Identification of Manner in 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: raheel.nawaz@cs.man.ac.uk, paul.thompson@manchester.ac.uk, sophia.ananiadou@manchester.ac.uk

Abstract

Due to the rapid growth in the volume of biomedical literature, there is an increasing requirement for high-performance semantic search systems, which allow biologists to perform precise searches for events of interest. Such systems are usually trained on corpora of documents that contain manually annotated events. Until recently, these corpora, and hence the event extraction systems trained on them, focussed almost exclusively on the identification and classification of event arguments, without taking into account how the textual context of the events could affect their interpretation. Previously, we designed an annotation scheme to enrich events with several aspects (or dimensions) of interpretation, which we term *meta-knowledge*, and applied this scheme to the entire GENIA corpus. In this paper, we report on our experiments to automate the assignment of one of these meta-knowledge dimensions, i.e. *Manner*, to recognised events. Manner is concerned with the rate, strength intensity or level of the event. We distinguish three different values of manner, i.e., *High, Low* and *Neutral*. To our knowledge, our work represents the first attempt to classify the manner of events. Using a combination of lexical, syntactic and semantic features, our system achieves an overall accuracy of 99.4%.

Keywords: meta-knowledge, event, manner

1.   Introduction

In recent years, a number of biomedical corpora have been created, whose rich annotation includes not only named entities of interest, but also the bio-events in which these entities participate, e.g., GENIA (Kim et al. 2008), BioInfer (Pyysalo et al. 2007), and GREC (Thompson et al. 2009). Such corpora constitute important resources for training domain-specific information extraction (IE) systems, which in turn can allow sophisticated semantic-based searching to be carried out over documents (Ananiadou et al. 2010). In general, the event annotation involves the identification of individual events and their participants. However, the annotation pays little or no attention to the additional information present in the textual context, which is vital for the correct interpretation of events. We refer to this additional information as *meta-knowledge*. We have previously defined an annotation scheme for capturing the key meta-knowledge aspects of a bio-event (Nawaz et al. 2010). The scheme has subsequently been applied to the GENIA event corpus, which comprises 1000 MEDLINE abstracts containing 36,858 bio-events (Thompson et al. 2011b)., to create the GENIA meta-knowledge (GENIA-MK) corpus.

In this paper, we describe the design and evaluation of a machine learning system that can automate the assignment of one aspect (or *dimension)* of meta-knowledge to bio-events, i.e. *Manner.* This is the most domain-specific dimension of our scheme, which encodes the rate, level, strength or intensity of the event (in biological terms). The identification of such information is considered to be highly important for the correct interpretation of biomedical events (Tsai et al. 2007). To our knowledge, our system is the first that is able to automatically identify and classify information about manner in biomedical text, through the assignment of three possible values to events, i.e., *High, Low* and *Neutral*, with the latter being the default value. Given that non-default manner values are assigned to around 5% of events in the GENIA event, a majority class baseline system would achieve an accuracy of 95%. Through the employment a combination of several different feature types, i.e., syntactic, semantic, lexical, lexico-semantic and lexico-syntactic, our system is able to perform considerably better than the baseline, with an overall accuracy of 99.4% and micro averaged F-scores of 98.3%.

The remainder of this paper is structured as follows: In section 2, we provide background information to the current work, starting with a brief description of bio-events, followed by a more detailed explanation of event manner, and concluding with some statistics regarding its annotation in the GENIA Event corpus. In section 3, we describe our work on designing the new classifier that is able to predict appropriate manner values for events. Firstly, we provide an analysis of the different types of explicit textual cues that can help to predict different manner values. Secondly, we describe the set of features employed by the system, together with the learning algorithm used. In section 4, we present and discuss the results achieved by our classifier, while in section 5, we summarise our work and propose some directions for future work.

2.   Background

This section provides a brief introduction to the bio-events, event manner and its annotation in the enriched GENIA event corpus.

