelemans (2009) have trained a system to undertake this task. The scheme described by Wilbur et al. (2006) applies annotation to *fragments* of sentences, which are created on the basis of changes in the meta-knowledge expressed. The scheme consists of multiple annotation dimensions which capture aspects of both certainty and rhetorical/pragmatic intent, amongst other things. Training a system to automatically annotate these dimensions is shown to be highly feasible (Shatkay et al., 2008).

Event-Level Annotation: Explicit annotation of meta-knowledge at the event-level is currently rather minimal within biomedical corpora. Whilst several corpora contain annotations to distinguish positive and negative events (e.g. Sanchez-Graillet & Poesio, 2007; Pyysalo et al., 2007), the annotation of the GENIA Event Corpus (Kim et al., 2008) is slightly more extensive, in that it additionally annotates certainty level. To our knowledge, no existing bio-event corpus has attempted annotation that concerns rhetorical/pragmatic intent.

1.3 The Need for an Event-Centric Meta-Knowledge Annotation Scheme

In comparison to meta-knowledge annotation carried out at the text-span level, the amount of annotation carried out at the event level is quite sparse. The question thus arises as to whether it is possible to use systems trained on text-span annotated corpora to assign meta-knowledge to bio-events, or whether new annotation at the event level is required.

Some corpora seem better suited to this purpose than others – whilst sentence-level annotations are certainly too granular for an event-centric view of the text, sentence fragments, such as those identified by Wilbur et al. (2006), are likely to correspond more closely to the extent of text that describes an event and its slots. Likewise, knowing the scopes of negative and speculative keywords within a sentence may be a useful aid in determining whether they affect the interpretation of a particular event.

However, the information provided in these corpora is still not sufficiently precise for event-level meta-knowledge annotation. Even within a text fragment, there may be several different bio-events, each with slightly different meta-knowledge interpretations. In a similar way, not all events that occur within the scope of a negation or speculation keyword are necessarily affected by it.

Based on these observations, we have developed a meta-knowledge annotation scheme that is specifically tailored to bio-events. Our scheme annotates various different aspects or dimensions of meta-knowledge. A close examination of a large number of relevant bio-events has resulted in a scheme that has some similarities to previously proposed schemes, but has a number of differences that seem especially relevant when dealing with events, e.g. the annotation of the manner of the event. The scheme is intended to be general enough to allow integration with existing bio-event annotation schemes, whilst being detailed enough to capture important subtleties in the nature of the meta-knowledge expressed about the event.

1.4 Lexical Markers of Meta-Knowledge

Most of the existing corpora mentioned above annotate text spans or events with particular categories (e.g. certainty level or general information type) in different meta-knowledge dimensions. However, what they do not normally do is to annotate lexical clues or keywords used to determine the correct values.

A number of previous studies have demonstrated the importance of lexical markers (i.e., words or phrases) that can accompany statements in scientific articles in determining the intended

interpretation of the text (e.g. Hyland, 1996; Rizomilioti 2006). We also performed a similar study (Thompson et al., 2008) although, in contrast to other studies, we took a multi-dimensional approach to the categorisation of such lexical items, acknowledging that several types of important information may be expressed through different words in the same sentence. As an example, let us consider the example sentence in Figure 3.

The DNA-binding properties of mutations at positions 849 and 668 may indicate that the catalytic role of these side chains is associated with their interaction with the DNA substrate.

Figure 3. Example Sentence

The author's pragmatic/rhetorical intent towards the statement that *the catalytic role of these side chains is associated with their interaction with the DNA substrate* is encoded by the word *indicate*, which shows that the statement represents an analysis of the evidence stated at the beginning of the sentence, i.e., that the mutations at positions 849 and 668 have DNA-binding properties. Furthermore, the author's *certainty level* (i.e., their degree of confidence) towards this analysis is shown by the word *may*. Here, the author is uncertain about the validity of their analysis.

