Linguistically Motivated Negation Processing: an Application for the Detection of Risk Indicators in Unstructured Discharge Summaries

Caroline Hagege

Abstract—**The paper proposes a linguistically motivated approach to deal with negation in the context of information extraction. This approach is used in a practical application: the automatic detection of cases of hospital acquired infections (HAI) by processing unstructured medical discharge summaries. One of the important processing steps is the extraction of specific terms expressing risk indicators that can lead to the conclusion of HAI cases. This term extraction has to be very accurate and negation has to be taken into account in order to really understand if a string corresponding to a potential risk indicator is attested positively or negatively in the document. We propose a linguistically motivated approach for dealing with negation using both syntactic and semantic information. This approach is first described and then evaluated in the context of our application in the medical domain. The results of evaluation are also compared with other related approaches dealing with negation in medical texts.**

Index terms—**Negation detection, discharge summaries, dependency parsing.**

I. Introduction

NEGATION is commonly used in natural language texts and is a challenge for general tasks of information extraction. In the medical domain, in particular, specific efforts for the annotation (see [13]), the description (see [10]), and the processing of negation (see [9] and [11]) have been made in the recent years. One evident application of processing negation is to make it possible to distinguish factual information from non-factual information expressed in the texts. This processing will benefit the classical tasks of information extraction like question-answering, summarization (where usually one wants to give priority to positive information). Furthermore, according to [8], explicit knowledge of what is negated may be also useful for a wide range of scenarios in the medical and biomedical domain.

In this paper, we present an algorithm which is able to detect negated information in French hospital discharge summaries. This work has been developed within a larger system detecting occurrences of hospital acquired infections (HAI) in texts. An overall presentation of the project can be found in [12]. One of the processing steps of the system is the extraction of terms and expressions which correspond to risk indicators for HAI. We want thus to be able to distinguish between factual and non-factual risk indicators. We focus in this paper on the negation processing task, which is integrated in the full system.

II. Related Work

The abundant literature on the treatment of negation in medical and biomedical texts shows that this is a crucial problem. The BioScope corpus [13] is manually annotated with uncertain and negated information. This work reports that around 13% of sentences in the corpus contain negation. Existing systems dealing with negation in the medical domain use either machine learning algorithms as [9] or rule-based approaches. For all these systems the general goal is the same: finding negative triggers and their scope.

The Negfinder system presented in [11] is a rule-based system that first identifies medical terms, and then replaces them by an UMLS concept identifier. Then, a lexer and context-free negation grammars are applied. The output consists in the original text in which concepts and negation information is marked. [3] also presents a system based on regular expressions aiming at the detection of the presence or absence of a medical finding in texts. More recently, [6] describes a system in which negation together with temporality and experiencer contextual values are processed.

In all these approaches, the general method is to first define a set of trigger expressions. These expressions usually consist in a wide range of simple or complex lexicalized linguistic chunks that may induce a polarity change to the neighboring textual expressions. Then, once these expressions are found in text, a way to determine the scope of the negative trigger is proposed.

Another syntax-based method is presented in [4], where the authors describe the NegHunter system, which detects negation in Clinical Practice Guidelines. NegHunter considers

Manuscript received October 21, 2010. Manuscript accepted for publication January 19, 2011.

The author is with XRCE – Xerox Research Centre Europe, 6 Chemin de Maupertuis, 38240 Meylan, France. (e-mail: Caroline.Hagege@xrce.xerox.com).

a more restricted and universal set of negative triggers compared with the previously cited approaches. As we will see in section 3.2, we adopt a similar approach for detecting negative triggers but with some differences that will be explained later.

Our method for dealing with negation is also rule-based and generalizes the approaches presented above making the following assumptions:

- Negative trigger expressions presented in the related literature can be generalized using parsing and some lexical semantics.
- Negative triggers should be general enough to be used for processing negation in different contexts and domains.
- Syntax is not enough to determine the scope of negation.

