The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts

György Szarvas[1], Veronika Vincze[1], Richárd Farkas[2] and János Csirik[2]

[1]Department of Informatics
University of Szeged
H-6720, Szeged, Árpád tér 2.

[2]Research Group on Artificial Intelligence
Hungarian Academy of Science
H-6720, Szeged, Aradi vértanúk tere 1.

{szarvas, vinczev, rfarkas, csirik}@inf.u-szeged.hu

Abstract

This article reports on a corpus annotation project that has produced a freely available resource for research on handling negation and uncertainty in biomedical texts (we call this corpus the BioScope corpus). The corpus consists of three parts, namely medical free texts, biological full papers and biological scientific abstracts. The dataset contains annotations at the token level for negative and speculative keywords and at the sentence level for their linguistic scope. The annotation process was carried out by two independent linguist annotators and a chief annotator – also responsible for setting up the annotation guidelines – who resolved cases where the annotators disagreed. We will report our statistics on corpus size, ambiguity levels and the consistency of annotations.

1   Introduction

Detecting uncertain and negative assertions is essential in most Text Mining tasks where in general, the aim is to derive factual knowledge from textual data. This is especially so for many tasks in the biomedical (medical and biological) domain, where these language forms are used extensively in textual documents and are intended to express impressions, hypothesised explanations of experimental results or negative findings. Take, for example, the clinical coding of medical reports, where the coding of a negative or uncertain disease diagnosis may result in an over-coding financial penalty. Another example from the biological do-main is interaction extraction, where the aim is to mine text evidence for biological entities with certain relations between them. Here, while an uncertain relation or the non-existence of a relation might be of some interest for an end-user as well, such information must not be confused with real textual evidence (reliable information). A general conclusion is that for text mining, extracted information that is within the scope of some negative / speculative (hedge or soft negation) keyword should either be discarded or presented separately from factual information.

Even though many successful text processing systems (Friedman et al., 1994, Chapman et al. 2001, Elkin et al. 2005) handle the above-mentioned phenomena, most of them exploit hand-crafted rule-based negation/uncertainty detection modules. To the best of our knowledge, there are no publicly available standard corpora of reasonable size that are usable for evaluating the automatic detection and scope resolution of these language phenomena. The availability of such a resource would undoubtedly facilitate the development of corpus-based statistical systems for negation/hedge detection and resolution.

Our study seeks to fill this gap by presenting the BioScope corpus, which consists of medical and biological texts annotated for negation, speculation and their linguistic scope. This was done to permit a comparison between and to facilitate the development of systems for negation/hedge detection and scope resolution. The corpus described in this paper has been made publicly available for research purposes and it is freely downloadable[1].

---

[1] www.inf.u-szeged.hu/rgai/bioscope

1.1 Related work

Chapman et al. (2001) created a simple regular expression algorithm called NegEx that can detect phrases indicating negation and identify medical terms falling within the negative scope. With this process, a large part of negatives can be identified in discharge summaries.

Mutalik et al. (2001) earlier developed Negfinder in order to recognise negated patterns in medical texts. Their lexer uses regular expressions to identify words indicating negation and then it passes them as special tokens to the parser, which makes use of the single-token look-ahead strategy. Thus, without appealing to the syntactic structure of the sentence, Negfinder can reliably identify negated concepts in medical narrative when they are located near the negation markers.

Huang and Lowe (2007) implemented a hybrid approach to automated negation detection. They combined regular expression matching with grammatical parsing: negations are classified on the basis of syntactic categories and they are located in parse trees. Their hybrid approach is able to identify negated concepts in radiology reports even when they are located at some distance from the negative term.

The Medical Language Extraction and Encoding (MedLEE) system was developed as a general natural language processor in order to encode clinical documents in a structured form (Friedman et al., 1994). Negated concepts and certainty modifiers are also encoded within the system, thus it enables them to make a distinction between negated/uncertain concepts and factual information which is crucial in information retrieval.

Elkin et al. (2005) use a list of negation words and a list of negation scope-ending words in order to identify negated statements and their scope.

Although a fair amount of literature on uncertainty (or hedging) in scientific texts has been produced since the 1990s (e.g. Hyland, 1994), speculative language from a Natural Language Processing perspective has only been studied in the past few years. Previous studies (Light et al., 2004) showed that the detection of hedging can be solved effectively by looking for specific keywords which imply speculative content.

