Dependency and relational structure in treebank annotation

Cristina BOSCO, Vincenzo LOMBARDO
Dipartimento di Informatica, Università di Torino
Corso Svizzera 185
10149 Torino,
Italia,
bosco,vincenzo@di.unito.it

Abstract

Among the variety of proposals currently making the dependency perspective on grammar more concrete, there are several treebanks whose annotation exploits some form of Relational Structure that we can consider a generalization of the fundamental idea of dependency at various degrees and with reference to different types of linguistic knowledge.

The paper describes the Relational Structure as the common underlying representation of treebanks which is motivated by both theoretical and task-dependent considerations. Then it presents a system for the annotation of the Relational Structure in treebanks, called Augmented Relational Structure, which allows for a systematic annotation of various components of linguistic knowledge crucial in several tasks. Finally, it shows a dependency-based annotation for an Italian treebank, i.e. the Turin University Treebank, that implements the Augmented Relational Structure.

1 Introduction

Different treebanks use different annotation schemes which make explicit two distinct but interrelated aspects of the structure of the sentence, i.e. the function of the syntactic units and their organization according to a part-whole paradigm. The first aspect refers to a form of *Relational Structure* (RS), the second refers to its *constituent* or *Phrase Structure* (PS). The major difference between the two structures is that the RS allows for several types of relations to link the syntactic units, whilst the PS involves a single relation "part-of". The RS can be seen as a generalization of the dependency syntax with the syntactic units instantiated to individual words in the dependency tree (Mel'čuk, 1988). As described in many theoretical linguistic frameworks, the RS provides a useful interface between syntax and a semantic or conceptual representation of predicate-argument structure. For example, Lexical Functional Grammar (LFG) (Bresnan, 1982) collocates relations at the interface between lexicon and syntax, Relational Grammar (RG) (Perlmutter, 1983) provides a description of the sentence structure exclusively based on relations and syntactic units not structured beyond the string level.

This paper investigates how the notion of RS has been applied in the annotation of treebanks, in terms of syntactic units and types of relations, and presents a system for the definition of the RS that encompasses several uses in treebank schemata and can be viewed as a common underlying representation. The system, called Augmented Relational Structure (ARS) allows for an explicit representation of the three major components of linguistic structures, i.e. morpho-syntactic, functional-syntactic and semantic. Then the paper shows how a dependency-based annotation can descend on ARS, and describes the ARS-based annotation of a dependency treebank for Italian, the Turin University Treebank (TUT), which is the first available treebank for Italian, with a few quantitative results.

The paper is organized as follows. The next section investigates both the annotation of RS in treebanks and the major motivations for the use of RS from language-specific issues and NLP tasks implementation; then we present the ARS system; finally, we show the dependency annotation of the TUT corpus.

2 Annotation of the Relational Structure

In practice, all the existing treebank schemata implement some form of relational structure. Annotation schemata range from pure (dependency) RS-based approaches to RS-PS combinations (Abeillé, 2003).

Some treebanks consider the relational information as the exclusive basis of the annotation.

The Prague Dependency Treebank ((Hajičová and Ceplová, 2000), (Böhmová et al., 2003)) implements a three level annotation scheme where both the analytical (surface syntactic) and tectogrammatical level (deep syntactic and topic-focus articulation) are dependency-based; the English Dependency Treebank (Rambow et al., 2002) implements a dependency-based mono-stratal analysis which encompasses surface and deep syntax and directly represents the predicate-argument structure. Other projects adopt mixed formalisms where the sentence is split in syntactic subunits (phrases), but linked by functional or semantic relations, e.g. the Negra Treebank for German ((Brants et al., 2003), (Skut et al., 1998)), the Alpino Treebank for Dutch (van der Beek et al., 2002), and the Lingo Redwood Treebank for English (Oepen et al., 2002). Also in the Penn Treebank ((Marcus et al., 1993), (Marcus et al., 1994)) a limited set of relations is placed over the constituency-based annotation in order to make explicit the (morpho-syntactic or semantic) roles that the constituents play.