Take for instance the following expressions:

Absence d'origine évidente de cette septicémie.
(Absence of evident origin of this septicemia)
Absence de signes de septicémie.
(Absence of signs of septicemia)

Both expressions have the same syntactic structure: a nominal head *absence* followed by a modifying prepositional chunk headed by *origine* in the first expression and headed by *signes* in the second expression. These heads are in turn modified by another prepositional chunk headed by *septicémie* in both cases. We are interested in the fact that a patient has or has not septicemia. In the first case, the expression entails that the patient has septicemia and the negation carried out by *absence* indicates that the origin of septicemia is not clear. In the second case, the patient has no septicemia at all, and the negation carried out by *absence* has to be applied to the whole expression *signes de septicémie.* These examples show clearly the limits of a purely syntax-based approach for dealing with negation. In these examples syntactic structures are exactly the same and only the introduction of lexical semantics makes it possible to process these sentences in an appropriate manner. The need of a semantic processing of negation is also expressed in the approach presented in [4]. In this paper, the authors explain that their approach for dealing with negation is a first processing step which has to be completed with further semantic processing.

III. Negation Processing Methodology

This section details our negation processing methodology. Negation detection is integrated within a more general linguistic processor presented in [1], which deals with discharge summaries for the final purpose of HAI detection.

A. *General Motivation*

Our parser takes as input a text (discharge summary) and provides as output a linguistic representation of this text consisting in tokenization, part-of-speech tagging, chunking and the establishment of dependency links between the linguistic units of the text. Linguistic units consist in feature structures that carry morphological (inflectional), syntactic (part-of-speech, and some sub-categorization information) and some lexical semantic features. The parser also provides a Java API used for the implementation of extensions (as co-reference, temporal processing etc.).

For this specific task we use the French version of the parser that has been tuned for the processing of medical texts (introduction of dedicated medical lexicon and terminology, specific POS disambiguation rules for the medical domain). Our purpose is to detect automatically occurrences of hospital acquired infections in these texts, and one important step is to recognize in the texts risk indicators that may lead to the conclusion of an HAI. These risk indicators can be either medical terms or more complex expressions involving numerical values. Our linguistic processor uses a specialized lexicon (for simple terms) together with local and syntactic rules (for complex terms and numerical expressions) in order to mark all risk indicators belonging to the following classes (see [5] for more details):

- INFECTIOUS_DISEASE
  PRESENCE_OF_FEVER
  DIAGNOSIS
  VIRAL_DISEASE
  ANTIBIOTIC_ADMINISTRATION
  INTERVENTION
  PRESENCE_OF_INFECTIOUS_GERMS

For accurate detection, we want to be able to state if a textual occurrence of a risk indicator is negated or not. We will discard negative occurrences from the list of the potential risk indicators.

B. *Negative Seeds*

Related approaches dealing with negation usually have a first processing step consisting in the detection of what is usually called negative triggers. [10] describes negative cues found in the BioScope corpus. Negative triggers used by the NegEx algorithm presented in [3] and are publicly available. They consider expressions like *without any evidence of, without evidence, without indication of, without sign of* as negative triggers. If we examine these expressions carefully, we can see that they correspond to the following pattern: they are prepositional phrases (PP) introduced by *without* and the nominal head of these PP correspond to one of the nouns *evidence, indication,* or *sign.* All these nouns are in the same semantic field (for instance in the synonym dictionary available on-line at http://dico.isc.cnrs.fr/dico/en/search).

Instead of considering negative triggers, we decided to consider only what we call negative seeds. Negative seeds consist in a small set of linguistic units, which have the property of negating the syntactic heads they are linked to. The negative seeds are very general and universal and can be used as negation introductors for all kinds of documents and domains. We differ here from [4] (which also considers more general negative triggers) by not considering any verb or noun like *absence* for which negation is induced by the lexical semantics attached to the word. The list of our 14 negative seeds is given here exhaustively. Negative seeds are presented according their distributional properties.

