Automatic annotation of speculation in biomedical texts: new perspectives and large-scale evaluation

Julien Desclés **Olfa Makkaoui** **Taouise Hacène**

julien.descles@gmail.com olfa_makkaoui@yahoo.fr taouise@gmail.com

LaLIC Université Paris-Sorbonne Maison de la Recherche 28 rue Serpente, 75006 Paris

Abstract

One emergent field in text mining tools applied to biological texts is the automatic detection of speculative sentences. In this paper, we test on a large scale BioExcom, a rule-based system which annotates and categorizes automatically speculative sentences ("prior" and "new"). This work enables us to highlight a more restrictive way to consider speculations, viewed as a source of knowledge, and to discuss the criteria used to determine if a sentence is speculative or not. By doing so, we demonstrate the efficiency of BioExcom to extract these types of speculations and we argue the importance of this tool for biologists, who are also interested in finding hypotheses.

1 Introduction

In one of the first studies about biological speculations in Natural Language Processing, Light et al. (2004) have reported that biologists can have a dual posture concerning this subject:

"*In the context of information retrieval, an example information need might be "I am looking for speculations about the X gene in liver tissue." One of the authors spoke at a research department of a drug company and the biologists present expressed this sort of information need. On the other hand, one of the authors has also encountered the opposite need: "I am looking for definite statements about transcription factors that interact with NF Kappa B." Both these information needs would be easier to fulfill if automated annotation of speculative passages was possible.*" (Light et al., 2004)

In this quotation, the term "*speculation*" has not exactly the same meaning in these two statements (depending if it comes from compu-tational linguists or from biologists). Indeed, because biologists are almost interested in knowing only factual statements, the information extraction tools have to remove all uncertain statements, identified as hedging or speculation, or at least to present them separately from definite results. Consequently, the vast majority of Natural Language Processing tools dealing with the identification of speculation have followed this very large meaning of the word "*speculation*", in order to avoid extracting uncertain information as factual information (Kilicoglu and Bergler, 2008; Medlock, 2008; Szarvas, 2008; Morante and Daelemans, 2009; Özgur and Radev, 2009). To help improve the information extraction tools, a corpus, called BioScope, has been annotated for speculation, negation and its linguistic scopes in biomedical texts (Szarvas et al., 2008).

However, when a biologist says he is interested in knowing all speculations about a biological entity (gene or protein for example) or a biological process, this claim concerns another meaning of the word "*speculation*". The latter is in this case more restrictive than previously, and close to the notion of hypothesis and uncertain proposal. This interest of biologists can be explained by different reasons. Firstly, since speculations give meaning to results, they sometimes carry more useful information than factual sentences. In addition, speculative sentences emphasize important data, which can be very useful in data-collection papers (genomic and post-genomic papers, see (Brent and Lok, 2005)). Finally, speculations can also give current trends or directions, by enabling the researchers to anticipate future experimental discoveries, or by suggesting other ways to envision biological problems and giving new ideas for future experiments (Blagosklonny and Pardee, 2002). Hence, despite its importance for biologists, the need to find speculation according to this view has been neglected until now in Natural Language

Processing. To our knowledge, the only work focusing specifically on this issue is the development of the rule-based system BioExcom (Desclés et al., 2009). Since BioExcom has obtained good results for detecting speculations but in a relatively small scale evaluation, it seems useful to test this tool on a large, unknown corpus like BioScope. Furthermore, it is important to compare in greater detail these two different approaches to characterize speculative sentences and see more precisely in what they differ.

We performed an automatic annotation of the BioScope corpus by BioExcom and we measured raw performance. We observed that the vast majority of the divergences between BioExcom results and BioScope were due to the criteria used for detecting a speculative sentence. We manually treated the diverging sentences in order to correctly evaluate BioExcom according to its own criteria.

The contributions of this paper are the following:

• We present an original approach for considering speculative sentences in bio text mining. It concerns the definition of a speculation, the criteria used to find it and its importance for biologists.
• We demonstrate the efficiency of BioExcom to recognize these statements on a large-scale evaluation with good results.
• According to this new approach we provide an annotated corpus freely available in order to be used by researchers.

