Ontology-based linguistic annotation

Philipp Cimiano, Siegfried Handschuh
Institute AIFB
University of Karlsruhe
{cimiano,handschuh}@aifb.uni-karlsruhe.de

Abstract

We propose an ontology-based framework for linguistic annotation of written texts. We argue that linguistic annotation can be actually considered a special case of semantic annotation with regard to an ontology such as pursued within the context of the Semantic Web. Furthermore, we present CREAM, a semantic annotation framework, as well as its concrete implementation OntoMat and show how they can be used for the purpose of linguistic annotation. We demonstrate the value of our framework by applying it to the annotation of anaphoric relations in written texts.

1 Introduction

Linguistic annotation is crucial for the development and evaluation of natural language processing (NLP) tools. In particular machine-learning based approaches to part-of-speech tagging, word sense disambiguation, information extraction or anaphora resolution - just to name a few - rely on corpora annotated with the corresponding phenomenon to be trained and tested on. In this paper, we argue that linguistic annotation can to some extent be considered a special case of semantic annotation with regard to an ontology. Part-of-Speech (POS) annotation for example can be seen as the task of choosing the appropriate tag for a word from an ontology of word categories (compare for example the Penn Treebank POS tagset as described in (Marcus et al., 1993)). The annotation of word senses such as used by machine-learning based word sense disambiguation (WSD) tools corresponds to the task of selecting the correct semantic class or concept for a word from an underlying ontology such as WordNet (Resnik, 1997). Annotation by template filling such as used to train machine-learning based information extraction (IE) systems as (Ciravegna, 2001) can be seen as the task of finding and marking all the attributes of a given ontological concept in a text. An ontological concept in this sense can be a launching event, a management succession event or a person together with attributes such as name, affiliation, position, etc. The annotation of anaphoric or bridging relations is actually the task of identifying the semantic relation between two linguistic expressions representing a certain ontological concept.

Most linguistic annotation tools make use of schema specifying what can actually be annotated. These schema can in fact be understood as a formal representation of the conceptualization underlying the annotation task. Ontologies are formal specifications of a conceptualization (Gruber, 1993) so that it seems straightforward to formalize annotation schemes as ontologies and make use of semantic annotation tools such as OntoMat (Handschuh et al., 2001) for the purpose of linguistic annotation.

The structure of this paper is as follows: Section 2 presents the ontology-based framework for linguistic annotation, and section 3 shows how the framework can be applied to the annotation of anaphoric relations. Section 4 presents CREAM, a semantic annotation framework for the Semantic Web as well as its concrete implementation OntoMat. Finally, section 5 discusses related work, and section 6 concludes the paper.

2 The Ontology-based linguistic annotation framework

An ontology is a formal specification of a conceptualization (Gruber, 1993). A conceptualization can be understood as an abstract representation of the world or domain we want to model for a certain purpose. The ontological model underlying this work is basically the one in (Bozsak et al., 2002). According to this model, an ontology is defined as follows:

Definition 1 (Ontology)
An ontology is a structure $O := (C, \leq_C, R, \leq_R)$ *consisting of (i) two disjoint sets* C *and* R *called concept identifiers and relation identifiers respectively, (ii) a partial order* $\leq_C$ *on* C *called concept hierarchy or taxonomy, (iii) a function* $\sigma : R \rightarrow C \times C$[1] *called signature and (iv) a partial order* $\leq_R$ *on* R *called relation hierarchy.*

In addition, the underlying ontological model also allows to define axioms:

Definition 2 (L-Axiom System)
Let L *be a logical language. An* L*-axiom system for an ontology* O *as defined above is a pair (AI,*α*) where (i) AI is a set whose elements are called axiom identifiers and (ii)* α*: AI* $\rightarrow$ L *is a mapping. The elements of A:=*α*(AI) are called axioms.*

In our ontological framework, a relation r can for example be defined as symmetric, i.e. SYM(r). Now if F-logic (Kifer et al., 1995) is used as the underlying logical language such as in (Staab and Mädche, 2000), the translation of the SYM *axiom identifier* is as follows:

$$\forall R, X, Y\ Y[R \twoheadrightarrow X] \leftarrow SYM(R) \quad (1)$$
$$and\ X[R \twoheadrightarrow Y]$$

In addition, we will also distinguish special type of relations which we will call *attributes*. These are relations with a plain datatype as range, i.e. relations $a \in A$ with signatures of the type $\sigma : A \rightarrow C \times P$, where P is a plain datatype such as a string, an integer, etc.