2.1 Bio-Events

In its most general form, a **textual event** can be described as an action, relation, process or state expressed in the text (Sauri & Pustejovsky 2009). More specifically, an event is a structured semantic representation of a certain piece of information contained within the text. Events are usually anchored to particular text fragments that are central to the description of the event, e.g., *event-trigger*, *event-participants* and *event-location*, etc. A **bio-event** is a textual event specialised for the biomedical domain, in that it constitutes a dynamic bio-relation involving one or more participants (Kim et al. 2008). These participants can be bio-entities or (other) bio-events, and each is assigned a semantic role/slot like *theme* and *cause,* etc. Bio-events and bio-entities are also typically assigned semantic types/classes from particular taxonomies/ontologies. Consider sentence (1):

(1) *"The c-jun mRNA was slightly augmented by LTB4"*.

This sentence contains a single bio-event of type *positive_regulation*, whose event-trigger is the verb *augmented*. Figure 1 shows a typical structured representation of this bio-event. The event has two participants: *c-jun mRNA* (a bio-entity of type *RNA_molecule*) and *LTB4* (a bio-entity of type *organic_molecule*). Each participant is assigned a semantic role label to characterise the part it plays in the description of the event.

TRIGGER: *augmented*
TYPE: **positive_regulation**
THEME: *c-jun mRNA* : RNA_molecule
CAUSE: *LTB4* : organic_molecule

Figure 1: Typical bio-event representation

A system trained to extract event representations automatically from texts can allow structured searches to be performed over bio-events, using different types of semantic restrictions, in terms of semantic role types, named entity types, etc. Such advanced search functionality can help biologists to locate relevant information much more quickly than is possible using the traditional method of keyword searches over unstructured documents (Miyao et al. 2006).

It is important to notice that sentence (1) expresses information about the *manner* of the event, although this is not encoded in the typical event representation shown in Figure 1. In sentence (1), the manner of the event is conveyed through the use of the adverb *slightly*, which denotes that the event occurred with a lesser intensity than would be expected by default. Thus, we can say that the *positive_regulation* event occurred with *Low* manner. If the word *slightly* was replaced with *significantly*, then the event would have *High* manner. The detection of manner information can be useful for several tasks, e.g., in comparing results obtained by different authors, or to help to detect possible contradictions or inconsistencies in the results reported in different papers.

2.2 Manner of Bio-Events

The term "manner" could correspond to any information about *how* an event occurs, and so is not in itself domain-specific. Indeed, manner annotated a general adjunct-like argument type in the PropBank corpus, (Palmer et al. 2005), which provides a semantic annotation of general language verbs that appear in the Penn Treebank (Marcus et al. 1994). However, since adjuncts are considered to be general phrases that are not closely associated with any particular verb, they are not normally specified in semantic frame resources that are developed for general language.

In contrast, manner is considered to be highly important for the correct interpretation of biomedical relations and events (Tsai et al. 2007). Accordingly, in the GREC corpus, *Manner* was annotated as one of 13 fixed semantic roles that can characterise the semantic arguments of verbs and nominalisations in biomedical texts. The annotations were extracted as semantic frames and linked with syntactic frames in the BioLexicon (Thompson et al. 2011a), thus allowing the identification of verbs that are particularly likely to specify manner information in biomedical texts.

In the GREC corpus and the BioLexicon, the characterisation of manner arguments can be quite wide-ranging. They can correspond to the intensity of an event, as in sentence (1). However, they can also correspond to a process or method that is employed by the agent to bring about the event (normally a noun phrase after the preposition *by*), an adverb relating to a process that describes how the event is carried out, information about the direction of an event, etc.

In our model of bio-event interpretation (Nawaz et al. 2010), each dimension of event meta-knowledge comprises a fixed set of values, e.g., there are 2 possible values for *Polarity*, and 3 for *Certainty Level*. Thus, while the BioLexicon can help to identify diverse phrases that are related to the manner of an event, the *Manner* dimension in our meta-knowledge scheme aims to provide a useful *classification* of events according to the type of manner that they express. Given the wide range of information that can come under the general heading of manner, our meta-knowledge scheme focusses on a restricted view of the manner of biological processes, which lends itself to a reasonably straightforward division into a set of distinct categories, and which are feasible to recognise automatically.