Another possibility is to treat the problem as a classification task and train a statistical model to discriminate speculative and non-speculative assertions. This approach requires the availability of labeled instances to train the models on. Medlock and Briscoe (2007) proposed a weakly supervised setting for hedge classification in scientific texts where the aim is to minimise human supervision needed to obtain an adequate amount of training data. Their system focuses on locating hedge cues in text and thus they do not determine the scopes (in other words in a text they define the scope to be a whole sentence).

1.2 Related resources

Even though the problems of negation (mainly in the medical domain) and hedging (mainly in the scientific domain) have received much interest in the past few years, open access annotated resources for training, testing and comparison are rare and relatively small in size. Our corpus is the first one with an annotation of negative/speculative keywords and their scope. The authors are only aware of the following related corpora:

- The *Hedge classification corpus* (Medlock and Briscoe, 2007), which has been annotated for hedge cues (at the sentence level) and consists of five full biological research papers (1537 sentences). No scope annotation is given in the original corpus. We included this publicly available corpus in ours, enriching the data with annotation for negation cues and linguistic scope for both hedging and negation.

- The *Genia Event corpus* (Kim et al., 2008), which annotates biological events with negation and three levels of uncertainty (1000 abstracts).

- The *BioInfer corpus* (Pyysalo et al., 2007), where biological relations are annotated for negation (1100 sentences in size).

In the two latter corpora biological terms (relations and events) have been annotated for both negation and hedging, but linguistic cues (i.e. which keyword modifies the semantics of the statement) have not been annotated. We annotated keywords and their linguistic scope, which is very useful for machine learning or rule-based negation and hedge detection systems.

2 Annotation guidelines

This section describes the basic principles on the annotation of speculative and negative scopes in biomedical texts. Some basic definitions and technical details are given in Section 2.1, then the general guidelines are discussed in Section 2.2 and the most typical keywords and their scopes are illustrated with examples in Section 2.3. Some special cases and exceptions are listed in Section 2.4, then the annotation process of the corpus is described and discussed in Section 2.5. The complete annotation guidelines document is available from the corpus homepage.

2.1 Basic issues

In a text, just sentences with some instance of speculative or negative language are considered for annotation. The annotation is based on linguistic principles, i.e. parts of sentences which do not contain any biomedical term are also annotated if they assert the non-existence/uncertainty of something.

As for speculative annotation, if a sentence is a statement, that is, it does not include any speculative element that suggests uncertainty, it is disregarded. Questions inherently suggest uncertainty – which is why they are asked –, but they will be neglected and not annotated unless they contain speculative language.

Sentences containing any kind of negation are examined for negative annotation. Negation is understood as the implication of the non-existence of something. However, the presence of a word with negative content does not imply that the sentence should be annotated as negative, since there are sentences that include grammatically negative words but have a speculative meaning or are actually regular assertions (see the examples below).

In the corpus, instances of speculative and negative language – that is, keywords and their scope – are annotated. Speculative elements are marked by angled brackets: *<or>*, *<suggests>* etc., while negative keywords are marked by square brackets: [*no*], [*without*] etc. The scope of both negative and speculative keywords is denoted by parentheses. Also, the speculative or negative cue is always included within its scope:

This result (<suggests> that the valency of Bi in the material is smaller than + 3).

Stable appearance the right kidney ([without] hydronephrosis).

In the following, the general guidelines for speculative and negative annotation are presented.

2.2 General guidelines

During the annotation process, we followed a *min-max* strategy for the marking of keywords and their scope. When marking the keywords, a minimalist strategy was followed: the minimal unit that expressed hedging or negation was marked as a keyword. However, there are some cases when hedge or negation can be expressed via a phrase rather than a single word. Complex keywords are phrases that express uncertainty or negation together, but they cannot do this on their own (the meaning or the semantics of its subcomponents are significantly different from the semantics of the whole phrase). An instance of a complex keyword can be seen in the following sentence:

Mild bladder wall thickening (<raises the question of> cystitis).

On the other hand, a sequence of words cannot be marked as a complex keyword if it is only one of those words that express speculative or negative content (even without the other word). Thus prepositions, determiners, adverbs and so on are not annotated as parts of the complex keyword if the keyword can have a speculative or negative content on its own:

The picture most (<likely> reflects airways disease).