The choice of a RS-based annotation schema can depend on theoretical linguistic motivations (a RS-based schema allows for an explicit, fine-grained representation of several linguistic phenomena), task-dependent motivations (the RS-based schema represents the linguistic information involved in the task(s) at hand), language-dependent motivations (the relational structure is traditionally considered as the most adequate representation of the object language).

Theoretical motivations for exploiting representations based on forms of RS was developed in the several RS-based theoretical linguistic frameworks (e.g. Lexical Functional Grammar, Relaional Grammar and dependency grammar), which allow for capturing information involved at various level (e.g. syntactic and semantic) in linguistic structures, and grammatical formalisms have been proposed with the aim to capture the linguistic knowledge represented in these frameworks. Since the most immediate way to build wide-coverage grammars is to extract them directly from linguistic data (i.e. from treebanks), the type of annotation used in the data is a factor of primary importance, i.e. a RS-based annotation allows for the extraction of a more descriptive grammar[1].

Task-dependent motivations rely on how the annotation of the RS can facilitate some processing aspects of NLP applications. The explicit representation of predicative structures allowed by the RS can be a powerful source of disambiguation. In fact, a large amount of ambiguity (such as coordination, Noun-Noun compounds and relative clause attachment) can be resolved using such a kind of information, and relations can provide a useful interface between syntax and semantics. (Hindle and Rooth, 1991) had shown the use of dependency in Prepositional Phrase disambiguation, and the experimental results reported in (Hockenmaier, 2003) demonstrate that a language model which encodes a rich notion of predicate argument structure (e.g. including long-range relations arising through coordination) can significantly improve the parsing performances. Moreover, the notion of predicate argument structure has been advocated as useful in a number of different large-scale language-processing tasks, and the RS is a convenient intermediate representation in several applications (see (Bosco, 2004) for a survey on this topic). For instance, in Information Extraction relations allows for recognizing different guises in which an event can appear regardless of the several different syntactic patterns that can be used to specify it (Palmer et al., 2001)[2]. In Question Answering, systems usually use forms of relation-based structured representations of the input texts (i.e. questions and answers) and try to match those representations (see e.g. (Litkowski, 1999), (Buchholz, 2002)). Also the in-depth understanding of the text, necessary in Machine Translation task, requires the use of relation-based representations where an accurate predicate argument structure is a critical factor (Han et al., 2000)[3].

Language-dependent motivations rely on the fact that the dependency-based formalisms has been traditionally considered as the most adequate for the representation of free word order languages. With respect to constituency-based

---

[1]See (Mazzei and Lombardo, 2004a) and (Mazzei and Lombardo, 2004b) for experiments of LTAG extraction from TUT.

[2]Various approaches to IE (Collins and Miller, 1997) address this issue by using relational representations, that is forms of "concept nodes" which specifies a trigger word (usually a Verb) and also forms of mapping between the syntactic and the semantic relations of the trigger.

[3]The system presented in (Han et al., 2000) generates the dependency trees of the source language (Korean) sentences, then directly maps them to the translated (English) sentences.

formalisms, free word order languages involves a large amount of discontinuous constituents (i.e. constituents whose parts are not contiguous in the linear order of the sentence). In practice, a constituency-based representation was adopted for languages with rather fixed word order patterns, like English (Penn Treebank), while a dependency representation for languages which allow variable degrees of word order freedom, such as Czech (see Prague Dependency Treebank) or Italian (as we will see later, TUT). Nevertheless, in principle, since the representation of a discontinuous constituent X can be addressed in various ways (e.g. by introducing lexically empty elements co-indexed with the moved parts of X), the presence to a certain extent of word order freedom does not necessarily mean that a language has to be necessarily annotated according to a relation-based format rather than a constituency-based one. Moreover, free word order languages can present difficulties for dependency-based as well as for constituency-based frameworks (e.g. non-projective structures). The development of dependency-based treebanks for English (see English Dependency Treebank) together with the inclusion of relations in constituency-based treebanks (see Penn Treebank) too, confirms the strongly prevailing relevance of motivations beyond the language-dependent ones.