We took as our starting point the relatively narrow definition of manner proposed in (Sanchez-Graillet & Poesio 2007) for a specific type of bio-event, i.e., protein-protein interactions (PPI). According to them, manner may reveal levels of interaction or certainty of the reported interaction, and is indicated by *manner cues* (adjectives or adverbs) that affect the PPI *trigger* (the

word or phrase indicating the presence of a PPI). Based on our analysis of bio-events, our definition of manner is a slightly modified version of the one provided in (Sanchez-Graillet & Poesio 2007). Firstly, we do not include aspects of certainty, since we treat *Certainty Level* as a separate meta-knowledge dimension. Secondly, we extend the other part of the definition slightly, to cover information concerned with the rate, strength or intensity of the event, as well as the level. This expanded interpretation is needed, given that our meta-knowledge annotation scheme is intended to be applicable to a wider range of events than only PPIs, whose varying semantics mean that expressions of manner can have subtly different interpretations according to the type of event they modify. Based on a manual examination of a large number of events in the GENIA corpus, we found that events can normally be ascribed to one of the following three categories of manner:

- **High:** Event has explicit indication of higher than default rate, level, strength or intensity. Cue expressions are typically adjectives or adverbs such as *high*, *strongly*, *rapidly*, *potent*, etc.
- **Low:** Event expresses lower than default rate, level, strength or intensity. Cue expressions are typically adjectives and adverbs such as *slightly*, *partially*, *small*, etc.
- **Neutral:** The default category, for events with no explicit indication of either *High* or *Low* manner. In rare cases, *Neutral* manner is explicitly indicated, using clue words such as *normal* or *medium*, etc.

When combined with polarity (which is another of our meta-knowledge annotation dimensions), annotation of event manner can help to capture subtle variations between the interpretations of different events. That is to say, a distinction can be made between "low interaction" and "no interaction". Historically, certain cues of *Low* manner (like *low*, *little*, *small*, etc.) have been treated as negation indicators. In the field of sentiment analysis, these cues have been considered a special class of negative polarity indicators, which have been referred to as both *diminishers* (Wiegand et al. 2010) and *negative polarity shifters* (Wilson et al. 2005). The same types of cues have been treated as negation triggers in the field of biomedical text mining (Pyysalo et al. 2007; Kim et al. 2008). However, in the context of bio-events, there is a clear and important distinction between a *Low* manner event and a negated (i.e., non-existent) event. This view has been confirmed by biologists who were consulted and involved in the creation of the GENIA-MK corpus.

2.3 Annotation of Manner in the Enriched GENIA Event Corpus

Analysis of the meta-knowledge annotations in the GENIA-MK corpus revealed that 1,392 events (4%) are expressed with *High* manner, 323 events (1%) are expressed with *Low* manner, and the remaining 35,143 events (95%) were found to be of *Neutral* manner.

Amongst events with an explicit indication of manner, *High* manner marking is much more common, accounting for 81% of cases. However, the significance of identifying instances of *Low* manner cannot be overlooked, since, as described above, it can help to distinguish between truly negative events and those that occur at a low level or with low intensity. Interestingly, the overall frequency of events expressed with a non-default manner is only 1% less than the frequency of negated events (Thompson et al. 2011b). While negation detection has received significant attention in the literature (Morante & Sporleder 2010), manner identification in biomedical text remains an understudied area of research.

3. Automated Identification of Event Manner

Since manner is considered an important part of biomedical event descriptions, it follows that training a system to classify events according to the type of manner they express is an important task. To our knowledge, the automatic classification of manner-related information has not previously been attempted in biomedical text, either at the level of events or for larger units of text.

3.1 Analysis of Manner Cues

The textual context of an event and the syntactic structure of the sentence in which the event is contained can both play important roles in determining the most appropriate manner value to assign to an event. Accordingly, these are both taken into account by the set of features used by our classifier, as explained in the next section. However, the single most important factor is the presence of an explicit cue expression in a sentence. Thus, we carried out a detailed analysis of the manner cues identified in the GENIA-MK corpus. Some of the key findings are as follows:

3.1.1 Cue Frequency

While a total of 273 *High* and 103 *Low* manner cues have been identified, most of these cues (72%) appear just once or twice, and only a handful (9%) appear 10 or more times. Moreover, this small set of the most frequent cues occur in the textual context of the majority (61%) of events that are expressed with a non-default manner. These statistics demonstrate that although a relatively small set of cues accounts for a majority of *High/Low* events, much larger cue sets need to be considered in order to achieve optimum results for automated manner identification.