Whilst our previous work served to demonstrate that the different aspects of meta-knowledge that can be specified lexically within texts require a multi-dimensional analysis to correctly capture their subtleties, it showed that the presence of particular lexical items is not the only important feature for determining meta-knowledge categories. In particular, their presence does not guarantee that the "expected" interpretation can be assumed (Sándor, 2007). In addition, not all types of meta-knowledge are indicated through explicit markers. Mizuta & Collier (2004) note that *rhetorical zones* may be indicated not only through explicit lexical markers, but also through features such as the main verb in the clause and the position of the sentence within the article or abstract.

For these reasons, we perform annotation on *all* relevant instances, regardless of the presence of lexical markers. This will allow systems to be trained that can learn to determine the correct meta-knowledge category, even when lexical markers are not present. However, due to the proven importance of lexical markers in determining certain meta-knowledge dimensions, our annotation scheme annotates such markers, whenever they are present.

2 Annotation Scheme

The annotation scheme we present here is a slightly modified version of our original meta-knowledge annotation scheme (Nawaz et al., 2010). The modified scheme consists of five meta-knowledge dimensions, each with a set of complete and mutually-exclusive categories, i.e., any given bio-event belongs to exactly one category in each dimension. Our chosen set of annotation dimensions has been motivated by the major information needs of biologists discussed earlier, i.e., the ability to distinguish between different intended interpretations of events.

In order to minimise the annotation burden, the number of possible categories within each dimension has been kept as small as possible, whilst still respecting important distinctions in meta-knowledge that have been observed during our corpus study.

The advantage of using a multi-dimensional scheme is that the interplay between different values of each dimension can reveal both subtle and substantial differences in the types of meta-knowledge expressed in the surrounding text. Therefore, in most cases, the exact rhetorical/pragmatic intent of an event can only be determined by considering a combination of the values of different dimensions. This aspect of our scheme is further discussed in section 3.

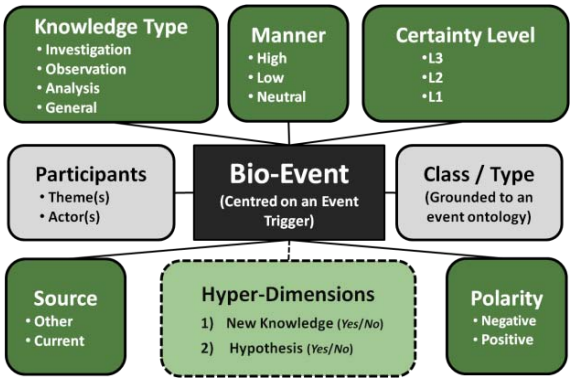


Figure 4. Bio-Event Annotation

Figure 4 provides an overview of the annotation scheme. The boxes with the light-coloured (grey) background correspond to information that is common to most bio-event annotation schemes, i.e., the participants in the event, together with an indication of the class or type of

the event. The boxes with the darker (green) backgrounds correspond to our proposed meta-knowledge annotation dimensions and their possible values. The remainder of this section provides brief details of each annotation dimension.

2.1 Knowledge Type (KT)

This dimension is responsible for capturing the general information content of the event. Whilst less detailed than some of the previously proposed sentence-level schemes, its purpose is to form the basis of distinguishing between the most critical types of rhetorical/pragmatic intent, according to the needs of biologists. Each event is thus classified into one of the following four categories:

Investigation: Enquiries or investigations, which have either already been conducted or are planned for the future, typically marked by lexical clues like *examined*, *investigated* and *studied*, etc.

Observation: Direct observations, often represented by lexical clues like *found*, *observed* and *report*, etc. Simple past tense sentences typically also describe observations. Such events represent experimental knowledge.

Analysis: Inferences, interpretations, speculations or other types of cognitive analysis, typically expressed by lexical clues like *suggest*, *indicate*, *therefore* and *conclude* etc. Such events, if they are interpretations or reliable inferences based on experimental results, can also constitute another type of (indirect) experimental knowledge. Weaker inferences or speculations, however, may be considered as hypotheses which need further proof through experiments.

General: Scientific facts, processes, states or methodology. This is the default category for the knowledge type dimension.

