

A New framework for Annotating Semantic Relations in Corpora

Mani Ezzat, Thierry Poibeau

▶ **To cite this version:**

Mani Ezzat, Thierry Poibeau. A New framework for Annotating Semantic Relations in Corpora. Recent Advances in Natural Language Processing, 2011, Bulgaria. hal-00666771

HAL Id: hal-00666771
https://hal.science/hal-00666771

Submitted on 6 Feb 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A New Scheme for Annotating Semantic Relations between Named Entities in Corpora

Mani EZZAT[(1)(2)]
[(1)]INaLCO-CRIM
49bis av. Belle Gabrielle
75012 Paris
[(2)] ARISEM
1-5 rue Carnot
91883 Massy
mani.ezzat@gmail.com

Thierry POIBEAU[(3)]
[(3)] LaTTiCe-CNRS and ENS
1, rue Maurice Arnoux
92120 Montrouge, France
thierry.poibeau@ens.fr

Abstract

Although several studies have developed models and type hierarchies for named entity annotation, no such resource is available for semantic relation annotation, despite its utility for various applications (e.g. question answering, information extraction). In this paper, we show that there are two issues in semantic relation description, one concerning knowledge engineering (what to annotate?) and the other concerning language engineering (how to deal with modality and modifiers?). We propose a new annotation scheme, making it possible to have both a precise and tractable annotation. A practical experiment shows that annotators using our scheme were able to quickly annotate a large number of sentences with very high inter-annotator agreement.

1 Introduction

A large number of natural language applications (e.g. information extraction, question answering, automatic summarization) require a precise analysis of the linguistic content of the text. Since the Message Understanding Conferences in the 1990s, there is a general agreement on the different steps required to perform this analysis: *i)* relevant elements (mostly named entities) are first recognized and tagged, then *ii)* relations between these elements are extracted. This generic schema does not preclude the existence of other steps in the analysis (*e.g.* anaphora resolution, discourse structure recognition), but the recognition of basic elements and relations between them is nevertheless a shared basis among a large number of systems (Jurafsky and Martin, 2009).

This of course explains why there has been an increasing amount of research both on named entity recognition and on relation analysis in the last 20 years (MUC6, 1995; Appelt and Martin, 1999). However, the maturity of these two tasks differs to a large extent. As for named entity recognition, a large number of tools, data and gold standard are available for very different languages. The success rate is often above .9 or even .95 F-measure for major categories (person's names, location's names) in newspapers (Collins and Singer, 1999). Entity types are to a certain extent normalized and formalized in large hierarchies (see for example the hierarchy proposed by Sekine which is now a de facto standard (Sekine et al., 2002)).

In comparison, it is interesting to observe that only a few annotated corpora and no real gold standard exist for semantic relations[1]. A first explanation is that relation analysis largely depends on the task and on the kind of corpora being analyzed. However, we do not think that this is enough to explain the current situation: for example question-answering systems are supposed to address any kinds of questions and thus require a generic approach for relation analysis.

It is of course difficult to normalize the set of all possible relations. The clusters of verbs described in Wordnet (synsets, clusters of near-synonym verbs) (Fellbaum, 1998) or Framenet (clusters of verbs sharing the same argument structure) (Fillmore et al., 2003) are a good basis and our goal is not to propose a new classification of verbs and/or events. Nevertheless, annotation schemes proposed so far do not go beyond simple events themselves. From this perspective, they are inadequate in that they do not provide enough room between a yes or no option (the relation can be identified or not), whereas texts constantly report relations along with modalities, negations, etc.

This is the reason why, in this paper, we propose a tractable annotation scheme allowing one to annotate relations more accurately, with a level of generality that makes our scheme both tractable and extensible. We do not focus on event them-

[1]One of our reviewers suggested previous studies (like (Carlson et al., 2002; Poesio and Artstein, 2008), among several others). However, none of these propose a general scheme for semantic relation annotation. They generally deal with a specific theory (e.g. Rhetorical Structure Theory (Carlson et al., 2002)) or a specific phenomenon (e.g. anaphora resolution (Poesio and Artstein, 2008)). Recent frameworks like ACE take profits of all these studies but a large number of problems remains unsolved, see (ACE, 2008a).

Proceedings of Recent Advances in Natural Language Processing, pages 275–281,
Hissar, Bulgaria, 12-14 September 2011.

selves, but we propose to annotate contextual information for a more thorough analysis of relations expressed in texts. Contextual information includes negations, modalities and reported speech, which are surprisingly poorly represented in most schemes.