1. Determiners (*aucun, ni, pas de, point de, nul*) which negate the nominal head they determine (e.g. *aucune* infection). This corresponds to the *DETERM* dependency calculated by the parser
2. Adjectives (*nul, inexistant, négatif*) which either negate the nominal head they qualify (e.g. infection *inexistante*) or, when they are used as subject complements, they negate the subject of the copulative verb (e.g. l'infection est *inexistante*). This corresponds to the *NMOD_POSIT1* dependency calculated by the parser.
3. Discontinuous negation adverbs (*ne ...pas, ne...aucun, ne...point, ne...plus*) which negate the verbal predicate situated either in the discontinuous part of these adverbs (for simple forms) or on the right of the adverb (for participial verbal forms). This corresponds to a *VMOD_POSIT1* dependency calculated by the parser. A restriction is added in order to avoid taking as negative seeds these adverbs when they are modified by other adverbs as *presque*, *quasiment* (almost). For instance, *Il **n'**a presque **pas** de fièvre* means *he has almost no fever*.
4. A simple adverb (*non)* which negates its head. This head can be an adjective, a past participle and sometimes a noun.
5. A preposition (*sans*) which always negates the nominal head of the prepositional phrase they introduce. This corresponds to the *PREPD* dependency calculated by the parser.

All these linguistic elements change the polarity of the syntactic head that is in a direct dependency relation with them.

For instance, in the following expressions negative triggers are indicated in bold and negated syntactic heads are underlined.

*Le patient **n'**a **p**as présenté de fièvre*
(The patient did not show any fever)
***Aucun** signe d'infection à ce jour.*
(No sign of infection this day)

Concretely, during parsing, these negative seeds create a unary relation *NEGAT* on the verbal or nominal head associated to them. Taking the two examples mentioned above, two unary relations are thus calculated: *NEGAT(présenté) and NEGAT(signe).*

It is important to note, that since we work with dependency relations, the fact that the negated head is on the right or on the left of the negative seed is not a concern for us.

C. Semantic Fields and Their Polarity

We also consider a subset of lexical units belonging to specific semantic fields. As our final purpose is to be able to distinguish if a term mentioned in text is attested or not attested, we are interested in words belonging to semantic fields denoting the existence, the evidence, the continuation of a fact or an event. More specifically, we consider the following semantic fields:

- existence/non existence
  evidence/non evidence
  continuity/break
  augmentation/diminution

Nouns and verbs belonging to these fields will have an a-priori polarity associated to them. Intuitively, a noun stating the existence a fact (like *sign* or *existence)* will have a positive polarity, and on the contrary, a noun like *absence* will carry a negative polarity. The collection of lexical units belonging to these fields has been compiled using the online synonym dictionary for French developed by the Caen University (http://www.crisco.unicaen.fr/cgi-bin/cherches.cgi). We established a list of 122 verbs and nouns. They are coded in the lexicon of our linguistic processor using Boolean features corresponding to the above mentioned semantic fields.

For instance, the verb *attester (attest)* and the noun *preuve (proof)* belong to the *evidence* semantic field. As such, they have the boolean feature *[evidence:+]* associated to the corresponding lexical entries. The noun *persistance (persistence)* is of *continuity* semantic field and bears the feature *[continuity:+]*, and the noun *fin (end)* corresponds to a noun of non-continuity semantic field bearing thus the feature *[continuity:-]*. Note that these semantic fields may be only relevant to one specific reading of a semantically ambiguous lexical unit. However, because we deal with a specific domain, semantic ambiguity is here limited.

We can in our linguistic processor generalize over features carried by the lexical entries. For instance, we can state that any feature *[evidence:+]* implies a feature *[polarity:+]*. These kinds of generalizations are performed in configuration files read by the parser. As a result, all the lexical entries coded with the above-mentioned features related with semantic fields will have an associated *polarity* feature which can have the value + (for positive polarity) or the value – (for negative polarity).

We can then propagate polarities in order to finally detect what is negated or not.

D. Polarity Propagation

Syntactic negation (expressed by the *NEGAT*relation introduced by negation seeds) and a-priori polarities of lexical heads are then combined in order to propagate negative/positive polarity information from one head to its complement.

Two simple rules for polarity propagation are used:

1. If a linguistic head has a NEGAT and if it has an a-priori polarity associated, then this polarity is inverted.
2. If a linguistic head has no NEGAT relation but bears information on polarity and if its modifier also has an explicit polarity, then polarities are combined (as it is explained later), and a new polarity is given to the modifier.

1) Polarity Inversion due to Syntactic Negation

The first rule corresponds to the fact that a syntactic negation marker inverts the polarity of the negated lexical unit. It can be illustrated by the following expression: *aucun signe d'infection* (*no sign of infection*).