2.2 Certainty Level (CL)

The value of this dimension is almost always indicated through the presence/absence of an explicit lexical marker. In scientific literature, it is normally only applicable to events whose *KT* corresponds either to *Analysis* or *General*. In the case of *Analysis* events, *CL* encodes confidence in the truth of the event, whilst for *General* events, there is a temporal aspect, to account for cases where a particular process is explicitly stated to occur most (but not all) of the time, using a marker such as *normally*, or only occasionally, using a marker like *sometimes.* Events corresponding to direct *Observations* are not open to judgements of certainty, nor are *Investigation* events, which refer to things which have not yet happened or have not been verified.

Regarding the choice of values for the *CL* dimension, there is an ongoing discussion as to whether it is possible to partition the epistemic scale into discrete categories (Rubin, 2007). However, the use of a number of distinct categories is undoubtedly easier for annotation purposes and has been proposed in a number of previous schemes. Although recent work has suggested the use of four or more categories (Shatkay et al., 2008; Thompson et al., 2008), our initial analysis of bio-event corpora has shown that only three levels of certainty seem readily distinguishable for bio-events. This is in line with Hoye (1997), whose analysis of general English showed that there are at least three articulated points on the epistemic scale.

We have chosen to use numerical values for this dimension, in order to reduce potential annotator confusions or biases that may be introduced through the use of labels corresponding to particular lexical markers of each category, such as *probable* or *possible*, and also to account for the fact that slightly different interpretations apply to the different levels, according to whether the event has a *KT* value of *Analysis* or *General*.

L3: No expression of uncertainty or speculation (default category)

L2: High confidence or slight speculation.

L1: Low confidence or considerable speculation; typical lexical markers include *may*, *might* and *perhaps.*

2.3 Source

The source of experimental evidence provides important information for biologists. This is demonstrated by its annotation during the creation of the Gene Ontology (Ashburner et al., 2000) and in the corpus created by Wilbur et al. (2006). The *Source* dimension can also help in distinguishing new experimental knowledge from previously reported knowledge. Our scheme distinguishes two categories, namely:

Other: The event is attributed to a previous study. In this case, explicit clues (citations or phrases like *previous studies* etc.) are normally present.

Current: The event makes an assertion that can be (explicitly or implicitly) attributed to the current study. This is the default category, and is assigned in the absence of explicit lexical or contextual clues.

2.4 Polarity

This dimension identifies negated events. Although certain bio-event corpora are annotated with this information, it is still missing from others. The indication of whether an event is negated is vital, as the interpretation of a negated event instance is completely opposite to the interpretation of a non-negated (positive) instance of the same event.

We define negation as the absence or non-existence of an entity or a process. Negation is typically expressed by the adverbial *not* and the nominal *no*. However, other lexical devices like negative affixals (*un-* and *in-*, etc.), restrictive verbs (*fail*, *lack*, and *unable*, etc.), restrictive nouns (*exception*, etc.), certain adjectives (*independent*, etc.), and certain adverbs (*without*, etc.) can also be used.

2.5 Manner

Events may be accompanied by a word or phrase which provides an indication of the rate, level, strength or intensity of the interaction. We refer to this as the *Manner* of the event. Information regarding manner is absent from the majority of existing bio-event corpora, but yet the presence of such words can be significant in the correct interpretation of the event. Our scheme distinguishes 3 categories of *Manner*, namely:

High: Typically expressed by adverbs and adjectives like *strongly*, *rapidly* and *high*, etc.

Low: Typically expressed by adverbs and adjectives like *weakly*, *slightly* and *slow*, etc.

Neutral: Default category assigned to all events without an explicit indication of manner.

3 Hyper-Dimensions

Determining the pragmatic/rhetorical intent behind an event is not completely possible using any one of our explicitly annotated dimensions. Although the *Knowledge Type* value forms the basis for this, it is not in itself sufficient. However, a defining feature of our annotation scheme is that additional information can be inferred by considering combinations of some of the explicitly annotated dimensions. We refer to this additional information as "latent" or "hyper" dimensions of our scheme. We have identified two such hyper-dimensions.

3.1 New Knowledge

The isolation of events describing new knowledge can be important in certain tasks undertaken by biologists, as explained earlier. Events with the *Knowledge Type* of *Observation* could correspond to new knowledge, but only if they represent observations from the current study, rather than observations cited from elsewhere. In a similar way, an *Analysis* drawn from experimental results in the current study could be treated as new knowledge, but generally only if it represents a straightforward interpretation of results, rather than something more speculative.