We first show why semantic relation annotation is difficult. We then present previous schemes that have been proposed in different frameworks, esp. the Message Understanding Conferences (MUC) (Grishman and Sundheim, 1996) and the Automatic Content Extraction (ACE) conferences, as well as their limitations. We then propose our own scheme and present two experiments showing that annotators using our scheme were able to quickly annotate a large number of sentences with a very high accuracy.

2 Why is Relation Annotation a Difficult Task?

We consider different issues related to semantic relation analysis for event detection. Note that we do not focus on the analysis of lexical relations themselves (e.g. synonymy, meronymy, hyponymy, etc.) since there has been a huge body of research on this topic so far (Cruse, 1986). We consider that lexical semantics is outside the scope of this study, even if this kind of knowledge plays a prominent role in relation analysis (and therefore, in various tasks like information extraction or question answering).

In our view, there are two main issues in relation annotation. The first one is a knowledge engineering problem, the second one a linguistic representation problem.

2.1 A Knowledge Engineering Problem

In most annotation schemes, one has to take a binary decision, *i.e.* whether to annotate or not the relation. There are of course some clear cases. For example, if one is interested in companies acquiring other companies, the following sentences should obviously be considered as positive examples:

- *Google has bought Irish company Green Parrot Pictures in an attempt to improve the quality of video uploaded to YouTube.*
- *Google Buys Mobile Ad Company for $750M*
- *Google buys YouTube for $1.65 billion*

However, most cases are not that clear. Since relations refer to semantic concepts and since those concepts can be difficult to grasp, some examples cannot be tagged accurately without a proper representation of the domain. Some examples are impossible to classify, since the text does not provide enough information to decide if the event (*the purchase*) has been completed or not:

- *Under the Note Purchase Agreement: (a) Dolphin Fund II acquired convertible notes of the Issuer in the aggregate principal amount of $988,900, which convertible notes were convertible, as of January 15, 2003 into 3,826,270 shares of Common Stock*

In this example, the text is complex, refers to domain specific concepts and does not even give the key to the annotator: it is not explicitly said if the result of the transaction means a transfer of the control of the company or not.

All these refer to knowledge engineering problems: most of the time, a good command of domain knowledge is necessary to be able to annotate accurately the different examples in the text. As seen above, this knowledge is not enough when some information is missing or when the text is underspecified.

2.2 A linguistic engineering problem

The linguistic side of the problem is of course not completely disconnected from the knowledge engineering point of view. Let's consider the following examples:

- *Rumors Swirling Around A Google Acquisition of Groupon...*
- *Is Google Buying Groupon For Several Billion Dollars?*

One can see that the first sentence does not refer to a pure fact since the main information is introduced by the phrase "Rumors Swirling Around". In the second example, it is the fact that information appears with a question mark that makes it uncertain.

More generally, relation annotation is inseparable from the analysis of hedge expressions. According to J. Watts (Watts, 2003), hedge expressions are "linguistic expressions which weaken the illocutionary force of a statement: by means of attitudinal predicates (*I think, I don't think, I mean*) or by means of adverbs such as *actually*, etc.". Modal auxiliaries (*may, would...*) should also be include in this list.

- *Google May Acquire Groupon for $6 Billion*
- *If Google would acquire Salesforce.com, it wouldn't be about CRM only.*

In the previous examples, modal auxiliaries make it clear that these sentences are not about facts but possibilities.

For some kinds of events, one can easily find speculations (*e.g.* rumors in the financial domain). Speculations can also use the negative form:

- *Google will NOT acquire Twitter in 2011.*

- *Why Google Will Not Acquire Twitter*

All these examples show that texts are not just about facts but include a lot of other phenomena (modals, negation, etc.) that make annotation a difficult task.

This is of course not new, and a lot of studies have tried to address some of these complex linguistic questions (*e.g.* analyzing the scope of modalities or negations). However, these questions are not directly addressed by most existing annotation schemes, especially the most popular ones.

3 Existing Schemes for Semantic Relation Annotation

Semantic relation analysis is a traditional task for the language understanding community. Despite the lack of generic resources (as seen in the introduction), a large number of works involve relation annotation. As a consequence, relation annotation has been identified as a separable and re-usable task from the Message Understanding Conferences on.