2   Related work

Hyland (1995) has extensively studied hedging, from a linguistic perspective, (the term of hedging has been introduced by Lakoff (1972)) in biological papers. Hedging represents an absence of certainty and is employed to indicate either a lack of commitment to the truth value of an accompanying proposition; either a desire not to express that commitment categorically. Three main functions are outlined for hedging: weakening the strength of a statement, signalling uncertainty and expressing deference to the reader.

From a Natural Language Processing perspective, the first work was carried out by Light et al. (2004). After a linguistic study of the use of speculative language in MEDLINE abstracts, the authors tested the possibility of manually annotating the speculative sentences by experts and linguists whilst providing small annotation guidelines. They concluded that humans can reliably annotate speculative sentences but that it is not possible to distinguish between "high" speculations and "low" speculations. Furthermore they performed an experiment with different automated methods based principally on the retrieval of keywords. Wilbur et al. (2006) defined five qualitative dimensions for scientific text annotations. Two of them concerned speculative statements (certainty and evidence) and they defined various guidelines for annotating text using them.

Medlock and Briscone (2007) provided much more detailed guidelines for hedge detection. A linguist and a domain expert without any input into the guideline development process, labelled a publicly available dataset (FlyBase dataset, consisting of 6 papers on *Drosophila melanogaster*) in order to perform a probabilistic acquisition model and to test it. A separation of the acquisition and classification phases in semi-supervised machine learning was used.

Svarzas (2008) managed to devise a hedge classification in biomedical texts based on a weakly supervised selection of keywords. To evaluate their system, they manually annotated four papers of BMC Bioinformatics with the same criteria as Medlock and Briscone (2007). They obtained an F-Measure of 85% on the FlyBase dataset and an F-Measure of 75% on the BMC Bioinformatics dataset, when the training was carried out with the FlyBase dataset, demonstrating that the probability of their hedge classifiers is limited.

To recognize speculative sentences, Kilicoglu and Bergler (2008) also used speculative keywords from prior linguistic work and expanded them by WordNet, UMLS Specialist Lexicon, and by introducing syntactic patterns. In order to determine the speculative strength of sentences, they assigned speculative cues to weights in two methods. They worked on the two publicly available corpora, obtaining an F-Measure of 85% on the FlyBase data set and an F-Measure of 82% on the BMC Bioinformatics data.

The BioScope corpus is a manually annotated corpus for speculation and negation keywords (token level), and their linguistic scopes (sentence level) (Szarvas et al., 2008). This corpus is publicly available[1] and the annotation

---

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

process has been performed by two independent annotators and a chief linguist. In particular, the corpus consists of 9 full texts papers (five papers from the FlyBase dataset and four papers from the journal BMC Bioinformatics), and 1273 abstracts (from the Genia Corpus (Collier et al., 1999)). The annotation guidelines are provided and speculation is identified to uncertainty and hedging. However the criteria used here are not detailed very accurately, unlike the detailed work of Medlock and Briscone (2007).

In order to detect hedging but also their scope, two recent works were recently published. Morante and Daelemans (2009) present a machine learning system based on a previous system used to detect the scope of negation cues. Thus they show that the same scope finding system can be applied to both negation and hedging. They used the three subcropora of the BioScope corpus to test the efficiency of their system: the best F-Measures they obtained were 85% on the abstracts (10-fold cross-validation experiments), 72% on the Full Text Papers and 60% on the Clinical reports for hedge detection (training phase on the full abstract subcorpus). Özgur and Radev (2009) built also a system that detects speculations and their scopes in biomedical scientific texts. They performed a supervised classification task, where they classified the potential keywords as real speculation keywords or not by using a diverse set of linguistic features that represent the contexts of the keywords. They obtained F-Measure of 92% for the scientific abstracts (10-fold cross-validation experiments). The F-Measure for Full Text Papers (leave-one-out-cross-validation) was 83%.