Hence, we consider *New Knowledge* to be a hyper-dimension of our scheme. Its value (either *Yes* or *No*) is inferred by considering a combination of the value assignments for the *KT*, *Source* and *CL* dimensions.

Table 1 shows the inference table that can be used to obtain the value for the *New Knowledge* hyper-dimension from the assigned values of the *Source*, *KT* and *CL* dimensions. The symbol 'X' indicates a "don't care condition", meaning that this value does not have any impact on the result.

Source (Annotated)	**KT** (Annotated)	**CL** (Annotated)	**New Knowledge** (Inferred)
Other	X	X	No
X	X	L2	No
X	X	L1	No
Current	Observation	L3	Yes
Current	Analysis	L3	Yes
X	General	X	No
X	Investigation	X	No

Table 1. Inference-Table for New Knowledge Hyper-Dimension

3.2 Hypothesis

A further hyper-dimension of our scheme is *Hypothesis*. The binary value of this hyper-dimension can be inferred by considering the values of *KT* and *CL*. Events with a *KT* value of *Investigation* can always be assumed to be a hypothesis, However, if the KT value is *Analysis*, then only those events with a *CL* value of L1 or L2 (speculative inferences made on the basis of results) should be considered as hypothesis, to be matched with more definite experimental evidence when available. A value of L3 in this instance would normally be classed as new knowledge, as explained in the previous section.

Table 2 shows the inference table that can be used to get the value for the *Hypothesis* hyper-dimension.

KT (Annotated)	CL (Annotated)	Hypothesis (Inferred)
General	X	No
Observation	X	No
Analysis	L3	No
Analysis	L2	Yes
Analysis	L1	Yes
Investigation	X	Yes

Table 2. Inference-Table for Hypothesis Hyper-Dimension

4 Evaluation

The annotation scheme has been evaluated through a small annotation experiment. We randomly choose 70 abstracts from the GENIA Pathway Corpus, which collectively contain over 2600 annotated bio-events. Two of the authors independently annotated these bio-events using a set of annotation guidelines. These guidelines were developed following an analysis of the various bio-event corpora and the output of the initial case study (Nawaz et al., 2010).

The highly favourable results of this experiment further confirmed the feasibility and soundness of the annotation scheme. The remainder of this section discusses the results in more detail.

Dimension	Cohen's Kappa
Knowledge Type	0.9017
Certainty Level	0.9329
Polarity	0.9059
Manner	0.8944
Source	0.9520

Table 3. Inter-Annotator Agreement

4.1 Inter-Annotator Agreement

We have used the familiar measure of Cohen's kappa (Cohen, 1960) for assessing the quality of annotation. Table 3 shows the kappa values for each annotated dimension. The highest value of kappa was achieved for the *Source* dimension, while the *KT* dimension yielded the lowest kappa value. Nevertheless, the kappa scores for all annotation dimensions were in the *good* region (Krippendorff, 1980).

4.2 Category Distribution

Knowledge Type: The most prevalent category found in this dimension was *Observation*, with 45% of all annotated events belonging to this category. Only a small fraction (4%) of these events was represented by an explicit lexical clue (mostly sensory verbs). In most cases the tense, local context (position within the sentence) or global context (position within the document) were found to be important factors.

The second most common category (37% of all annotated events) was *General*. We discovered that most (64%) of the events belonging to this category were processes or states embedded in noun phrases (such as *c-fos expression*). More than a fifth of the *General* events (22%) expressed known scientific facts, whilst a smaller fraction (14%) expressed experimental/scientific methods (such as *stimulation* and *incubation* etc.). Explicit lexical clues were found only for facts, and even then in only 1% of cases.

Analysis was the third most common category, comprising 16% of all annotated events. Of the events belonging to this category, 44% were deductions (*CL=L1*), whilst the remaining 54% were hedged interpretations (*CL=L2/L3*). All *Analysis* events were marked with explicit lexical clues.

The least common category was *Investigation* (1.5% of all annotated events). All *Investigation* events were marked with explicit lexical clues.