The negative seed *aucun* has created a NEGAT relation on the word *signe*. This word has a feature *[polarity:+]* since it is a noun of the semantic field *evidence*. In this specific context, the polarity will be inverted and the new value of the feature *polarity* will become -.

In a similar way for the word *absence* in the context *pas d'absence de signe d'infection*, the negation seed *pas de* creates a NEGAT relation to the word *absence* which has an a-priori *[polarity:-]* feature. As a result, in this specific context, as polarity will be inverted, *absence* will finally bear the feature *[polarity:+]*

The following statement is added in the grammar rule files read by the parser. This statement says that anything having the feature *[polarity:+]* (expressed by the first line of the rule) and NEGAT relation (first condition in the second line), will have first the feature polarity suppressed (expressed the second condition *#1[polarity=~]*) and then the feature polarity is set to – (expressed by the last condition of the expression *#1[polarity=-]*). A similar statement changing *[polarity:-]* into *[polarity:+]* for negated lexical heads is also present in the grammar files.

```
| #1[polarity:+] |
if ( NEGAT(#1) & #1[polarity=~] & #1[polarity=-] ) ~
```

2) Polarity Combination from Head to Modifier

Rule 2) mentioned above expresses the fact that the polarity carried by a syntactic head may have influence on the polarity of its complement. Intuitively, in an expression like *lack of food*, where *lack* is the syntactic head and *food* the complement, the final status concerning the existence or nonexistence of *food* is ruled out by the fact that *lack* introduces semantically the idea of absence.

Polarity propagation is implemented taking advantage of the general syntactic dependencies computed by the parser.

Two possibilities can occur:

1. The argument or modifier of a lexical head with a polarity also has an a-priori polarity. In this case the polarity of the argument/modifier is changed according to table 1. Polarity propagation will then once again be applied on this argument/modifier.
2. The argument or modifier of a lexical head with a polarity has no a-priori polarity. In this case, the argument/modifier will be negated if the polarity of the lexical head is – or not negated if the polarity of the lexical head is +. Polarity propagation stops on this modifier.

Note that there is an order in the choice of arguments/modifiers for polarity propagation:

– Arguments are taken before modifiers (which means that for a verb, its object complement will be considered before any kind of modifying PP).

In case of multiple modifier choice, the left-most modifier is chosen.

Returning to the example *il n'y a pas d'absence de signe d'infection*, the syntactic negation of the verb *avoir* creates a *NEGAT(a)* relation. As *absence* has an a-priori polarity *[polarity:-]*, the initial polarity attached to *absence* is changed and becomes *[polarity:+]*. The word *absence* is in turn modified by the word *signe* which has an a-priori polarity set to +. The combination of both *[polarity:+]* gives a final *[polarity:+]* to the word *signe* (according to Table 1). Finally, *infection*, which has no a-priori polarity, modifies *signe.* It is not negated because it modifies a lexical unit with *[polarity:+]* feature.

TABLE I
POLARITY COMBINATION

HEAD / MODIFIER	Polarity:+	Polarity:-
Polarity:+	Polarity:+	Polarity:-
Polarity:-	Polarity:-	Polarity:+

E. Negation Focus

Negation focus is the final unary relation that is established when polarity propagation stops.

This propagation stops in two situations:

1. A lexical unit with a polarity has no complements or modifiers.

The complement or modifier of the lexical unit bearing a polarity has no a-priori polarity associated to it.

The first case can be illustrated by the following example:

Il ***n'****y a* ***pas*** *d'augmentation.*
(There is no increase)

The verb *a* bears a negative polarity because of the syntactic negation. Its complement *augmentation* has an a-priori positive polarity and receives a negative polarity during the polarity propagation process. Propagation then stops in the absence of any modifier of *augmentation*. In this case, because *augmentation* has a negative polarity which cannot be propagated, it will correspond to what we call the negation focus. The parser produces a unary dependency *NON(augmentation).*

2. The second case can be illustrated as follows:

On ne retrouve pas d'infection suite aux examens.
(No infection was found after the examinations).

The verb *retrouver* receives a negative polarity as it is involved in a unary *NEGAT* relation created by the negation seed *ne...pas*. The word *infection* has no a-priori polarity associated to it but it is a complement of *retrouver [polarity:-]*. The propagation stops on the word *infection* which is also the focus of the negation. A unary dependency *NON(infection)* is also created by the parser.