3   BioExcom

EXCOM[2] (Djioua et al., 2006; Alrahabi, 2010) is a rule-based system using the computational principles of the Contextual Exploration processing (Desclés, 2006). EXCOM does not need any morpho-syntactic analysis and only requires one pre-treatment step of corpus segmentation into segments (which are sentences or sometimes clauses according to the punctuation) (Figure 1). EXCOM uses declarative rules built by linguists or domain experts and based on the search for linguistic markers in the text. The latter are hierarchically organized

into the rules: they are either indicators (strong markers) or clues (complementary markers). Only the presence of indicators in the text triggers the associated CE rules, and then the additional clues can be searched for in the context defined by the rules, which is in our case the same sentence as the indicator (Figure 2).
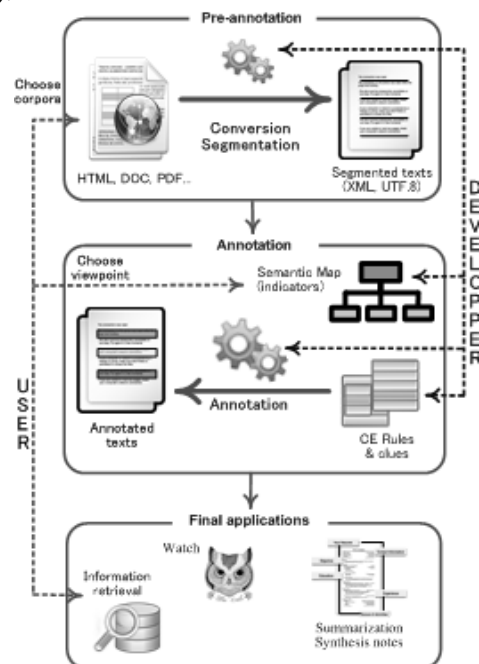

Figure 1: Overview of Excom processing (Alrahabi, 2010)

EXCOM allows the use of already annotated segments as markers, and also to order the rules and to use "negative" clues that cancel certain rules. The clues can be arranged between each other and versus the indicator (at its left, right or inside it). They are used to confirm, invalidate or specify an annotation carried by an indicator. If all the examined conditions of a rule are satisfied, EXCOM attributes the corresponding annotation to the designated segment. Generally the indicators are interconnected into graphs called "semantic maps" (Alrahabi and Desclés, 2008). EXCOM has been used for various tasks such as automatic summarization, relationships between concepts, categorization of bibliographic citations and reported speech.

BioExcom is the system that uses the annotation performed by EXCOM thanks to specific linguistic resources built for detecting speculations in biomedical papers (Desclés et al., 2009). Furthermore, BioExcom performs an indexation of the annotated segments in order to provide the possibility of searching for

---

[2] http://www.excom.fr/

34

specific terms. The rules used for the annotation processing are based on a precise semantic analysis of the multiple ways to express speculation performed by an expert in about seventy biological papers. BioExcom also categorizes speculative segments into "new speculation" (speculative sentences presented for the first time in the paper or not explicitly presented as prior speculation) and "prior speculation" (speculative sentences cited in the paper, but presented as having been proposed previously). BioExcom uses thirty rules, based on twelve indicator classes.
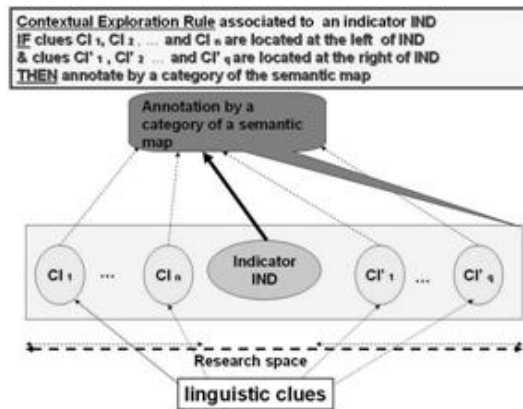

Figure 2: The contextual exploration principles: search for an indicator and then for some clues in a contextual space (a sentence or a clause in our case) according to some associated rules.

The criteria used to find speculations are described in detailed annotation guidelines, which are available on line[3]. Only the segments containing at least one linguistic element expressing speculation are considered. A speculation is defined as a proposal still not demonstrated about a biological issue and explicitly presented as uncertain in the paper. Thus, speculations are considered as potential sources of relevant information for biologists. Others types of statements such as deductions, conclusions, arguments and discussions are not considered as speculative but as intermediary statements, because they either present concepts more or less certain, or they do not make a proposal.

It is worth noting that contrary to other approaches, like Medlock (2008), the open questions are not speculative because they only ask a question about a biological problem without proposing a mechanism (sentence 1).