[1] Here we actually restrict the model to binary relations.

Our framework basically offers three ways of annotating a text with regard to an ontology:

- a linguistic expression appearing in a text can be annotated as an instance of a certain ontological concept $c \in C$
- a linguistic expression in a text can be annotated as an attribute instance of some other linguistic expression previously annotated as a certain concept $c \in C$
- the semantic relation between two linguistic expressions respectively annotated as instances of two concepts $c_1, c_2 \in C$ can be annotated as an instance of relation $r \in R$ if $\sigma(r) = (c_1, c_2)$

The advantages of an ontology-based linguistic annotation framework as described above are the following:

- The formalization of the annotation scheme as an ontology as well as the use of standard formalisms such as RDF (Lassila and Swick, 1999) or OWL [2] to encode it, allow to reuse the scheme across different annotation tools. This meets the interoperability requirement mentioned in (Ide, 2002).
- The specification of the annotation task, i.e. the annotation scheme, can be performed in an arbitrary ontology development environment and thus becomes completely independent of the annotation tool actually used.
- The ontology-based linguistic annotation model offers the kind of flexibility mentioned in (Ide, 2002) in the sense that it is general enough to be applied to a broad variety of annotation tasks.
- The fact that annotation is performed with respect to an ontological hierarchy offers annotators the possibility to choose the appropriate level of annotation detail such that they are never forced to overspecify, i.e. to annotate more specifically than they actually feel comfortable with.

[2] http://www.w3.org/TR/owl-ref/

In addition, a hierarchical annotation offers further possibilities regarding the computation of the agreement between different annotators as well as the evaluation of a system against a certain annotation. In this sense, instead of measuring only the categorial agreement between annotators with the kappa statistic (Carletta, 1996) or the performance of a system in terms of precision/recall, we could take into account the hierarchical organization of the categories or concepts by making use of measures considering the 'hierarchical distance' between two concepts such as proposed by (Hahn and Schnattinger, 1998) or (Mädche et al., 2002).

Furthermore, the use of an ontology-based and thus more semantic framework for linguistic annotation has two further, very interesting properties. On the one hand, the use of an ontology helps to constrain the possible relations between two concepts, thus reducing the amount of errors in the annotation process. For example when annotating *Coreference*-relations in a text, it seems obvious that an event and an entity will never be coreferring and in fact such an erroneous annotation can be avoided if the underlying ontological model actually forbids such an annotation (see below). Furthermore, by using axioms such as described above for example stating that *Coreference* is reflexive, symmetric and transitive - thus representing an equivalence relation - the evaluation of systems becomes much easier and more straightforward when using an inference machine such as (Decker et al., 1999). If an annotator for example annotates the following coreferences: Coreference(A,B) and Coreference(B,C) a system's answer such as Coreference(A,C) will actually be counted as correct due to the fact that *Coreference* is defined as a transitive relation within the ontology.

3 Annotating anaphoric relations

Before showing how our framework can be applied to the annotation of anaphoric relations in written texts, the assumptions underlying our model have to be explained. First, we aim at a more semantic annotation of anaphoric relations than for example described in (Müller and Strube, 2001) because we think that such a model can to some extent be subsumed by the one we propose. In fact, we will understand the term *anaphoric* in a much wider sense in line with (Krahmer and Piwek, 2000) and (van Deemter and Kibble, 2000). They argue for example that coreference is not a necessary property of anaphora such as proposed in (Müller and Strube, 2001). So annotating the relation between two expressions as anaphoric will correspond to the most general relation in our hierarchy. In particular, in our model *Identity* or *Coreference* will only be a special type of anaphoric relation (compare figure 2).

On the other hand, *bridging* will be defined in our framework in line with (Asher and Lascarides, 1999) as "the inference that two objects or events that are introduced in a text are related in a particular way that isn't explicitly stated". Thus *Coreference* or *Identity* can represent an anaphoric relation or more specifically a bridging reference depending on whether the identity relation is explicit or not. Consider the following minimal pair:

(2) John bought a car yesterday. The car was in a good state.