3.1.2 Cue Variation

While most cues for non-default manner consist of particular words and phrases, others constitute patterns, in which different numerical values may be substituted. An example is the expression *n-fold,* in which *n* represents a number. This expression accounts for 111

(over 8%) of the *High* events. However, a particular challenge lies in the fact that the exact form of expression can vary. Indeed, in the GENIA-MK corpus, 13 different variants of this numerical expression have been annotated as *High* cues. Some examples include *2-fold*, *4-6 fold*, *5- to 7-fold*, etc. Moreover, four non-numeric variants (*two-fold*, *threefold*, *two to threefold* and *two-three fold*) have also been annotated as *High* cues. These non-numeric variants account for a further 14 *High* events. Similarly, several variants of the numeric expression *n%* have also been annotated as both *High* and *Low* manner cues.

3.1.3   Cue Ambiguity

The presence of a *High/Low* cue in a sentence is not sufficient to assign a *High/Low* value to all events in the sentence. While a sentence contains, on average, four bio-events, the majority of manner cues affect only one event in the sentence. Therefore, the syntactic structure of the sentence needs to be considered to determine which, if any, events are being affected by the cue. The semantic context also plays an important role in determining the identity of some cue expressions. For example, depending on the context, numerical expressions (like *n-fold* and *n%*) may indicate a *High* manner, a *Low* manner or neither.

3.1.4   Combined Event-Triggers / Manner Cues

Whilst most manner cues are independent of event type, certain words can act simultaneously as both event-triggers (which denote the type of the event) and manner cues. For example, the word *overexpression* is an event-trigger that introduces an event of type *gene_expression*. Furthermore, the word tells us that the event occurred with *High* manner.

3.1.5   Effect of Negation

An expression of negation inverts the polarity of a manner cue. For example, the word *significant* acts a *High* cue, but its negated form (*no/not significant*) is a *Low* cue.

3.2   Classifier Design

In this section, we explain the various different types of features that are used by our classifier, together with an explanation of the learning algorithm that was employed.

3.2.1.   Features

We used a combination of syntactic, semantic, lexical, lexico-semantic and lexico-syntactic features. The Enju parser (Miyao & Tsujii 2008) was used to obtain the lexical and syntactic information required to construct these features. We also compiled master cue lists for the *High* and *Low* categories by extracting all *High/Low* cues identified in the GENIA-MK corpus. These cue lists were also used in the generation of features. A brief explanation of each feature set is as follows:

- **Syntactic features** include the POS of the event-trigger, event-participants and the *High/Low* cues found in the sentence.
- **Semantic features** are constructed from the semantic information that is annotated for the bio-event. They include the semantic type of the bio-event (e.g., *gene_expression*, *positive_regulation* etc.), the semantic type of each participant (e.g., *lipid*, *DNA molecule*, etc.) and the role of each participant (e.g., *theme* and *cause*, etc.). We have also used a *complexity* feature, which indicates whether a bio-event is simple or complex. The latter value means that the event has one or more participants which are bio-events themselves.
- **Lexical features** include the presence of a *High/Low* cue in the sentence, the cue itself, the presence of a negation indicator and its relative position with respect to the *High/Low* cue, etc. We used regular expressions to identify numeric cues, such as *n-fold* and *n%*.
- **Lexico-semantic features** are constructed using a combination of the "textual" bio-event information and information from the sentence containing the bio-event. The textual bio-event information includes the text fragment indicating the occurrence of the bio-event (i.e., the event-trigger), the text fragments identifying the event participants and the text fragments indicating any event attributes like location etc. The features used include the surface distances between the *High/Low* cue and the event-trigger, participants and event-location, whether the *High/Low* cue is part of the event-trigger, and whether the *High/Low* cue precedes or follows the event-trigger, etc.
- **Dependency (lexico-syntactic) features** are constructed using the textual bio-event information and the dependency relations in the sentence identified by the Enju parser. These features include the presence of direct and indirect dependency relations between the *High/Low* cue present in the sentence and the event-trigger and/or event-location, the types of the dependencies and the lengths of the dependency paths.
- **Constituency (lexico-syntactic) features** are based around the *command* (Langacker 1969) and *scope* relations, which are derived from the constituency parse tree. The concept of a command relation was first introduced by Langacker (1969) as a means for identifying the nodes affected by a given element in the constituency parse tree of a sentence. He defined an S-command relation as follows: 'a node X commands a node Y if neither X nor Y dominates the other and the S (sentence) node most immediately dominating X also dominates Y'. We used several command features including the existence of S-, VP- and NP-command relations between the *High/Low* cue and the event-trigger, and/or event-participants. The scope features consider whether the event-trigger

falls under the syntactic scope of the *High/Low* cue.