Complex keywords are not to be confused with the sequence of two or more keywords because they can express hedge or negation on their own, that is, without the other keyword as well. In this case, each keyword is annotated separately, as is shown in the following example:

Slightly increased perihilar lung markings (<may> (<indicate> early reactive airways disease)).

2.3 Scope marking

When marking the scopes of negative and speculative keywords, we extended the scope to the biggest syntactic unit possible (in contrast to other corpora like the one described in (Mutalik et al., 2001)). Thus, annotated scopes always have the

maximal length – as opposed to the strategy for annotating keywords, where we marked the minimal unit possible. Our decision was supported by two facts. First, since scopes must contain their keywords, it seemed better to include every element in between the keyword and the target word in order to avoid "empty" scopes, that is, scopes without a keyword. In the next example, *however* is not affected by the hedge cue but it should be included within the scope, otherwise the keyword and its target phrase would be separated:

(Atelectasis in the right mid zone is, however, <possible>).

Second, the status of modifiers is occasionally vague: it is sometimes not clear whether the modifier of the target word belongs to its scope as well. The following sentence can describe two different situations:

There is [no] primary impairment of glucocorticoid metabolism in the asthmatics.

First, the glucocorticoid metabolism is impaired in the asthmatics but not primarily, that is, the scope of *no* extends to *primary*. Second, the scope of *no* extends to *impairment* (and its modifiers and complements as well), thus there is no impairment of the glucocorticoid metabolism at all. Another example is shown here:

Mild viral <or> reactive airways disease is detected.

The syntactic structure of the above sentence is ambiguous. First, the airways disease is surely mild, but it is not known whether it is viral or reactive; or second, the airways disease is either mild and viral or reactive and not mild. Most of the sentences with similar problems cannot be disambiguated on the basis of contextual information, hence the proper treatment of such sentences remains problematic. However, we chose to mark the widest scope available: in other words, we preferred to include every possible element within the scope rather than exclude elements that should probably be included.

The scope of a keyword can be determined on the basis of syntax. The scope of verbs, auxiliaries, adjectives and adverbs usually extends to the right of the keyword. In the case of verbal elements, i.e. verbs and auxiliaries, it ends at the end of the clause (if the verbal element is within a relative clause or a coordinated clause) or the sentence, hence all complements and adjuncts are included, in accordance with the principle of maximal scope size. Take the following examples:

The presence of urothelial thickening and mild dilatation of the left ureter (<suggest> that the patient may have continued vesicoureteral reflux).

These findings that (<may> be from an acute pneumonia) include minimal bronchiectasis as well.

These findings (<might> be chronic) and (<may> represent reactive airways disease).

The scope of attributive adjectives generally extends to the following noun phrase, whereas the scope of predicative adjectives includes the whole sentence. For example, in the following two statements:

This is a 3 month old patient who had (<possible> pyelonephritis) with elevated fever.

(The demonstration of hormone receptor proteins in cells from malignant effusions is <possible>).

Sentential adverbs have a scope over the entire sentence, while the scope of other adverbs usually ends at the end of the clause or sentence. For instance,

(The chimaeric oncoprotein <probably> affects cell survival rather than cell growth).

Right upper lobe volume loss and (<probably> pneumonia).

The scope of conjunctions extends to all members of the coordination. That is, it usually extends to the both left and right:

Symptoms may include (fever, cough <or> itches).

Complex keywords such as *either ... or* have one scope:

Mild perihilar bronchial wall thickening may represent (<either> viral infection <or> reactive airways disease).

Prepositions have a scope over the following (noun) phrase:

Mildly hyperinflated lungs ([without] focal opacity).

When the subject of the sentence contains the negative determiners *no* or *neither*, its scope extends to the entire sentence:

Surprisingly, however, ([neither] of these proteins bound in vitro to EBS1 or EBS2).

The main exception that changes the original scope of the keyword is the passive voice. The subject of the passive sentence was originally the object of the verb, that is, it should be within its scope. This is why the subject must also be marked within the scope of the verb or auxiliary. For instance,

(A small amount of adenopathy <cannot be> completely <excluded>).

Another example of scope change is the case of raising verbs (*seem, appear, be expected, be likely* etc.). These can have two different syntactic patterns, as the following examples suggest:

It seems that the treatment is successful.

The treatment seems to be successful.