(1) *How endocytosis of DI leads to the activation of N remains to be elucidated.*

In order to better illustrate what a proposal is and how BioExcom performs annotation we can compare this sentence with the example (2) where BioExcom finds the indicator "*is not known*". Nevertheless, as the goal of this tool is not to extract all sentences expressing uncertainty, BioExcom searches for the clue "*whether*" at the right or at the left of the indicator. The presence of this clue indicates that, contrary to the sentence (1) with "*how*", the sentence (2) proposes a mechanism and consequently is a speculation.

(2) *Also, <u>whether</u> the signaling activity of Ser is similarly regulated by endocytosis <u>is not known</u>.*

The sentences which discuss a speculation without being informative about the content of the speculation are also not considered as speculations (sentence 3, extracted from the guidelines of Medlock (2008)). Indeed the goal of BioExcom is to detect a sentence explaining, at least partially, a speculation, and not to know, as the example (3), whether a speculation is purely speculative or is supported (or not) by facts. For the same reason, BioExcom extracts speculation without taking into account whether this speculation is denied or not.

(3) *The rescue of the D-mib mutant phenotype by ectopic expression of Neur strongly supports this interpretation*

At present, the evaluation process of BioExcom has only been performed on a relatively small scale (2 full-text papers read by five experts and containing 59 speculative sentences in total) and after the automatic annotation process (Desclés et al., 2009). Promising results have been reported for BioExcom in detecting speculations, by providing a high Precision of 98,3%, and a Recall of 95,1% (F-Measure of 96,7%). This rate is consistent with the goal of BioExcom to be a monitoring and decision support tool.

---

[3] http://www.bioexcom.net/

4   Raw evaluation

The Bioscope corpus (Szarvas et al., 2008) consists of three parts; namely medical free texts, biological full papers and biological scientific abstracts. However in this test, we only used two parts of the Bioscope corpus (full papers and abstracts) because we were preferentially interested in the biomedical scientific domain.

First, we cleaned the corpus from all element tags (angled brackets marking hedge keywords and parentheses marking scopes) and saved these documents into text files. The latter could then be automatically segmented into sentences or clauses and semantically annotated by BioExcom. As a result of this processing, BioExcom automatically extracted 341 segments from the Full Text Papers Corpus and 1489 segments from the Abstract Corpus (1830 segments in total)[4]. We could then compare our output files with the Bioscope Corpus which contained manual annotations. In this task we do not consider the categorization of BioExcom ("new speculation" and "prior speculation") and these annotated sentences are only considered as speculative. Thus, we obtained the results presented in Table 1. Consistent with the previous evaluation performed on BioExcom (Desclés et al., 2009), the Precision is high (approximately 93% in average, calculated from the total of segments of the two corpora). Nevertheless, the Recall dramatically falls to approximately 68% (in average) compared to the first evaluation (Recall of 93%).

	Precision	Recall	F-Measure
Full Text Papers	89,35	62,92	73,84
Abstracts	94,75	68,83	79,74

Table 1: Summary of raw results for BioExcom evaluation

Presented briefly are comments and some annotations performed by BioExcom which are in agreement with BioScope.

---

(4) *High conservation of residues near the inframe stop codon also <u>suggests</u> the importance of this region.*

(5) *Therefore, sets of genes identified from co-expression data <u>may</u> often contain multiple extraneous upstream sequences.*

(6) *To test the <u>hypothesis</u> that cortivazol acts in dex-resistant cells by making use of the residual GR found there, wild-type and dex-resistant clones were treated with various concentrations of cortivazol and the induction of GR mRNA was studied.*

(7) *<u>To determine</u> <u>whether</u> specific members of the NF-kappa B family contribute to this effect, we examined the abilities of different NF-kappa B subunits to act with Tat-I to stimulate transcription of HIV in Jurkat T-leukemia cells.*

If we compare the speculative keywords indicated in BioScope (underlined) and the markers used by BioExcom, we see some convergences ("*suggests*" and "*may*" for the sentences (4) and (5)), but also some divergences. In the sentence (6), BioExcom uses "*hypothesis that*" as an indicator in order to extract informative speculations and not only sentences containing the word "*hypothesis*". The example of the sentence (7) is much more illustrative for the differences: whereas only "*whether*" is taken into account as a keyword in BioScope, BioExcom uses "*to determine*" as an indicator and "*whether*" as a positive clue, allowing extracting only sentences containing a proposal (see example 2). The minimalist strategy (searching for the minimal unit that expresses speculation) followed by the annotators of BioScope for the keywords can explain these observations.