(3) John bought a car yesterday. The vehicle was in a good state.

In example (2), the anaphoric relation is explicit due to the matching heads of the NPs *a car* and *The car*. In (3) the anaphoric or bridging relation is not explicit as world knowledge such as that cars are vehicles is needed to resolve the reference. In the semantics-based model for the annotation of anaphoric relations we propose in this paper, both examples will in fact be annotated as instances of the *Coreference* or *Identity* relation. Consequently, we will completely omit the concept *bridging reference* in the ontology underlying the annotation. In fact, we claim that the classification of an anaphora as a bridging reference, direct anaphora, pronominal anaphora, etc. such as pursued in (Müller and Strube, 2001) can be seen as a byproduct of a more semantic classification as proposed here if additional grammatical information provided by the annotators is available. This grammatical information can be added to the concepts depicted in figure 2 in form of attributes specifying the grammatical form of the expression, i.e. whether it is for example a noun, an NP, a pronoun, a verb or a VP, as well as information about its head, gender or tense. The semantic classification proposed here together with the grammatical information modeled as attributes of a concept will then yield a classification as envisioned by (Müller and Strube, 2001). For example, if two expressions are annotated as *coreferring*, this semantic relation can be further classified as *nominal anaphora* if the referring expression is a pronoun, as *direct anaphora* if the heads of the expression match or as a *bridging reference* otherwise. On the other hand, all the *Non-Identity* relations modeled in the ontology underlying the annotation task will lead to a classification as a bridging reference (compare figure 2). However, it should be mentioned that we do not aim at such a 'grammatical' classification of anaphoric relations. We envision a task as in (Asher and Lascarides, 1999), where bridging reference resolution corresponds to the task of finding the discourse referent serving as antecedent as well as the semantic relation between this discourse referent and the one of the referring expression.

In our model, an expression can be antecedent for more than one referring expression, an assumption which seems to be commonly shared by many annotation schema. However, in our model a certain expression can also refer to more than one antecedent. (Poesio and Reyle, 2001) for instance show that the antecedent of a referring expression can in fact be ambiguous in a way that the overall interpretation of the expression or sentence is not affected. Furthermore, (Poesio and Reyle, 2001) argue that it is not clear whether the addressees of an utterance actually are aware of all the possible antecedents for a certain referring expression, if they underspecify the antecedent of a referring expression in case the overall interpretation is not affected or if they just choose one of the possible antecedents without being aware of the other ones. In any case, a model for the annotation of anaphoric or bridging relations should not *a priori* exclude that referring expressions can have more than one antecedent. Consequently, the annotation of the semantic relation between a referring expression and an antecedent can neither take place at the antecedent nor the referring expression such as in (Müller and Strube, 2001), but in a functional way, i.e. at a virtual edge between them.

The ontology underlying our annotation scheme is depicted schematically in figure 1 We distinguish two types of eventualities: events and states, and model the discourse relations described in (Lascarides and Asher, 1991) as semantic relations between them. In addition, we distinguish between three types of (meta-) entities: sets of entities, intensional entities (van Deemter and Kibble, 2000) and (real-world) entities together with the potential relations such as *member_of*, *part_of*, etc. between them as well as to other types: An *entity* for example can play a certain thematic role in some *event* (compare figure 1).

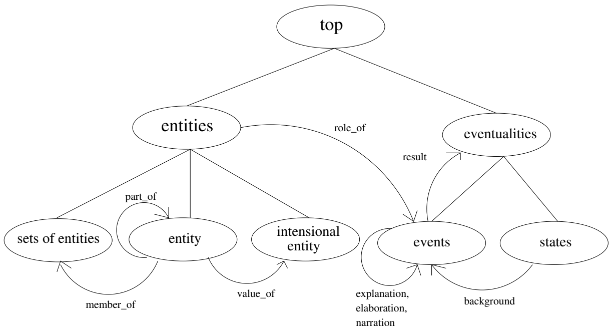

Figure 1: The ontology underlying the annotation scheme

With such a concept hierarchy as well as semantic relations with a precisely defined signature, we can for example overcome annotation problems of intensionality and predication as discussed in (van Deemter and Kibble, 2000). In order to profit from the benefits of a hierarchical annotation, we also define a hierarchy on the semantic relations (see figure 2). Thus if annotators for example feel that there is an anaphoric relation between two linguistic expressions, but can not specify the type of relation, they can choose the most general relation in the hierarchy, i.e. *anaphoric relation*. As mentioned in section 2, the idea is that annotators are never forced to overspecify and can annotate at the hierarchical level they feel comfortable with.