3.1 Early Work in Relation Annotation

Text understanding has been explored since the beginning of natural language processing, and involves since the beginning the recognition of semantic relations between textual entities.

During the 1970s, a number of applications tried to establish a link between texts and databases. This kind of analysis typically requires to be able to connect together different pieces of information. Ad hoc relations were defined and recognized in texts in order to fill databases and subsequently be able to access these databases with natural language queries (see for example the LUNAR system developed by Woods to access databases on materials collected on the moon (Woods, 1973)).

Semantic networks (e.g. conceptual graphs (Sowa and Way, 1986)) provided a framework to standardize the representation of this kind of information, but did not normalize the annotation itself.

3.2 The Message Understanding Conferences (MUC)

The Message Understanding Conferences refer to a series of evaluation campaigns organized by DARPA from 1987 to 1998 (MUC6, 1995; MUC7, 1998). The goal was for DARPA and other funding institutions to be able to track the progress of different strategies for information extraction (*i.e.* the extraction of structured knowledge from unstructured texts). We will not detail here the evolution of MUC during these 12 years, since good overviews are available elsewhere (Grishman and Sundheim, 1996).

What is interesting from our perspective is the fact that for MUC-6, in 1995, named entity recognition was recognized as an independent task. Three other tasks ("co-reference annotation", "template element" and "scenario template") were proposed for evaluation, and these were mainly based on the identification of relevant relations between named entities, and between named entities and their attributes.

Here, the evaluation was clearly task-oriented: a limited number of texts from the targeted domain were carefully selected for evaluation. Modifiers, negations and other hedge expressions were only marginally represented and not really integrated in the annotation framework. Most systems did not take these elements into account, with no major penalty. Of course, this kind of strategy can lead to major errors, which can be a serious problem when the system is used in the real world.

3.3 Automatic Content Extraction (ACE)

Automatic Content Extraction refers to a series of evaluation campaigns held between 2000 and 2008 and organized by the Linguistic Data Consortium (LDC). Contrary to what was done in the framework of MUC, the evaluation is not task-oriented but technology-oriented, in that it is supposed to provide general guidelines that are not limited to a given domain (Doddington et al., 2004; ACE, 2008b).

ACE considers for example issues related to modality (ACE, 2008b). A fact can be tagged as ASSERTED or as OTHER (all other cases). As we have seen in the previous section, there are far more than two cases to consider in order to be able to accurately tag texts. Moreover, the guidelines provide rather unclear rules like "If we think of the situations described by sentences as pertaining to possible descriptions of the world (or as 'possible worlds') then we can think of ASSERTED Relations as pertaining to situations in 'the real world', and we can think of OTHER Relations as pertaining to situations in 'some other world defined by counterfactual constraints elsewhere in the context'" (ACE, 2008a).

The authors give the following example: "*We are afraid Al-Qaeda terrorists will be in Baghdad*". Since "The presence of Al-Qaeda terrorists in Baghdad is a situation being described as holding in the counterfactual world defined by 'our' fears", the example should be consider as being ASSERTED. They also give an example that should not be considered as being ASSERTED: "*If the inspectors can get plane tickets today, then they will be in Baghdad on* Tuesday". This sentence is not ASSERTED because "the inspectors (they) are in Baghdad only in the worlds where they get plane tickets today" (ACE, 2008a).

So a fact is asserted when it is "interpreted relative to the 'Real' world" and not asserted (OTHER) when the fact "is taken to hold in a particular counterfactual world". Finally, "negatively defined relations (e.g. "*John is not in the house*") [should] not be annotated" following the ACE proposal.

In our view, there are several problems with this scheme:

1. there are more than two values to be considered. The distinction between ASSERTED and OTHER is not enough to get a fine grained description of relations in texts (for example, this annotation does not say if the event is completed or ongoing, if it is sure, probable or just possible) . Moreover, it seems important to annotate the source of the assertion when possible;

2. there is no reason to exclude negative events. Moreover, from an applicative point of view, this knowledge is often of paramount importance for the domain (*e.g.* knowing/speculating that Google will not buy Twitter in 2011 may have a major impact on investment people);

3. the notion of real world *vs* counterfactual world is not really operational for the task. It does not provide enough evidence for the annotator to make her decision.

Most recent frameworks do not seem to answer these issues, even for the "event detection" task; they often contain domain specific annotation (Aitken, 2002; Mcdonald et al., 2004; Jayram et al., 2006; Shen et al., 2007; Kim et al., 2008) or focus on a certain type of information (Morante and Daelemans, 2009). So we need to build on the ACE scheme in order to overcome some of its shortcomings.