5   Corrected evaluation

Our goal was to evaluate the performance of BioExcom according to its own definition of speculation. To analyze the observed low Recall (Table 1), we assumed that all sentences presenting an agreement between both methods (manual annotation in BioScope and automatic annotation by BioExcom) were correctly annotated and we checked manually all the segments (984 segments) presenting a divergence of annotation. This checking was performed by a biologist as chief evaluator (writer

---

[4] The results of the annotation of BioScope corpus by BioExcom are publicly available: http://www.bioexcom.net/. It is worth mentioning that a few sentences were not segmented exactly in the same way by BioExcom and in the BioScope corpus (approximately 2% of divergences). We based all our calculations on BioExcom segmentation.

of the annotation guidelines) and two independent linguists, not allowed to communicate with each other. None of these evaluators knew the annotation performed by BioExcom or in BioScope (blind study). The conflicts were resolved by discussions during regular meetings and, in case of important uncertainty for at least two annotators, the sentences (54 in total) were not taken into account. The BioScope corpus re-annotated according to the criteria of BioExcom is publicly available[5].

We can divide the segments presenting a disagreement between BioExcom and BioScope into two groups depending on the annotation performed either automatically by BioExcom or manually in the BioScope corpus (Figure 3).

The first group consists of 865 segments which were considered as speculations in BioScope (representing around 36% of the total of annotated sentences in BioScope) and which were not annotated by BioExcom (see Figure 2). After manual checking, we found that only around one third of these segments in the corpus Full papers were speculative according to the criteria of BioExcom. This proportion was around one fourth in the BioScope corpus Abstract. The goal of BioExcom to avoid annotating open questions (see examples (1-2)) or lack of knowledge (sentence (8)) can explain some of the absences of annotation by BioExcom.

(8) *The exact role of the ubiquitination pathway in regulating apoptosis is still <u>unclear</u>.*

In other cases, some differences linked to the semantic conception of speculation play a role. Thus, the following sentences are considered as speculative in BioScope:

(9) *Furthermore, genetic and transplantation studies <u>indicate that</u> both Neur and Mib act in a non-autonomous manner [18,21,22,23,25,29], <u>indicating that</u> endocytosis of Dl is associated with increased Dl signalling activity.*

(10) *It can be deduced that the erythroid ALAS precursor protein has a molecular weight of 64.6 kd, and is similar in size to the previously isolated human housekeeping ALAS precursor of molecular weight 70.6 kd.*

Although these sentences correspond to hedging as it has been defined by Hyland (1995), we argue that the sentences (9-10) can be characterized more as a demonstration or a deduction with the expressions "*indicate that*" and "*it can be deduced*" than a speculation. According to the criteria used for establishing BioExcom rules, these sentences do not correspond to a speculation because they present things more or less as certain (see also (Thompson et al., 2008)). In this view, the case of "*indicate that*" is interesting to be detailed. Whereas many studies use it as a linguistic marker of speculation, Kilicoglu and Bergler (2008) recently moderated its speculative meaning by highlighting the additional need to take into account its context. The linguistic analysis used to establish the rules in BioExcom is much more restrictive and does not consider it as a marker.

It should also be mentioned that we noticed a few sentences which were incorrectly annotated in BioScope as speculation. Thus, the sentence (11) is rather, in our opinion, a definite statement ("*or*" can be replaced by "*and*").