4 CREAM and OntoMat

CREAM is an annotation and authoring framework and OntoMat-Annotizer (OntoMat for short) is its concrete implementation. The framework itself was developed for the creation of ontology-based annotation in the context of the Semantic Web. Its main objective is thus the transformation of existing syn-

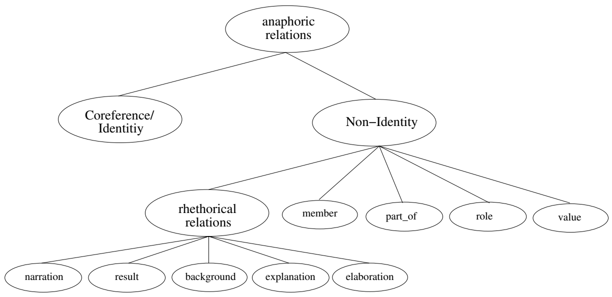

Figure 2: The hierarchical organization of the semantic relations.

tactic resources (*viz.* textual documents) into interlinked knowledge structures that represent relevant underlying information (Handschuh et al., 2001). However, with an apropriate ontology one can also take advantage of the framework and use it for linguistic annotation. In the subsequent section we will explain only the features that are relevant to this purpose.

4.1 CREAM Features

4.1.1 User Interface

OntoMat's *document viewer* visualizes the document contents. The user may easily provide annotations by selecting pieces of text and aligning it with parts of the ontology. The document viewer supports various formats[3] (HTML, PDF, XML, etc.). The *Ontology and Fact Browser* is the visual interface to the ontology and the annotated facts. The annotation framework needs guidance from the ontology. In order to allow for sharing of knowledge, newly created annotations must be consistent with a given ontology. Otherwise, if annotators instantiate arbitrary classes and properties the semantics of these properties remains void and the annotation thus useles.

Both the *Ontology and Fact Browser* and the *document editor/viewer* are intuitive to use: Drag'n'drop helps to avoid syntax errors and typos and a good visualization of the ontology helps the annotators to correctly choose the most appropriate class for an instance (compare figure 3).

4.1.2 Annotation

An annotation in our context is a set of instantiations of classes, relationships and attributes. This instances are not directly embedded into the text, but are pointing to appropriate fragments of the document. The link between the annotation and the document is done by using XPointer (DeRose et al., 2001) as a adressing mechanism. This has some advantages with regards to the flexibility of annotation as it allows (i) multiple annotation (ii) nested annotation and (iii) overlapping annotation of text segments.

4.1.3 Annotation Inference Server

The annotation inference server reasons on the instances and on the ontology. Thereby, it also takes into account the axioms modeled within the ontology and can thus be used in the evaluation of a certain system such as described in section 2. We use Ontobroker's F-Logic-based inference engine (Decker et al., 1999) as annotation inference server. The F-Logic inference engine combines ordering-independent reasoning in a high-level logical language with a well-founded semantics.

4.1.4 Storage

CREAM supports different ways of storing the annotation. This flexiblity is given by the XPointer technique which allows to separate the annotation from the document. Hence, the annotations can be stored together with the document. Alternatively or simultaneously it is also possible to store them remote, either in a separate file or in the annotation inference server.

4.2 Annotaing anaphoric relations

The ontology described in section 3 is available in the form of DAML+OIL[4] classes and properties, in OWL, as pure RDF-Schema and in F-Logic. In the following, we shortly explain how OntoMat can be used for the creation of instances consistent with the ontology described in section 3.

Figure 3 shows the screen for navigating the ontology and creating annotations in Ontomat. The right

[3]The current OntoMat implementation is restricted to HTML/XHTML and plain text. A support for PDF is in development.