4 A New Relation Annotation Scheme

Semantic relations correspond to a core event with most of time additional information related to the event. These additional pieces of information are most of the time encoded through negations, modalities and higher level clauses (for reported speech for example). Our contribution addresses these elements.

4.1 Basic Event Encoding

We consider that a semantic relation is part of the linguistic expression of an event. This relation is most of the time expressed by a predicate, either a verb (*Google buys YouTube*) or a noun (*the purchase of Youtube by Google...*). The predicate governs some arguments (*Google, Youtube*) that can be tagged more or less precisely (*arg1, arg2; agent, patient; buyer, target;* etc.). Linguistic descriptions of verb hierarchies provide an accurate basis for this kind of analysis (see Wordnet (Fellbaum, 1998) or Framenet (Fillmore et al., 2003), as detailed above). These hierarchies must be adapted with respect to the domain but they are anyway as far as it can be re-usable.

Existing frameworks like MUC or ACE provided precise guidelines for this kind of information. We build on these guidelines for our experiments.

4.2 Enunciative Modalities

The description of basic events must be completed in order to take into account the different issues we have described above (knowledge engineering as well as linguistic engineering issues). We consider three basic attributes directly associated with relations in order to express the degree of completeness of the event: COMPLETED, ONGOING, POSSIBLE.

- if the process is done and over, it is COMPLETED;
- if the process has begun is not yet accomplished, it is ONGOING;
- if the process has not begun, it is POSSIBLE.

Moreover, the event can be NEGATED (e.g. see *Google will NOT acquire Twitter in 2011*).

The event can also be reported directly or by different sources, which means we have to annotate the relation as being DIRECT (*Google Buys Mobile Ad Company for $750M*) or INDIRECT and, for the latter, we also have to annotate the SOURCE when possible (see for example "*Rumors Swirling Around A Google Acquisition of Groupon*" where the PROCESS is reported, therefore INDIRECT and the "*rumors*" are the source). Table 1 gives some examples along with their annotation.

More detailed annotation schemes are possible, especially to deal with different kinds of modalities (epistemic, deontic, etc.). We do not think it is appropriate to have a so fine grained description as these categories will be inappropriate for most language understanding applications. Note that this more fine grained categorization is not incompatible with our scheme. It just requires that some of the categories are refined.

5 Experiments

We present here a method to quickly extract potential relevant sentences from corpora using collocations. These sentences are then manually annotated in order to check the operability of our scheme.

Sentence	Annotation
Rumors Swirling Around A Google Acquisition of Groupon	POSSIBLE, INDIRECT, SOURCE='rumors'
Google will NOT acquire Twitter in 2011	POSSIBLE, DIRECT, NEGATED
Google Buys Mobile Ad Company for $750M	COMPLETED, DIRECT
Is Google Buying Groupon For Several Billion Dollars?	POSSIBLE, DIRECT
Google announced on Friday that it has entered into an agreement to acquire Widevine	ONGOING, INDIRECT, SOURCE='Google'

Table 1: English examples with annotations.

5.1 Extracting Potentially Relevant Sentences from Corpora

The extraction of relevant sentences from corpora is a long and labour intensive task. Most of the time, one must read a large number of texts in order to find only a few relevant sentences. This is both inefficient and time-consuming.

In order to reduce the time spent on this step, we have developed a series of tools allowing one to retrieve relevant documents and then identify potentially relevant sentences. Our approach is simple and easy to reproduce: the idea is to use collocations as a basis for filtering sentences from corpora. The approach can be compared to previous experiments described for example by Riloff with the AutoSlog system (Riloff, 1993). Information extraction patterns involve arguments that can be used to find relevant predicates and, in turn, relevant predicates can be used to find relevant arguments. The same strategy can be used to identify relevant sentences.

We reproduced this idea by first fixing named entity types. Sentences containing these types are then retrieved if named entities appear within a certain distance (in most experiments we used a sliding window with a distance inferior to 10 between the two named entities) (Freitag, 1998). This technique makes it possible to retrieve a certain number of sentences (the method can be parametrized to adjust the number of retrieved sentences). User studies (made with a representative sample of potential end-users who are not trained linguists) have proven that experts can describe the kind of relations they are looking for and the kind of entities these relations involve. They are practically able to use the tools we have developed and are able to perform their analysis a lot quicker with this approach.