In the first case, the scope of *seems* starts right with the verb. If this was the case in the second pattern, *the treatment* would not be included in the scope, but it should be like that shown in the first pattern. Hence in the second sentence, the scope must be extended to the subject as well:

It (<seems> that the treatment is successful).

(The treatment <seems> to be successful).

Sometimes a negative keyword is present in the text apparently without a scope: *negative* obviously expresses negation, but the negated fact – what medical problem the radiograph is negative for – is not part of the sentence. In such cases, the keyword is marked and the scope contains just the keyword:

([Negative]) chest radiograph.

In the case of elliptic sentences, the same strategy is followed: the keyword is marked and its scope includes only the keyword since the verbal phrase, that is, the scope of *not,* is not repeated in the sentence.

This decrease was seen in patients who responded to the therapy as well as in those who did ([not]).

Generally, punctuation marks or conjunctions function as scope boundary markers in the corpus, in contrast to the corpus described in (Mutalik et al., 2001) where certain lexical items are treated as negation-termination tokens. Since in our corpus the scope of negation or speculation is mostly extended to the entire clause in the case of verbal elements, it is clear that markers of a sentence or clause boundary determine the end of their scope.

2.4 Special cases

It seems unequivocal that whenever there is a speculative or negative cue in the sentence, the sentence expresses hedge or negation. However, we have come across several cases where the presence of a speculative/negative keyword does not imply a hedge/negation. That is, some of the cues do not denote speculation or negation in all their occurrences, in other words, they are ambiguous.

For instance, the following sentence is a statement and it is the degree of probability that is precisely determined, but it is not an instance of hedging although it contains the cue *probable*:

The planar amide groups in which is still digging nylon splay around 30 less probable event.

As for negative cues, sentences including a negative keyword are not necessarily to be annotated for negation. They can, however, have a speculative content as well. The following sentence contains *cannot*, which is a negative keyword on its own, but not in this case:

(A small amount of adenopathy <cannot be> completely <excluded>).

Some other sentences containing a negative keyword are not to be annotated either for speculation or for negation. In the following example, the negative keyword is accompanied by an adverb and their meaning is neither speculative nor negative. The sequence of the negative keyword and the adverb can be easily substituted by another adverb or adjective having the same (or a similar) meaning, which is by no means negative – as shown in the example. In this way, the sentence below can be viewed as a positive assertion (not a statement of the non-existence of something).

Thus, signaling in NK3.3 cells is not always (=sometimes) identical with that in primary NK cells.

As can be seen from the above examples, hedging or negation is determined not just by the presence

of an apparent cue: it is rather an issue of the keyword, the context and the syntactic structure of the sentence taken together.

2.5 Annotation process

Our BioScope corpus was annotated by two independent linguists following the guidelines written by our linguist expert before the annotation of the corpus was initiated. These guidelines were developed throughout the annotation process as annotators were often confronted with problematic issues. The annotators were not allowed to communicate with each other as far as the annotation process was concerned, but they could turn to the expert when needed and regular meetings were also held between the annotators and the linguist expert in order to discuss recurring and/or frequent problematic issues. When the two annotations for one subcorpus were finalised, differences between the two were resolved by the linguist expert, yielding the gold standard labeling of the subcorpus.

3 Corpus details

In this section we will discuss in detail the overall characteristics of the corpus we developed, including a brief description of the texts that constitute the BioScope corpus and some general statistics concerning the size of each part, distribution of negation/hedge cues, ambiguity levels and finally we will present statistics on the final results of the annotation work.

3.1 Corpus texts

The corpus consists of texts taken from 4 different sources and 3 different types in order to ensure that it captures the heterogenity of language use in the biomedical domain. We decided to add clinical free-texts (radiology reports), biological full papers and biological paper abstracts (texts from Genia).

Table 1 summarises the chief characteristics of the three subcorpora. The 3rd and 5th rows of the table show the ratio of sentences which contain negated or uncertain statements. The 4rd and 6th rows show the number of negation and hedge cue occurrences in the given corpus.

A major part of the corpus consists of clinical free-texts. We chose to add medical texts to the corpus in order to facilitate research on negation/hedge detection in the clinical domain. The radiology report corpus that was used for the clinical coding challenge (Pestian et al., 2007) organised by the Computational Medicine Center in Cincinatti, Ohio in 2007 was annotated for negations and uncertainty along with the scopes of each phenomenon. This part contains 1954 documents, each having a *clinical history* and an *impression* part, the latter being denser in negated and speculative parts.