[4]http://annotation.semanticweb.org/ontologies/AnaphOnto.daml

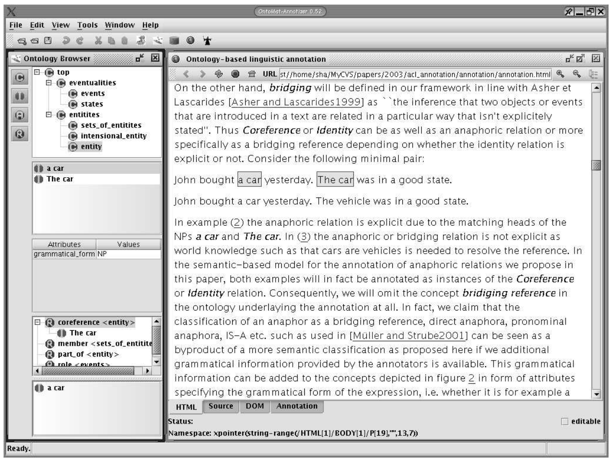

Figure 3: Annotation Tool Screenshot.

pane displays the document and the left panes shows the ontological structures contained in the ontology, namely classes, attributes and relations. In addition, the left pane shows the current semantic annotation knowledge base, i.e. existing class instances, attribute instances and relationship instances created during the semantic annotation. First of all, the user browses a document by entering the URL of the web document that he would like to annotate. Then he loads the corresponding ontology into the ontology browser. He selects a text fragment by highlighting it. There are two possibilities for the text fragment to be annotated: as an instance or as a relation. In the case of an instance, the user selects in the ontology the class where the text fragment fits in, e.g. for the expression "a car" in example 2, he would select the class *entity*. By clicking on the class, the annotation gets created and thus the text fragment will be displayed as an instance of the selected class in the ontology browser. The relationships between the created instances can be specified, e.g. the entity *The car* can be annotated as coreferring with the preceding entity *a car* as described in section 2. For this purpose, when selecting a certain class instance as well as a corresponding semantic relation from the ontology, OntoMat already presents the possible target class instances according to the range restrictions of the chosen relation. Hereby erroneous annotations of relations are avoided (compare section 2). Futhermore, literal attributes can be assigned to every created instance by typing them into the related attribute field. The choice of the predefined attributes depends on the class the instance belongs to. Thereby, instances of a certain concept can be annotated with grammatical information about how they are linguistically expressed, i.e. through an NP, a noun, a pronoun, a verb, etc. (compare section 3).

5 Discussion of Related Work

There is a vast amount of frameworks and tools developed for the purpose of linguistic annotation. However, in this paper we will focus on the discussion of frameworks for the annotation of anaphoric or discourse relations in written texts. In the annotation scheme proposed by (Müller and Strube, 2001) in the context of their annotation tool MMAX and in contrast to the one proposed in this paper, anaphoric relations are restricted to coreferring expressions, while bridging relations are restricted to non-coreferring ones. In line with (Krahmer and Piwek, 2000) and (van Deemter and Kibble, 2000) this is in our view a too strict definition of anaphora so that we propose a more relation-based classification of anaphoric and bridging relations. Furthermore, in (Müller and Strube, 2001), anaphoric relations are further differentiated according to the lexical items taking part in the relation. We have shown that under the assumption that the corresponding grammatical information is provided by the annotators, such a classification can be seen as a byproduct of a more semantic one such as outlined in this paper. In addition, (Müller and Strube, 2001) propose to specify antecedence with regard to equivalence classes rather than with regard to particular antecedents. However, this has the disadvantage that the information about the actual antecedent an annotator has selected is actually lost. Thus in our annotation proposal the fact that the *Coreference* relation forms equivalence classes is modeled by an underlying axiom system which can be exploited in the evaluation of a system against the annotation standard.

The annotation scheme proposed by Poesio et al. (Poesio and Vieira, 1998) is a product of a corpus-based analysis of definite description (DD) use showing that more than 50% of the DDs in their corpus are *discourse new* or *unfamiliar*. Thus in Poesio et al.'s annotation scheme definite descriptions are also explicitly annotated as discourse new.