For example, in the case of companies buying other companies, only sentences that contain at least two company names are extracted. This of course eliminates relevant sentences containing less than two company names (esp. sentences containing anaphora) but, after manual inspection, we assume we get a representative set of sentences anyway, since anaphora do not fundamentally change the deep semantic structure. So, even if anaphora are not taken into consideration here, they can be analyzed and integrated in subsequent steps without any problem.

For the company buyout task, the system provided more than 1000 potentially relevant sentences in a few minutes (extracted from a 2.9 million word corpus). It then took less than one hour for an expert to manually check these sentences and discard non relevant ones. More than 50% of the extracted sentences were relevant but this represents less than 5% of the corpus (and always less than 10% of the corpus, even with other domains and relations). This proves that the approach is both efficient and accurate.

5.2 Corpus annotation

Our experiment is based on the previous set of sentences extracted from different sources, mainly from financial newswires and newspapers (see table 2 for some examples). A reduced experiment has been done on English texts (see examples in table 1 and in section 2) but a larger experiment has been done on French, using texts from the same domain. This ensures that our annotation scheme is largely language independent.

This corpus is automatically analyzed using a state-of-the art named entity tagger[2]. Sentences containing two company names are extracted. As a result, one hundred sentences are extracted and these sentences are annotated according to the above scheme by two human annotators.

5.3 Inter-annotator agreement

Interannotator agreement is relatively straightforward to calculate, although there are dependencies between tags (e.g. SOURCE is relevant only in case of INDIRECT speech). For each sentence, we compare the set of tags added by annotator A and by annotator B. If the tags do not fully correspond,

[2]Tha ARISEM named entity recognizer.

Twitter dément la rumeur de rachat par Apple	NEGATED, INDIRECT, SOURCE='rumeur'
Areva a racheté pour 1,62 milliard d'euros la part de Siemens dans la co-entreprise Areva NP, ouvrant la voie à un rapprochement entre Siemens et le russe Rosatom, selon le journal allemand Die Welt, qui cite les porte-parole des deux groupes, s'exprimant dans un document qui sera publié lundi.	COMPLETED, INDIRECT, SOURCE='les porte-parole des deux groupes'
Selon Apple4us, un des plus gros blogs chinois au sujet d'Apple, la firme de Cupertino aurait racheté EditGrid, un service de tableurs en ligne basé à Hong Kong, pour une somme comprise entre 10 et 30 millions de dollars.	COMPLETED, INDIRECT, SOURCE='Apple4us'
Amazon aurait racheté la jeune pousse américaine Touchco basée à New York pour développer son offre de lecteurs de livres numériques Kindle.	POSSIBLE, DIRECT
Le possible rachat du Parisien-Aujourd'hui en France par le groupe Dassault inquiète.	POSSIBLE, DIRECT
La société Acom27 dirigée par Monsieur et Madame Garnot n'a absolument pas été rachetée par les éts Cochet.	NEGATED, DIRECT

Table 2: French examples used for evaluation.

we consider that there is a disagreement. Dependencies between tags are not taken into account. This is not a problem as it penalizes the evaluation, rather that the other way round (i.e. results are lower than they would be if we were taking into account these dependencies).

We then computed Cohen's kappa (Cohen, 1960) and obtained 0.94, which means a near perfect agreement, according to the usual interpretation of Cohen's kappa results (Fleiss, 1981). This proves that our method is both efficient and accurate.

Some sentences are hard to classify between DIRECT and INDIRECT, especially when the event is negated, for example when a company denies rumors (*Twitter dément la rumeur de rachat par Apple — Twitter denies the rumor of a buyout by Apple*). In this case, the experts agreed on NEGATED and INDIRECT. The cases of disagreement are rare and affect quite specific sentences (with negation or with a complex structure); they can all be solved after discussion between domain experts.

However, this scheme does not cover all possible cases and should be extended for specific needs. Since it is open (and built upon existing schemes) it can easily be extended to cover new cases and new applications.

6 Conclusion

In this paper, we have presented an annotation scheme that is more precise that what has been proposed for the MUC and the ACE conferences. Our scheme allows one to quickly annotate relations in texts without sacrificing accuracy.

We have proven this result through an experiment on texts from the financial domain, both in English and in French. Additionally, we have shown that it is possible to quickly retrieve relevant examples just by accessing the corpus with key collocations.