Certainty Level: *L3* was found to be the most prevalent category, corresponding to 93% of all events. The categories *L2* and *L1* occurred with frequencies of 4.3% and 2.5%, respectively. The relative scarcity of speculative sentences in scientific literature is a well documented phenomenon (Thompson et al., 2008; Vincze et al., 2008). Vincze et al. (2008) found that less than 18% of sentences occurring in biomedical abstracts are speculative. Similarly, we found that around 20% of corpus events belong to speculative sentences. Since speculative sentences contain non-speculative events as well, the frequency of speculative events is expected to be much less than the frequency of speculative sentences. In accordance with this hypothesis, we found that only 7% of corpus events were expressed with some degree of speculation. We also found that almost all speculated events had explicit lexical clues.

Polarity: Our event-centric view of negation showed just above 3% of the events to be negated. Similarly to speculation, the expected fre-

quency of negated events is lower than the frequency of negated sentences. Another reason for finding fewer negated events is the fact that, in contrast to previous schemes, we draw a distinction between events that are negated and events expressed with *Low* manner. For example, certain words like *limited* and *barely* are often considered as negation clues. However, we consider them as clues for *Low* manner. In all cases, negation was expressed through explicit lexical clues.

Manner: Whilst only a small fraction (4%) of events contains an indication of *Manner*, we found that where present, manner conveys vital information about the event. Our results also revealed that indications of *High* manner are three times more frequent than the indications of *Low* manner. We also noted that both *High* and *Low* manners were always indicated through the use of explicit clues.

Source: Most (99%) of the events were found to be of the *Current* category. This is to be expected, as authors tend to focus on current work in within abstracts. It is envisaged, however, that this dimension will be more useful for analyzing full papers.

Hyper-dimensions: Using the inference tables shown in section 3, we calculated that almost 57% of the events represent *New Knowledge*, and just above 8% represent *Hypotheses*.

5 Conclusion and Future Work

We have evaluated a slightly modified version of our meta-knowledge annotation scheme for bio-events, first presented in Nawaz et al. (2010). The scheme captures key information regarding the correct interpretation of bio-events, which is not currently annotated in existing bio-event corpora, but which we have shown to be critical in a number of text mining tasks undertaken by biologists. The evaluation results have shown high inter-annotator agreement and a sufficient number of annotations along each category in every dimension. These results have served to confirm the feasibility and soundness of the annotation scheme, and provide promising prospects for its application to existing and new bio-event corpora.

We are currently working on a large scale annotation effort, involving multiple independent annotators. Although our main objective is to enrich the entire GENIA event corpus with meta-knowledge information, we also plan to create a small corpus of full papers enriched with bio-event and meta-knowledge annotations.

Acknowledgments

The work described in this paper has been funded by the Biotechnology and Biological Sciences Research Council through grant numbers BBS/B/13640, BB/F006039/1 (ONDEX)

References

M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. *Nature Genetics* 25:25-29.

J. Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20: 37–46.

A. de Waard, B. Shum, A. Carusi, J. Park, M. Samwald and Á. Sándor. 2009. Hypotheses, Evidence and Relationships: The HypER Approach for Representing Scientific Knowledge Claims. In *Proceedings of the Workshop on Semantic Web Applications in Scientific Discourse*. Available at: http://oro.open.ac.uk/18563/

L. Hoye. 1997. *Adverbs and Modality in English*. London & New York: Longman

K. Hyland. 1996. Talking to the Academy: Forms of Hedging in Science Research Articles. *Written Communication* 13(2):251-281.

K. Hyland. 2005. *Metadiscourse: Exploring Interaction in Writing*. London: Continuum

J. Kim, T. Ohta and J. Tsujii. 2008. Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics* 9:10

K. Krippendorff. 1980. *Content Analysis: An Introduction to Its Methodology*. Beverly Hills: Sage Publications

H. Langer, H. Lungen and P. S. Bayerl. 2004. Text type structure and logical document structure. In *Proceedings of the ACL Workshop on Discourse Annotation*, pages 49-56

M. Light, X. T. Qui and P. Srinivasan. 2004. The language of bioscience: Facts, speculations, and statements in between. In *Proceedings of the BioLink 2004 Workshop on Linking Biological Literature, Ontologies and Databases: Tools for Users*, pages 17-24.