An important issue for polarity propagation and negation focus detection is the fact that an accurate PP attachment is necessary in order to get good results. For instance in a case like *Il n'y a pas de suspicion depuis la semaine dernière. (There has been no suspicion since last week)*, the temporal PP is attached to the main verb *avoir* and not to the noun *suspicion*. If an error of PP attachment occurs, polarity propagation would be wrong and the final negated element would be *semaine* and not *suspicion*.

IV. Experiment and Evaluation

In order to test our approach, we first trained the system in the following way. We perform two runs of the same set of texts.

– The first run extracts all risk indicators without using any information regarding negation (we disabled the lexical enrichment and the grammar rules for negation propagation). As a result, any occurrence of risk indicators is extracted regardless of the fact if they are negated or not.
– The second run uses the same system but enriched with negation processing. In this case, only non negated risk indicators (according to our system) are extracted.

The outputs of these two runs contain the initial text with the risk indicators annotated and colored. The two files are aligned and compared. Any difference between the two runs are examined and verified. During the training phase, we add some extra lexical entries, consider new negation seeds, and tune some rules.

After training, we perform a test in order to evaluate the accuracy in detecting negative risk indicators in our discharge summaries. We took a set of 110 unseen discharge summaries coming from different hospitals and different care units 42) documents for an intensive care unit, 50 documents for .(orthopedics and 18 documents for digestive surgery

These documents were first processed using the system without negation processing in order to detect and mark all possible occurrences of HAI risk indicators. All the marked occurrences were then verified manually and the annotator decided for each of them if they were negated or not. As we only treated negation and not modality, we consider that uncertainty is to be annotated as positive and not as negative. Furthermore, we did not take into account temporality. As a result, any mention of a future or past occurrence of a risk indicator is considered as positive if it is positively stated. We then process automatically the same documents with our system enriched with negation processing and compare the automatic and manual annotations. We obtained the following results:

TABLE II
EXPERIMENTAL RESULTS

	Manually annotated negative risk indicators	Manually annotated positive risk indicators
System annotated negative risk indicators	True positives (TP) 174	False positives (FP) 8
System annotated positive risk indicators	False negatives (FN) 6	True negatives (TN) 2,255

Precision, recall, specificity and accuracy are then calculated. We obtained the following figures:

Precision: TP/(TP+FP) 95.6%
Recall/Sensitivity = TP/(TP+FN) = 96.6%
Specificity = TN/(TN+FP) = 99.6%
Accuracy = (TP+TN)/(TP+TN+FP+FN) = 99.4%

These figures show that we obtain very good results. Compared with [11] which performs a comparable evaluation (verification of negated terms) our results are higher (specificity obtained was 97.7% and sensitivity was 95.3%). However, we do not know exactly the kind of texts that were processed in [11], and we only consider a subset of terms which may lead to less variety in expressing negation.

V. Discussion

The method we present shows to be very effective for detecting negative terms on the kind of medical documents we processed (French discharge summaries). The good results we obtain is explained by the fact that we make use of both syntactic and semantic information. Furthermore, because our underlying syntactic knowledge is expressed in terms of dependencies, the distance between words for finding the negation scope is not a concern. Our algorithm is completely integrated within our general purpose linguistic processor. However, the approach is easily adaptable to any other dependency parser. One of the advantages of our method is that it treats syntactic and lexically induced negation. Double negation, although not very frequent in medical narratives is processed naturally and straightforwardly, which is not the case in related approaches ([2] states in the discussion section that double negation is a problem for their system).

However, in this work we restricted the analysis to simple negation (negative conditional expressed by expressions like *either...or*, and uncertainty introduced by tense and modality are not considered). Examples of these more complex negation cases can be found in [10] and it would be interesting to enhance our system in order to take them into account.

Since we use an existing dependency parser and since the specific lexical and syntactic coding is very limited (addition of features on approximately 120 words and 6 additional rules in our grammar), this approach is easily portable for other languages for which we have a dependency parser[1]. Our results are however very dependent on the parser accuracy. PP attachment is one of the key issues, and it may lead to

[1]At least for all romance languages, English and German.

erroneous polarity propagation. Part-of-Speech disambiguation errors may also be a problem as they impact the computing of the dependencies used for polarity propagation. Furthermore, lexical semantic ambiguity can also be a concern if we enlarge this approach to other domains. This kind of ambiguity can lead to erroneous attribution of a-priori polarities, which will impact the correct computing of the negated element.