Another part of the corpus consists of full scientific articles. 5 articles from FlyBase (the same data were used by Medlock and Briscoe (2007) for evaluating sentence-level hedge classifiers) and 4 articles from the open access BMC Bioinformatics website were downloaded and annotated for negations, uncertainty and their scopes. Full papers are particularly useful for evaluating negation/hedge classifiers as different parts of an article display different properties in the use of speculative or negated phrases. Take, for instance, the *Conclusions* section of scientific papers that tends to contain significantly more uncertain or negative findings than the description of *Experimental settings and methods*.

Scientific abstracts are the main targets for various Text Mining applications like protein-protein interaction mining due to their public accessibility (e.g. through PubMed). We therefore decided to include quite a lot of texts from the abstracts of scientific papers. This is why we included the abstracts of the Genia corpus (Collier et al., 1999). This decision was straightforward for two reasons. First, the Genia corpus contains syntax tree annotation, which allows a comparison between scope annotation and syntactic structure. Being syntactic in nature, scopes should align with the bracket structure of syntax trees, while scope resolution algorithms that exploit treebank data can be used as a theoretical upper bound for the evaluation of parsers for resolving negative/hedge scopes. The other reason was that scope annotation can mutually benefit from the rich annotations of the Genia corpus, such as term annotation (evaluation) and event annotation (comparison with the biologist uncertainty labeling of events).

The corpus consists of more than 20.000 annotated sentences altogether. We consider this size to be sufficiently large to serve as a standard evaluation corpus for negation/hedge detection in the biomedical domain.

	Clinical	Full Paper	Abstract
#Documents	1954	9	1273
#Sentences	6383	2624	11872
Negation sentences	6.6%	13.76%	13.45%
#Negation cues	871	404	1757
Hedge sentences	13.4%	22.29%	17.69%
#Hedge cues	1137	783	2691

Table 1: Statistics of the three subcorpora.

3.2 Agreement analysis

We measured the consistency level of the annotation using inter-annotator agreement analysis. The inter-annotator agreement rate is defined as the $F_{\beta=1}$ measure of one annotation, treating the second one as the gold standard. We calculated agreement rates for all three subcorpora between the two independent annotators and between the two annotators and the gold standard labeling. The gold standard labeling was prepared by the creator of the annotation guide, who resolved all cases where the two annotators disagreed on a keyword or its scope annotation.

We measured the agreement rate of annotating negative and hedge keywords, and the agreement rate of annotating the linguistic scope for each phenomenon. We distinguished left-scope, right-scope and full scope agreement that required both left and right scope boundaries to match exactly to be considered as coinciding annotations. A detailed analysis of the consistency levels for the three subcorpora and the ambiguity levels for each negative and hedge keyword (that is, the ratio of a keyword being annotated as a negative/speculative cue and the number of all the occurrences of the same keyword in the corpus) can be found at the corpus homepage.

3.3 BioScope corpus availability

The corpus is available free of charge for research purposes and can be obtained for a modest price for business use. For more details, see the BioScope homepage:
www.inf.u-szeged.hu/rgai/bioscope.

4 Conclusions

In this paper we reported on the construction of a corpus annotated for negations, speculations and their linguistic scopes. The corpus is accessible for academic purposes and is free of charge. Apart from the intended goal of serving as a common resource for the training, testing and comparison of biomedical Natural Language Processing systems, the corpus is also a good resource for the linguistic analysis of scientific and clinical texts.

The most obvious conclusions here are that the usual language of clinical documents makes it much easier to detect negation and uncertainty cues than in scientific texts because of the very high ratio of the actual cue words (i.e. low ambiguity level), which explains the high accuracy scores reported in the literature. In scientific texts – which are nowadays becoming a popular target for Text Mining (for literature-based knowledge discovery) – the detection and scope resolution of negation and uncertainty is, on the other hand, a problem of great complexity, with the percentage of non-hedge occurrences being as high as 90% for some hedge cue candidates in biological paper abstracts. Take for example the keyword *or* which is labeled as a speculative keyword in only 11.32% of the cases in scientific abstracts, while it was labeled as speculative in 97.86% of the cases in clinical texts. Identifying the scope is also more difficult in scientific texts where the average sentence length is much longer than in clinical data, and the style of the texts is also more literary in the former case.