The MUC coreference scheme (Hirschman and Chinchor, 1997) is restricted to the annotation of coreference relations, where coreference is also defined as an equivalence relation. Though this annotation scheme may seem quite simple, we agree with (Hirschman and Chinchor, 1997) that it is complex enough when taking into account the agreement of the annotators on a task. In fact, it has been shown that the agreement of subjects annotating bridging (Poesio and Vieira, 1998) or discourse (Cimiano, 2003) relations can be too low for tentative conclusion to be drawn (Carletta, 1996). The motivation of the MUC coreference scheme was thus to develop an annotation scheme leading to a good agreement. On the other hand, our motivation is to show how our ontology-based framework can be applied to the annotation of anaphoric relations in written texts and from this perspective the MUC coreference annotation scheme would have been in fact too restricted to actually show all the advantages of our approach.

The UCREL (Fligelstone, 1992) and DRAMA (Passoneau, 1996) annotation schemes are more related to ours than the schemes above in the sense that they also provide a rich set of particular bridging relations that can be annotated. However, in contrast to the ontology-based framework presented in this paper, these bridging relations are not constrained with regard to the conceptual types of their arguments, so that erroneous annotations can not be avoided.

The coreference annotation scheme proposed within the MATE Workbench project consists of a core as well as an extended scheme (Davies et al., 1998). The core scheme is in principle identical with the MUC coreference scheme and is restricted to the annotation of coreference in the sense of (van Deemter and Kibble, 2000). The extended scheme also allows the annotation of bound anaphors, of the relationship between a function and its values, of different set, part and possession relations, of instantiation relations as well as of event relations. The MATE scheme is related to our ontology-based annotation scheme in the sense that relations are also annotated as triples via the *link*-tag (Davies et al., 1998). As in our framework, the MATE scheme also allows to mark up ambiguities of reference. However, in contrast to the MATE scheme our framework has no means to specify a preference order on these ambiguous antecedents. On the other hand, the MATE scheme also includes a reasonable and complete taxonomy of markables as well as some features relevant for the annotation of coreference in dialogues such as the treatment of hesitations, disfluencies and misunderstandings.

6 Conclusion

We have argued that many linguistic annotation tasks can be seen as a special case of semantic annotation with regard to an ontology and have proposed a novel ontology-based framework for this purpose. We have furthermore applied our framework to the annotation of anaphoric relations in written texts. For this purpose we have proposed a relatively complex annotation scheme for anaphoric relations in which we have deliberatively abstracted from important issues such as inter-annotator agreement. In fact, the main contribution of this paper is certainly not the annotation scheme proposed in section 2, but to show that relatively complex annotation schemes such as the one proposed can be modeled in our ontology-based framework in a straightforward manner. The main benefits of the approach presented here are that the annotation can be performed at different levels of detail with regard to a given taxonomy as well as that the possible relations between two different concepts are constrained by the underlying ontology, which could make the annotation less error-prone. Furthermore, we have shown how the modeling of axioms within the ontology can actually make the evaluation of a system more straightforward. The most important advantage is that by specifying the annotation scheme in form of an ontology and adhering to standards such as RDF or OWL, it can be easily exchanged between different parties and can also be developed independently of the annotation tool used, which meets the interoperability requirement mentioned in (Ide, 2002). In addition, our framework is flexible enough to be applied to various annotation tasks, which is also a requirement mentioned in (Ide, 2002). In the future, we hope to show that, with the necessary extensions, our model is also suitable for the annotation of multi-modal corpora as well as of speech signals.

References

N. Asher and A. Lascarides. 1999. Bridging. *Journal of Semantics*, 15:83–113.

E. Bozsak, M. Ehrig, S. Handschuh, A. Hotho, A. Mädche, B. Motik, D. Oberle, C. Schmitz, S. Staab, L. Stojanovic, N. Stojanovic, R. Studer, G. Stumme, Y. Sure, J. Tane, R. Volz, and V. Zacharias. 2002. Kaon - towards a large scale semantic web. In *Proceedings of the Third International Conference on E-Commerce and Web Technologies (EC-Web 2002)*. Springer Lecture Notes in Computer Science.