VI. Conclusion

We have presented a method for dealing with negation in unstructured medical discharge summaries written in French. The method we propose makes use of both syntactic and semantic information and is integrated within a larger linguistic processor for unstructured texts. This approach is suitable to other languages and should be easily adaptable, as coding effort to integrate negation processing in the parser is limited. One of the advantages of our approach is that it treats in a homogeneous way negation expressed syntactically and negation induced lexically.

The next step will be to test this approach to medical texts that are not discharge summaries and even to texts in other domains. We believe that we can extend this approach to other domain-dependent texts (possibly with some changes in the lexical coding). We also would like to apply this approach to the treatment of English medical texts in order to take advantage of already existing annotated resources for the evaluation and comparison of the results with other existing systems. Finally, we would like to enlarge negation detection to a more general system of factuality detection, which will take into account modality, conditionality and uncertainty.

References

[1] S. Aït-Mokhtar, J.P. Chanod, and C. Roux, "Robustness beyond Shallowness: Incremental Deep Parsing," *Natural Language Engineering*, 8, pp.121-144, 2002.

[2] P.L. Bikin, S.H. Brown, B.A. Bauer, C.S. Husser, W. Carruth, L.R Bergstrom, and D.L. Wahner-Roedler, "A controlled trial of automated classification of negation from clinical notes," *BMC Medical Informatics and Decision Making*, S:13, 2005.

[3] W.W Chapman, W. Bridewell, P. Habury, G.F. Cooper, and B.G. Buchanan, "A simple algorithm for identifying negated findings and diseases in discharge summaries," *Journal of Biomedical Informatics*, 34, pp. 301-310, 2001.

[4] S. Gindl, K. Kaiser, and S. Miksh, "Syntactical Negation Detection in Clinical Practice Guidelines," in Andersen, S.K.; Klein, G.O.; Schulz, S.; Aarts, J.; Mazzoleni, M.C. (eds.) *eHealth Beyond the Horizon – Get IT There. Proc. of the 21st International Congress of the European Federation for Medical Informatics (MIE 2008)*, Göteborg, Sweden, IOS Press, 2008, pp. 187-192.

[5] C. Hagège, P. Marchal, Q. Gicquel, S. Darmoni, S. Pereira, M-H. Metzger, "Linguistic and Temporal Processing for Discovering Hospital Acquired Infection from Patient Records," in *Proceedings of the 2nd International Workshop on Knowledge Representation for Health Care (KR4HC-2010)*, Lisbon, Portugal, 2010.

[6] H. Harkema, J. N. Dowling, T. Thomblade, and W. Chapman, "ConText: An Algorithm For Determining Negation, Experiencer, and Temporal Status from Clinical Reports," *Journal of Biomedical Informatics*, 42, pp. 839-851, 2009.

[7] Y. Huang and H.J. Lowe, "A Novel Hybrid Approach to Automated Negation Detection in Clinical Radiology Reports," *Journal of the American Medical Informatics Association (JAMIA)*, vol. 14(3), pp. 304-311, 2007.

[8] Y. Krallinger, "Importance of negations and experimental qualifiers in biomedical literature," in *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, Uppsala, Sweden, 2010, pp. 46-49.

[9] R. Morante and W. Daelemans, "A metalearning approach to processing the scope of negation," in *Proceedings of CoNLL 2009*, 2009, pp. 21-29.

[10] R. Morante, "Descriptive analysis of negation cues in biomedical texts," in *Proceedings of the Seventh International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, 2010, pp. 1429-1436.

[11] P.G. Mutalik, A. Deshpande, and P.M. Nadkarni, "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 (JAMIA)*, vol. 8(6), pp. 598-609, 2001.

[12] D. Proux, P. Marchal, F. Segond, I. Kergoulay, S. Darmoni, S. Pereira, Q. Gicquel, M-H. Metzger, "Natural Language Processing to detect Risk Patterns related to Hospital Acquired Infections," in *Proceedings of RANLP 2009*, Borovetz, Bulgaria, 2009, pp. 865-881.

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