(11) *Tandem copies of this 67-bp MnlI-AluI fragment, when fused to the chloramphenicol acetyltransferase gene driven by the conalbumin promoter, stimulated transcription in B cells but not in Jurkat T cells or HeLa cells.*

There are also some sentences annotated in Bioscope which are undoubtedly speculations but they were not identified by BioExcom. Thus, BioExcom has not annotated some speculative sentences because of the lack of some accurate markers into its linguistic resources. This is the case in the sentence (12): "*seems*" is recognized by BioExcom only when it is associated with some adjectives ("*seems probable*" for example). We can also cite the more ambiguous case of "*predicted*" (sentence (13)) which is also absent from the linguistic resources of BioExcom. Although a prediction can be considered as having a more particular status because it presents a low degree of uncertainty, we can consider that it should be taken into account because the predicted proposal remains hypothetical.
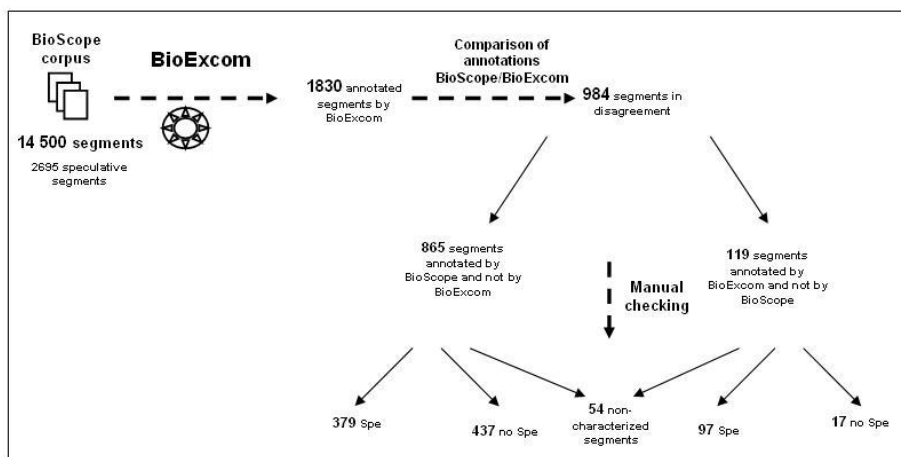
Figure 3: Results of the annotation of BioScope by BioExcom and of the manual checking (the calculations are based on BioExcom segmentation). Spe = speculative segment; no Spe = non speculative segment.

(12) *A transcription factor, NF-AT, which is essential for early T-cell gene activation, <u>seems</u> to be a specific target of cyclosporin A and FK506 action because transcription directed by this protein is blocked in T cells treated with these drugs, with little or no effect on other transcription factors such as AP-1 and NF-kappa B.*

(13) *Expression of full-length and mutant cDNA clones in bacteria reveal that the single HMG motif, which is <u>predicted</u> to contain two extended alpha-helical segments, is sufficient to direct the sequence-specific binding of TCF-1 alpha to DNA.*

Beside the lack of accurate markers, the absence of specific markers in some sentences does not allow the detection of some speculations by BioExcom. For example, in BioExcom, the ambiguity of the indicator "*could*" (past form or conditional form) is lifted by the presence of some positive clues expressing conditionality or possibility, such as "*if*" or "*whether*". But in the sentence (14), "*could*" has no clues to be disambiguated and although it is a speculation, BioExcom did not annotate it.

(14) *This method could be quite useful to detect not only CAG repeats in SBMA but also other polymorphic dinucleotide and trinucleotide repeats.*

The second group consists of segments annotated by BioExcom and not in BioScope

(119 segments, see Figure 2). Around 80% of these sentences appeared to be speculative after manual examination. As an illustration, the following sentence is clearly a speculation ("*We hypothesize that*") but is not annotated in BioScope.

(15) *We hypothesize that a mutation of the hGR glucocorticoid-binding domain is the cause of cortisol resistance.*

Finally, based on these results, we decided to recalculate the Precision, Recall and F-Measure to be more in agreement with the notion of speculation as it has been described by Desclés et al. (2009) (last lines in Tables 2 and 3). Corrected Precision, Recall and F-Measure are respectively around 99%, 83% and 90% (averages calculated from the total of segments of the two corpora). These results are close to the first evaluation performed by Desclés et al. (2009), even if Recall is still lower than previously.

Obviously, our results are not directly comparable with the prior studies because BioExcom does not use exactly the same criteria to recognize the speculative sentences and consequently we re-annotated the BioScope corpus according to the criteria of BioExcom. One other difference is that the source used for linguistic resources and rules of BioExcom is completely different from the source of the corpus used for the evaluation, aside from the studies using directly BioScope like Morante and Daelemans (2009) or Özgur and Radev (2009). Nevertheless, considering that there are a few studies using a part of the BioScope cor-

pus, it can be interesting to mention that Bio-Excom achieves good performances in particular for Precision rate (Table 2 and 3).