F. Lisacek, C. Chichester, A. Kaplan and A. Sandor. 2005. Discovering Paradigm Shift Patterns in Biomedical Abstracts: Application to Neurodegenerative Diseases. In *Proceedings of SMBM 2005*, pages 212-217

L. McKnight and P. Srinivasan. 2003. Categorization of sentence types in medical abstracts. In *Proceedings of the 2003 Annual Symposium of AMIA*, pages 440-444.

B. Medlock and T. Briscoe. 2007. Weakly supervised learning for hedge classification in scientific literature. In *Proceedings of ACL 2007*, pages 992- 999.

Y. Miyao, T. Ohta, K. Masuda, Y. Tsuruoka, K. Yoshida, T. Ninomiya and J. Tsujii. 2006. Semantic Retrieval for the Accurate Identification of Relational Concepts in Massive Textbases. In *Proceedings of COLING-ACL 2006*, pages 1017-1024.

Y. Mizuta and N. Collier. 2004. Zone identification in biology articles as a basis for information extraction. In *Proceedings of the joint NLPBA/BioNLP Workshop on Natural Language for Biomedical Applications*, pages 119-125.

R. Morante and W. Daelemans. 2009. A metalearning approach to processing the scope of negation. In *Proceedings of CoNLL 2009*, pages 21-29.

R. Nawaz, P. Thompson, J. McNaught and S. Ananiadou. 2010. Meta-Knowledge Annotation of Bio-Events. In *Proceedings of LREC 2010*, pages 2498-2507.

K. Oda, J. Kim, T. Ohta, D. Okanohara, T. Matsuzaki, Y. Tateisi and J. Tsujii. 2008. New challenges for text mining: mapping between text and manually curated pathways. *BMC Bioinformatics* 9(Suppl 3): S5.

S. Pyysalo, F. Ginter, J. Heimonen, J. Bjorne, J. Boberg, J. Jarvinen and T. Salakoski. 2007. BioInfer: a corpus for information extraction in the biomedical domain. *BMC Bioinformatics* 8:50.

V. Rizomilioti. 2006. "Exploring Epistemic Modality in Academic Discourse Using Corpora." *Information Technology in Languages for Specific Purposes* 7, pages 53-71

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

P. Ruch, C. Boyer, C. Chichester, I. Tbahriti, A. Geissbühler, P. Fabry, J. Gobeill, V. Pillet, D. Rebholz-Schuhmann and C. Lovis. 2007. Using argumentation to extract key sentences from biomedical abstracts. *International Journal of Medical Informatics* 76(2-3):195-200.

O. Sanchez-Graillet and M. Poesio. 2007. Negation of protein-protein interactions: analysis and extraction. *Bioinformatics* 23(13):i424-i432

Á. Sándor. 2007. Modeling metadiscourse conveying the author's rhetorical strategy in biomedical research abstracts. *Revue Française de Linguistique Appliquée* 200(2):97-109.

H. Shatkay, F. Pan, A. Rzhetsky and W. J. Wilbur. 2008. Multi-dimensional classification of biomedical text: toward automated, practical provision of high-utility text to diverse users. *Bioinformatics* 24(18): 2086-2093.

S. Teufel, J. Carletta and M. Moens. 1999. An annotation scheme for discourse-level argumentation in research articles. In *Proceedings of EACL 1999*, pages 110-117.

S. Teufel, A. Siddharthan and C. Batchelor. 2009. Towards discipline-independent argumentative zoning: Evidence from chemistry and computational linguistics. In *Proceedings of EMNLP-09*, pages 1493-1502

P. Thompson, S. Iqbal, J. McNaught and S. Ananiadou. 2009. Construction of an annotated corpus to support biomedical information extraction. *BMC Bioinformatics* 10: 349.

P. Thompson, G. Venturi, J. McNaught, S. Montemagni and S. Ananiadou. 2008. Categorising Modality in Biomedical Texts. In *Proceedings of the LREC 2008 Workshop on Building and Evaluating Resources for Biomedical Text Mining*, pages 27-34.

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

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