The perspectives are twofold. First, we need to annotate a larger number of texts from different domains to ensure the utility of our scheme. Second, we need to explore different specializations of this scheme, as different needs will probably be expressed in the future to get a more precise annotation, concerning modalities for example.

References

ACE. 2008a. *ACE (Automatic Content Extraction) English Annotation Guidelines for Relations*. Linguistic Data Consortium, Univ. Pennsylvania.

ACE. 2008b. Automatic Content Extraction 2008 Evaluation Plan (ACE08) — Assessment of Detection and Recognition of Entities and Relations Within and Across Documents. In *Proceedings of the Automatic Content Extraction conference*, Gaithersburg.

J. Aitken. 2002. Learning information extraction rules : An inductive logic programming approach. In *15th European Conference on Artificial Intelligence*, Lyon.

Douglas Appelt and David Martin. 1999. Named Entity Extraction from Speech: Approach and Results Using the TextPro System. In *Proceedings of the DARPA Broadcast News Workshop*, pages 51–54, Herndon.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2002. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In Jan van Kuppevelt and Ronnie Smith, editors, *Current Directions in Discourse and Dialogue*. Kluwer Academic Publishers.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.

M. Collins and Y. Singer. 1999. Unsupervised models for named entity classification. In *Proceedings of Conf. on Empirical Methods in Natural Language Processing*, Univ. of Maryland.

D.A. Cruse. 1986. *Lexical semantics*. Cambridge University Press, Cambridge.

G. Doddington, A. Mitchell, M. Przybocki, L. Ramshaw, S. Strassel, and R Weischedel. 2004. Automatic Content Extraction (ACE) Program. In *Proceedings of 4th International Conference on Language Resources and Evaluation (LREC)*, pages 837–840, Lisbon.

Chritiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Charles J. Fillmore, Christopher Johnson, and Miriam Petruck. 2003. Background to Framenet. *International Journal of Lexicography*, 16.3: 235–250.

Joseph L. Fleiss. 1981. *Statistical methods for rates and proportions*. John Wiley, New York.

Dayne Freitag. 1998. Multistrategy learning for information extraction. In *Proceedings of International Conference on Machine Learning (ICML)*, Madison.

Ralph Grishman and Beth Sundheim. 1996. Message understanding conference 6 – A brief history. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pages 466–471, Copenhagen.

T.S. Jayram, R. Krishnamurthy, S. Raghavan, S. Vaithyanathan, and H. Zhu. 2006. Avatar information extraction system. *IEEE Data Engineering Bulletin*.

Daniel Jurafsky and James H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Prentice-Hall.

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

Daniel M. Mcdonald, Hsinchun Chen, Hua Su, and Byron B. Marshall. 2004. Extracting gene pathway relations using a hybrid grammar: the arizona relation parser. *Bioinformatics*, 20:3370–3378.

Roser Morante and Walter Daelemans. 2009. Learning the scope of hedge cues in biomedical texts. *Proceedings of the Workshop on BioNLP BioNLP 09*, page 28.

MUC6. 1995. Proceedings of the 6th Message Understanding Conference.

MUC7. 1998. Proceedings of the 7th Message Understanding Conference.

Massimo Poesio and Ron Artstein. 2008. Anaphoric Annotation in the ARRAU Corpus. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Lisbon.

Ellen Riloff. 1993. Automatically Constructing a Dictionary for Information Extraction Tasks. In *Proceedings of AAAI 1993 (Association for the Advancement of Artificial Intelligence)*, Washington.

Satoshi Sekine, Kiyoshi Sudo, and Chikashi Nobata. 2002. Extended Named Entity Hierarchy. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)*, pages 1818–1824, Las Palmas.

Warren Shen, AnHai Doan, Jeffrey F. Naughton, and Raghu Ramakrishnan. 2007. Declarative information extraction using datalog with embedded extraction predicates. In *Proceedings of the 33rd international conference on Very large data bases*, VLDB '07, pages 1033–1044. VLDB Endowment.

John F. Sowa and Eileen C. Way. 1986. Implementing a semantic interpreter using conceptual graphs. *IBM Journal of Research and Development*, 30(1):57–69.

Richard J. Watts. 2003. *Politeness*. Cambridge university Press, Cambridge.

W. A. Woods. 1973. Progress in natural language understanding: an application to lunar geology. In *Proceedings of the June 4-8, 1973, national computer conference and exposition*, AFIPS '73, pages 441–450, New York, NY, USA. ACM.