J. Carletta. 1996. Asessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.

P. Cimiano. 2003. Ontology-driven discourse analysis in GenIE. In *Proceedings of the 8th International Conference on Applications of Natural Language to Information Systems*.

F. Ciravegna. 2001. Adaptive information extraction from text by rule induction and generalization. In *Proceedings of tht 17th International Joint Conference on Artificial Intelligence (IJCAI 2001)*.

S. Davies, M. Poesio, F. Bruneseaux, and L. Romary, 1998. *Annotating Coreference in Dialogues: Proposal for a scheme for MATE*. http://www.cogsci.ed.ac.uk/p̃oesio/anno_manual.html.

S. Decker, M. Erdmann, D. Fensel, and R. Studer. 1999. Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information. In R. Meersman et al., editors, *Database Semantics: Semantic Issues in Multimedia Systems*, pages 351–369. Kluwer.

S. DeRose, E. Maler, and R. Daniel. 2001. XML Pointer Language (XPointer). Technical report, W3C. Working Draft 16 August 2002.

S. Fligelstone. 1992. Developing a scheme for annotating text to show anaphoric relations. In G. Leitner, editor, *New Directions in Corpus Linguistics*, pages 153–170. Mouton de Gruyter.

T.R. Gruber. 1993. Toward principles for the design of ontologies used for knowledge sharing. In *Formal Analysis in Conceptual Analysis and Knowledge Representation*. Kluwer.

U. Hahn and K. Schnattinger. 1998. Ontology engineering via text understanding. In *Proceedings of the 15th World Computer Congress 'The Global Information Society on the Way to the Next Millenium' (IFIP'98)*.

S. Handschuh, S. Staab, and A. Mädche. 2001. CREAM — Creating relational metadata with a component-based, ontology-driven annotation framework. In *Proceedings of K-Cap 2001*, pages 76–83. ACM Press.

L. Hirschman and N. Chinchor. 1997. MUC-7 coreference task definition. In *Proceedings of the 7th Message Understanding Conference (MUC-7)*.

Nancy Ide. 2002. Linguistic annotation framework. Technical Report ISO TC37/SC4/WG1 N11, ISO TC37 SC4 Language Resource Management, 08.

M. Kifer, G. Lausen, and J. Wu. 1995. Logical foundations of object-oriented and frame-based languages. *Journal of the ACM*, 42:741–843.

E. Krahmer and P. Piwek. 2000. Varieties of anaphora: Introduction. In *ESSLLI 2000 Reader*.

A. Lascarides and N. Asher. 1991. Discourse relations and defeasible knowledge. In *Meeting of the Association for Computational Linguistics*, pages 55–62.

O. Lassila and R. Swick. 1999. Resource description framework (RDF) model and syntax specification. Technical report, W3C. W3C Recommendation. http://www.w3.org/TR/REC-rdf-syntax.

A. Mädche, V. Pekar, and S. Staab. 2002. Ontology learning part one - on discovering taxonomic relations from the web. In *Web Intelligence*. Springer.

M. Marcus, B. Santorini, and M.A. Marcinkiewicz. 1993. Building a large annotated corpus of english: The penn treebank. *Computational Linguistics*, 19(2):313–330.

C. Müller and M. Strube. 2001. Annotating anaphoric and bridging relations with MMAX. In *Proceedings of the 2nd SIGdial Workshop on Discourse and Dialogue*, pages 90–95.

R. Passoneau, 1996. *DRAMA: Instructions for Applying Reference Annotation for Multiple Applications*.

M. Poesio and U. Reyle. 2001. Underspecification in anaphoric reference. In *Proceedings of the IWCS-4*.

M. Poesio and R. Vieira. 1998. A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2):183–216.

Philip Resnik. 1997. Selectional preference and sense disambiguation. In *Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How?*

S. Staab and A. Mädche. 2000. Ontology engineering beyond the modeling of concepts and relations. In R. V. Benjamins, A. Gomez-Perez, N. Guarino, and M. Uschold, editors, *Proceedings of the 14th Conference on Artificial Intelligence, Workshop on Applications and Problem-Solving Methods*.

Kees van Deemter and Rodger Kibble. 2000. On coreferring: Coreference in MUC and related annotation schemes. *Computational Linguistics*, 26(4).