The types of knowledge that many applications actually need are RS-based representations where predicate argument structure and the associated morphological and syntactic information can operate as an interface to a semantic-conceptual representation. All these types of knowledge have in common the fact that they can be described according to the dependency paradigm, rather than according to the constituency paradigm. The many applications (in particular those referring to the Penn Treebank) which use heuristics-based translation schemes from the phrase structure to lexical dependency ("head percolation tables") (Rambow et al., 2002) show that the access to comprehensive and accurate extended dependency-based representations has to be currently considered as a critical issue for the development of robust and accurate NLP technologies.
Now we define our proposal for the representation of the RS in treebank annotation.

3 The Augmented Relational Structure

A RS consists of syntactic units linked by relations. An Augmented Relational Structure (ARS) organizes and systematizes the information usually associated in existing annotations to the RS, and includes not only syntactic , but also linguistic information that can be represented according to a dependency paradigm and that is proximate to semantics and underlies syntax and morphology. Therefore the
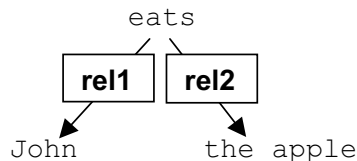

Figure 1: A simple RS.

ARS expresses the relations and syntactic units in terms of multiple components. We describe the ARS as a dag where each relation is a feature structure including three components. Each component of the ARS-relations is use-
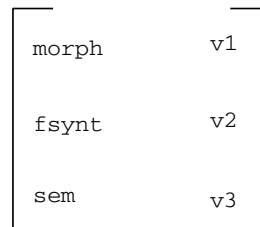

Figure 2: An ARS relation.

ful in marking both similarities and differences among the behavior of units linked by the dependency relations.
The morpho-syntactic component of the ARS describes morpho-syntactic features of the words involved in relations, such as their grammatical category. This component is useful for making explicit the morpho-syntactic variants of predicative structures. Instances of this kind of variants often occur in intransitive, transitive and di-transitive predicative structures, e.g. *esplosione-esplodere* (explosion - to explode) are respectively nominal and verbal variants of the intransitive structure "something explodes". By referring to the TUT, we can evaluate the frequency of this phenomenon: in 1,500 sentences 944 Verbs occur (for a total of 4169 occurrences) and around the 30% of them are

present in the nominal variant too[4].

The functional-syntactic component identifies the subcategorized elements, that is it keeps apart arguments and modifiers in the predicative structures. Moreover, this component makes explicit the similarity of a same predicative structure when it occurs in the sentence in different morpho-syntactic variants. In fact, the functional-syntactic components involved, e.g., in the transitive predicative structure "someone declares something", are the same in both the nominal (*dichiarazione [di innocenza]$_{OBJ}$ [di John]$_{SUBJ}$* - John's declaration of innocence) and verbal realization (*[John]$_{SUBJ}$ dichiara [la sua innocenza]$_{OBJ}$* - John declares his innocence) of such a predication, i.e. SUBJ and OBJ. The distinction between arguments and modifiers has been considered as quite problematic in the practice of the annotation, even if relevant from the applicative and theoretical point of view[5], and is not systematically annotated in the Penn Treebank (only in clear cases, the use of semantic roles allows for the annotation of argument and modifier) and in the Negra Treebank. This distinction is instead usually marked in dependency representations, e.g. in the English Dependency Treebank and in the Prague Dependency Treebank.

The semantic component of the ARS-relations specifies the role of words in the syntax-semantics interface and discriminates among different kinds of modifiers and oblique complements. We can identify at least three levels of generality: verb-specific roles (e.g. Runner, Killer, Bearer); thematic roles (e.g. Agent, Instrument, Experiencer, Theme, Patient); generalized roles (e.g. Actor and Undergoer). The use of specific roles can cause the loss of useful generalizations, whilst too generic roles do not describe with accuracy the data. An example of annotation of semantic roles is the tectogrammatical layer of the Prague Dependency Treebank.