In our study we found that hedge detection is a more difficult problem than identifying negations because the number of possible cue words is higher and the ratio of real cues is significantly lower in the case of speculation (higher keyword/non-keyword ambiguity). The annotator-agreement table also confirms this opinion: the detection of hedging is more complicated than negation even for humans.

Our corpus statistics also prove the importance of negation and hedge detection. The ratio of negated and hedge sentences in the corpus varies in the subcorpora, but we can say that over 20% of the sentences contains a modifier that radically influences the semantic content of the sentence.

One of the chief construction principles of the BioScope corpus was to facilitate the training/development of automatic negation and hedge detection systems. Such systems have to solve two sub-problems: they have to identify real cue words (note that the probability of any word being a keyword can be different for various domains) and

then they have to determine the linguistic scope of actual keywords.

These automatic hedge and negation detection methods can be utilised in a variety of ways in a (biomedical) Text Mining system. They can be used as a preprocessing tool, i.e. each word in a detected scope can be removed from the documents if we seek to extract true assertions. This can significantly reduce the level of noise for processing in such cases where only a document-level labeling is provided (like that for the ICD-9 coding dataset) and just clear textual evidence for certain things should be extracted. On the other hand, similar systems can classify previously extracted statements according to their certainty or uncertainty, which is generally an important issue in the automatic processing of scientific texts.

Acknowledgments

References

Wendy W. Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper and Bruce G. Buchanan. 2001. A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries. *Journal of Biomedical Informatics*, 34(5):301–310.

N. Collier, H. S. Park, N. Ogata, Y. Tateishi, C. Nobata, T. Ohta, T. Sekimizu, H. Imai, K. Ibushi, and J. Tsujii. 1999. The GENIA project: corpus-based knowledge acquisition and information extraction from genome research papers. *Proceedings of EACL-99*.

Peter L. Elkin, Steven H. Brown, Brent A. Bauer, Casey S. Husser, William Carruth, Larry R. Bergstrom and Dietlind L. Wahner-Roedler. 2005. A controlled trial of automated classification of negation from clinical notes. *BMC Medical Informatics and Decision Making* 5:13 doi:10.1186/1472-6947-5-13.

C. Friedman, P.O. Alderson, J.H. Austin, J.J. Cimino, and S.B. Johnson. 1994. A general natural-language text processor for clinical radiology. *Journal of the American Medical Informatics Association*, 1(2):161–174.

Yang Huang and Henry J. Lowe. 2007. A Novel Hybrid Approach to Automated Negation Detection in Clinical Radiology Reports. *Journal of the American Medical Informatics Association*, 14(3):304–311.

Ken Hyland. 1994. *Hedging in academic writing and EAP textbooks. English for Specific Purposes*, 13(3):239–256.

Jin-Dong Kim, Tomoko Ohta, and Jun'ichi Tsujii. 2008. Corpus annotation for mining biomedical events from literature. *BMC Bioinformatics* 2008, 9:10.

Marc Light, Xin Ting Qui, and Padmini Srinivasan. 2004. The language of bioscience: Facts, speculations, and statements in between. In *Proceedings of BioLink 2004 Workshop on Linking Biological Literature, Ontologies and Databases: Tools for Users*. Boston, Massachusetts, Association for Computational Linguistics, 17–24.

Ben Medlock and Ted Briscoe. 2007. Weakly supervised learning for hedge classification in scientific literature. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*. Prague, Association for Computational Linguistics, 992–999.

Pradeep G. Mutalik, Aniruddha Deshpande, and Prakash M. Nadkarni. 2001. Use of General-purpose Negation Detection to Augment Concept Indexing of Medical Documents: A Quantitative Study Using the UMLS. *Journal of the American Medical Informatics Association*, 8(6):598–609.

John P. Pestian, Chris Brew, Pawel Matykiewicz, DJ Hovermale, Neil Johnson, K. Bretonnel Cohen, and Wlodzislaw Duch. 2007. A shared task involving multi-label classification of clinical free text. In *Biological, translational, and clinical language processing*. Prague, Association for Computational Linguistics, 97–104.

Sampo Pyysalo, Filip Ginter, Juho Heimonen, Jari Björne, Jorma Boberg, Jouni Järvinen, and Tapio Salakoski. 2007. BioInfer: a corpus for information extraction in the biomedical domain. *BMC Bioinformatics* 2007, 8:50 doi:10.1186/1471-2105-8-50.