		Precision	Recall	F-Measure
1	Fruit-fly dataset			85,08
1	BMC dataset			74,93
2	Fruit-fly dataset	85	86	85
2	BMC dataset	80	85	82
3	BioScope	75,35	68,18	71,59
4	BioScope	90,81	76,17	82,82
5	BioScope (corrected)	97,63	77,46	86,39

Table 2: Results reported in different publications and concerning the recognition of speculations in Scientific Full Text Papers, representing a part of the BioScope corpus: (1) (Szarvas, 2008), (2) (Kilicoglu and Bergler, 2008), (3) (Morante and Daelemans, 2009), (5) (Özgur and Radev, 2009), (5) BioExcom in this study

	Corpus	Precision	Recall	F-Measure
1	BioScope	90,81	79,84	84,77
2	BioScope	95,56	88,22	91,69
3	BioScope (corrected)	99,39	83,93	91,01

Table 3: Results reported in different publications and concerning the recognition of speculations in Scientific Abstracts, representing a part of the BioScope corpus: (1) (Morante and Daelemans, 2009) (2) (Özgur and Radev, 2009), (3) BioExcom in this study

6 Conclusion and Perspectives

Our aim was to test on a large scale (the manually annotated BioScope corpus) the rule based system BioExcom that automatically annotates speculations in biomedical papers. We observed an important disagreement between the two annotations and as a result, treated it manually. We put forward three principal reasons for the differences in annotation:

• The lack of certain linguistic markers in BioExcom (false negative): some of them have to be added in the system (for example "*seem*" or "*it appears*"). Some other markers are too ambiguous to be relevant (for example "*could*" without a positive clue).

• An error in the annotation of BioExcom or in BioScope (false positives): this is relatively rare, especially for BioExcom, which favors the Precision.

• The difference of criteria used to determine whether a sentence is speculative or not: this is the main reason and we discuss it hereafter.

Obtaining a good manual annotation is a recurrent problem in semantic annotation (Uren et al., 2006) and some studies have studied the disagreement in human judgment (Veronis, 1998). This phenomenon is undoubtedly found here since Szarvas et al. (2008) and Desclés et al. (2009) have reported the difficulty even for experts to detect speculations. But the disagreement between BioExcom and BioScope has to be seen almost as a conflict concerning the meaning given to the word "*speculation*" and if we agree on its definition, we demonstrated the efficiency of BioExcom.

Indeed, in order to provide good tools to biologists, computational linguists seek to extract only definite statements and consequently try to remove or to present separately hedging. Some of them have also tried to indicate the degree of uncertainty, in order to better characterize hedging. Despite the promising use of weighting of hedging cues (Kilicoglu and Bergler, 2008), the task of determining the strength of speculation appears to be difficult, even for biologists (Light et al., 2004). In another hand biologists can also consider speculation as a source of knowledge (for example, Thompson et al. (2008) categorize speculation as a type of knowledge), giving actual trends, new ideas and hypothesis useful for them. In this view, BioExcom extracts them according to more restrictive criteria and categorizes them into "prior" and "new" in order to better highlight speculations. Knowing the new speculations of a paper can reveal some of the real new output of it and so help to decide which paper is worth spending time. The categorization into prior speculation highlights the emergence of an idea which is taken into consideration by the scientific community and thus can also, at least partially, give an indication of its importance among the huge amount of speculations in the literature.

The availability of the corpora (raw BioExcom annotations and re-annotations of BioScope according to these criteria) could help to better take into account these views. In particular, these corpora will be useful to improve the rules of BioExcom for detecting speculations. And many of the other sentences belonging to hedging and discarded during the re-annotation process (previously annotated or not in BioScope) will serve to develop other semantic categories such as *demonstration*/*conclusion* or *deduction*.

References

Alrahabi M (2010) EXCOM-2: a platform for automatic annotation of semantic categories: Conception, modeling and implementation. Application to the categorisation of quotations in French and Arabic, *PhD*, University of Paris-Sorbonne.