3.2.2. Learning Algorithm

We built a classifier using the Random Forest (Breiman 2001) algorithm. This algorithm develops an ensemble/forest of Decision Trees from randomly sampled subspaces of the input features. Once the forest has been created, new objects are classified by first obtaining individual classifications from each tree and then using a majority vote to attain the final classification. The Random Forest algorithm has been successfully used for various text mining and bioinformatics tasks (Chen & Liu 2005; Qi et al. 2005). We used the WEKA (Hall et al. 2009) implementation of the Random Forest algorithm, which is based on (Breiman 2001). Our optimization settings included: (1) setting the number of trees in the forest to 10, (2) setting the number of features used to build individual trees to $\log(N+1)$, where N is the total number of features, (3) setting no restrictions on the depth of individual trees.

4. Results and Discussion

We conducted a series of experiments using different cue lists and feature combinations. All results were 10-fold cross validated. The best results, as shown in Table 1, were achieved using all feature sets (mentioned in section 3.2), the 50 most frequent *High* cues and the 25 most frequent *Low* cues.

Although reasonable results (71% F-score) were achieved for the *Low* category, the results for the *High* category were significantly better. This is partly because the number of training examples available for the *High* category is 4 times higher than those available for the *Low* category. Moreover, the *Low* cues are more diverse and scattered than the *High* cues. The best results were achieved for the *Neutral* category. However, this is to be expected, given that the vast majority of training examples belong to this category. In order to evaluate the overall classifier performance, we calculated the macro and micro averages. The micro averaged results were significantly higher than the macro averaged results. This is because the best classified category (*Neutral*) is also the most abundant by a significant margin.

As mentioned above, since 95% of all events belong to the *Neutral* category, a classifier which assigns the *Neutral* category to all instances will achieve an accuracy of 95%. Therefore, this figure provides a natural baseline for measuring the overall accuracy of the classification system. Our classification system achieved an overall accuracy of 99.4%, which is significantly higher than the baseline.

For the *High* category, the recall is 7% lower than precision. This difference is almost double (13%) for the *Low* category. An error analysis revealed that, for both categories, the main factor contributing towards reduced recall was the inability of the system to identify the *High/Low* cues present in the sentence. As mentioned above, cues are mainly identified via *High/Low* cue lists. Given the ambiguous nature of *High/Low* cues, the size of these lists introduces a precision-recall trade-off, i.e., larger cue lists improve recall at the expense of precision. Thus, the optimum results (as shown in Table 1) were achieved using cut-down versions of the master cue lists. The use of shorter cue lists (i.e., the 50 most frequent *High* cues and the 25 most frequent *Low* cues) enhanced the classification performance (F-score) by 5% for the *High* category and by 7% for the *Low* category. However, it imposed implicit upper-limits of 91% and 79% on the recall for the *High* and *Low* categories, respectively.

Category	Precision	Recall	F-Score
High	85.1%	77.7%	81.2%
Low	78.7%	65.4%	71.4%
Neutral	99.1%	99.4%	99.2%
Macro Avg	87.6%	80.8%	83.9%
Micro Avg	98.4%	98.3%	98.3%

Table 1: Classification Results (10-fold CV)

A significant proportion (32%) of misclassified events belonged to sentences with complex syntactic structures, e.g., where the event-trigger and the *High/Low* cue belonged to different clauses. These misclassifications can be partly attributed to parsing limitations, especially in terms of identifying complex dependency relations.

5. Conclusion

We have analysed the problem of the identification of manner in bio-events and have presented a machine learning based solution to this problem. We have shown that the manner of bio-events can be automatically identified with a high degree of accuracy. Our classification system achieves an overall accuracy of over 99% and macro and micro averaged F-scores of 84% and 98% respectively. Given the level of accuracy achieved by our system, we plan to apply use it to enrichment other bio-event corpora with manner information automatically. We also plan to integrate our manner identification system with the event extraction systems, such as the one presented in (Miwa et al. 2010). The resulting system will be able to extract bio-events with the specified manner type from textual sources.