Alrahabi M, Desclés JP (2008) Automatic annotation of direct reported speech in Arabic and French, according to semantic map of enunciative modalities. *In 6th International Conference on Natural Language Processing, GoTAL*, Gothenburg, Sweden, pp 41-51

Blagosklonny MV, Pardee AB (2002) Conceptual biology: unearthing the gems. *Nature* 416: 373

Brent R, Lok L (2005) Cell biology. A fishing buddy for hypothesis generators. *Science* 308: 504-506

Collier, N., Park, H., Ogata, N., Tateishi, Y., Nobata, C., Ohta, T., Sekimizu, T., Imai, H., and Tsujii, J. (1999), "The GENIA project: corpus-based knowledge acquisition and information extraction from genome research papers", *Proc. Annual Meeting of the European Association for Computational Linguistics* (EACL-99), Bergen, Norway, June 8-12, pp.271-272.

Desclés JP, (2006) Contextual Exploration Processing for Discourse Automatic Annotations of Texts. *In FLAIRS 2006*, invited speaker, Melbourne, Florida, pp 281-284

Desclés J, Alrahabi M, Desclés JP (2009), BioExcom: Automatic Annotation and categorization of speculative sentences in biological literature by a Contextual Exploration processing, *In Proceedings of the 4th Language & Technology Conference* (LTC), Poznań, Poland

Djioua B, Flores JG, Blais A, Desclés JP, Guibert G, Jackiewicz A, Le Priol F, Nait-Baha L, Sauzay B (2006) EXCOM: an automatic annotation engine for semantic information. *In FLAIRS 2006*, Melbourne, Florida, 11-13 mai, Melbourne, Florida, pp 285-290

Hyland K (1995) The author in the text: Hedging Scientific Writing. *Hong Kong papers in linguistics and language teaching* 18: 33-42

Kilicoglu H, Bergler S (2008) Recognizing speculative language in biomedical research articles: a linguistically motivated perspective. *BMC Bioinformatics* 9 Suppl 11: S10

Lakoff G (1972) Hedges: a study in meaning criteria and the logic of fuzzy concepts. *In Chicago Linguistics Society Papers*, 8:183-228

Light M, Qiu XY, Srinivasan P (2004) The Language of Bioscience: Facts, Speculations, and Statements in Between. *In HLT-NAACL*, ed, *Workshop On Linking Biological Literature Ontologies And Databases*, pp 17-24

Medlock B, Briscone T (2007) Weakly Supervised Learning for Hedge Classification in Scientific Literature. *In AfC Linguistics, ed, Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, Prague, Czech Republic, pp 992-999

Morante R, Daelemans W (2009) Learning the scope of hedge cues in biomedical texts, *In Proceedings of the Workshop on BioNLP*, pp 28-36, Boulder, Colorado, USA, June 2009, ACL

Özgür A, Radev DR (2009) Detecting Speculations and their Scopes in Scientific Text, *In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp 1398-1407, Singapore, 6-7 August, 2009

Szarvas G (2008) Hedge classification in biomedical texts with a weakly supervised selection of keywords, *In Proceedings of ACL-08: HLT*, pp 281-289, Columbus, Ohio, USA, June 2008

Szarvas G, Vincze V, Farkas R, Csirik J (2008) The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts. *In BioNLP ACL-2008* workshop

Thompson P, Venturi G, McNaught J, Montemagni S, Ananiadou S (2008) Categorising modality in biomedical texts. *Proceedings of the LREC 2008 Workshop on Building and Evaluating Resources for Biomedical Text Mining.* Marrakech, Morocco 2008.

Uren V, Cimiano P, Iria J, Handschuh S, Vargas-Vera M, Motta E, Ciravegna F (2006) Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *In Web Semantics: Science, Services and Agents on the World Wide Web* Volume 4, Issue 1, January 2006, Pages 14-28

Véronis, J. (1998) A study of polysemy judgements and inter-annotator agreement, Programme and advanced papers of the *Senseval workshop* (pp. 2-4). Herstmonceux Castle (England)

Wilbur WJ, Rzhetsky A, Shatkay H (2006) New directions in biomedical text annotation: definitions, guidelines and corpus construction. *BMC Bioinformatics* 7: 35