6. Acknowledgements

7. References

Ananiadou, S., Thompson, P. and Nawaz, R. (2010). Improving Search Through Event-based Biomedical Text Mining. *First International Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts (AMICUS 2010), CLARIN/DARIAH 2010*. Vienna, Austria.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1): 5-32.

Chen, X.-W. and Liu, M. (2005). Prediction of Protein–Protein Interactions Using Random Decision Forest Framework. *Bioinformatics*, 21(24): 4394-4400.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H. (2009). The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11(1).

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

Langacker, R. (1969). On Pronominalization and the Chain of Command. *Modern Studies in English*. D. Reibel and S. Schane. Englewood Cliffs, NJ., Prentice-Hall: 160-186.

Marcus, M.P., Santorini, B. and Marcinkiewicz, M.A. (1994). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2): 313-330.

Miwa, M., Saetre, R., Kim, J.-D. and Tsujii, J. (2010). Event Extraction with Complex Event Classification Using Rich Features. *Journal of Bioinformatics and Computational Biology*, 8(1): 131-146.

Miyao, Y., Ohta, T., Masuda, K., Tsuruoka, Y., Yoshida, K., Ninomiya, T. and Tsujii, J.i. (2006). Semantic Retrieval for the Accurate Identification of Relational Concepts in Massive Textbases. *ACL '06: 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL*, Association for Computational Linguistics: 1017-1024.

Miyao, Y. and Tsujii, J.i. (2008). Feature Forest Models for Probabilistic HPSG Parsing. *Computational Linguistics*, 34(1): 35-80.

Morante, R. and Sporleder, C., Eds. (2010). *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*. Uppsala, Sweden.

Nawaz, R., Thompson, P. and Ananiadou, S. (2010). Evaluating a Meta-Knowledge Annotation Scheme for Bio-Events. *Workshop on Negation and Speculation in Natural Language Processing (NeSp-NLP 2010), ACL 2010*. Uppsala, Sweden: 69-77.

Palmer, M., Gildea, D. and Kingsbury, P. (2005). The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1): 71-106.

Pyysalo, S., Ginter, F., Heimonen, J., Bj̈orne, J., Boberg, J., J̈arvinen, J. and Salakoski, T. (2007). BioInfer: A Corpus for Information Extraction in the Biomedical Domain. *BMC Bioinformatics*, 8(50).

Qi, Y., Klein-Seetharaman, J. and Bar-Joseph, Z. (2005). Random Forest Similarity for Protein-Protein Interaction Prediction from Multiple Sources. *Pacific Symposium on Biocomputing*: 531-542.

Sanchez-Graillet, O. and Poesio, M. (2007). Negation of protein-protein interactions: analysis and extraction. *Bioinformatics*, 23(13): i424-32.

Sauri, R. and Pustejovsky, J. (2009). FactBank: A Corpus Annotated with Event Factuality. *Language Resources and Evaluation*, 43: 227–268.

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

Thompson, P., McNaught, J., Montemagni, S., Calzolari, N., Del Gratta, R., Lee, V., Marchi, S., Monachini, M., Pezik, P., Quochi, V., Rupp, C.J., Sasaki, Y., Venturi, G., Rebholz-Schuhmann, D. and Ananiadou, S. (2011a). The BioLexicon: a large-scale terminological resource for biomedical text mining. *BMC Bioinformatics*, 12: 397.

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

Tsai, R., Chou, W.-C., Su, Y.-S., Lin, Y.-C., Sung, C.-L., Dai, H.-J., Yeh, I., Ku, W., Sung, T.-Y. and Hsu, W.-L. (2007). BIOSMILE: A semantic role labeling system for biomedical verbs using a maximum-entropy model with automatically generated template features. *BMC Bioinformatics*, 8: 325.

Wiegand, M., Balahur, A., Roth, B., Klakow, D. and Montoyo, A. (2010). A Survey on the Role of Negation in Sentiment Analysis. *Workshop on Negation and Speculation in Natural Language Processing (NeSp-NLP 2010), ACL 2010*. Uppsala, Sweden: 60-68.

Wilson, T., Wiebe, J. and Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-level Sentiment Analysis. *Human Language Technologies Conference/Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*. Vancouver, Canada: 347–354.