ARS features a mono-stratal approach. By following this strategy, the annotation process can be easier, and the result is a direct representation of a complete predicate argument structure, that is a RS where all the information (morpho-syntactic, functional-syntactic and semantic) are immediately available. An alternative approach has been followed by the Prague

Dependency Treebank, which is featured by a three levels annotation. This case shows that the major difference between the syntactic (analytic) and the semantic (tectogrammatical) layer consists in the inclusion of empty nodes for recovering forms of deletion ((Böhmova et al., 1999), (Hajičová and Ceplová, 2000)). But this kind of information does not necessarily requires a separated semantic layer and can be annotated as well in a mono-stratal representation, like the English Dependency Treebank (Rambow et al., 2002) does.

The tripartite structure of the relations in ARS guarantees that different components can be accessed separately and analyzed independently (like in (Montemagni et al., 2003) or in (Rambow et al., 2002)). Furthermore, the ARS allows for forms of annotation of relations where not all the features are specified too. In fact, the ARS-relations which specify only a part of components allow for the description of syntactic grammatical relations which do not correspond with any semantic relation, either because they have a void semantic content or because they have a different structure from any possible corresponding semantic relation (i.e. there is no semantic relation linking the same ARS-units linked by the syntactic one). Typical relations void of semantic content can link the parts of expressions not compositionally interpretable (idioms), for instance *together* with *with* in *together with*. While a classic example of a nonisomorphic syntactic and semantic structure is one which involves the meaning of quantifiers: a determiner within a NP extends its scope beyond the semantic concept that results from the interpretation of the NP. Another example is the coordination where the semantic and syntactic structure are often considered as non isomorphic in several forms of representation.

The ARS-relations including values for both functional-syntactic and semantic components may be used in the representation of grammatical relations which participate into argument structures and the so-called oblique cases (see Fillmore and (Hudson, 1990)), i.e. where the semantic structures are completely isomorphic to the syntactic structures. For example, a locative adjunct like *in the garden* in *John was eating fish in the garden* is represented at the syntactic level as a Prepositional Phrase playing the syntactic function locative in the Verb Phrase (in the Penn Treebank it could be annotated as a PP-LOC); the semantic concept

---

[4]This statistics does not take into consideration the possible polysemic nature of words involved.

[5]See, for instance, in LFG and RG.

corresponding to *the garden* plays the semantic role LOCATION in the "eating" event stated by the sentence.

4 TUT: a dependency-based treebank for Italian

The TUT is the first available treebank of Italian (freely downloadable at http://www.di.unito.it/~tutreeb/). The current release of TUT includes 1,500 sentences corresponding to 38,653 tokens (33,868 words and 4,785 punctuation marks). The average sentence length is of 22,57 words and 3,2 punctuation marks.

In this section, we concentrate on the major features of TUT annotation schema, i.e. how the ARS system can describe a dependency structure.

4.1 A dependency-based schema

In Italian the order of words is fixed in non verbal phrases, but verbal arguments and modifiers can be freely distributed without affecting the semantic interpretation of the sentence. A study on a set of sentences of TUT shows that the basic word order for Italian is Subject-Verb-Complement (SVC), as known in literature (Renzi, 1988), (Stock, 1989), but in more than a quarter of declarative sentences it is violated (see the following table[6]). Although the

Permutations	Occurrences
S V C	74,26%
V C S	11,38%
S C V	7,98%
C S V	3,23%
V S C	2,29%
C V S	0,77%

Table 1: Italian word order

SVC order is well over the others, the fact that all the other orders are represented quantitatively confirms the assumption of partial configurationality intuitively set in the literature. The partial configurationality of Italian can be considered as a language-dependent motivation for the choice of a dependency-based annotation for an Italian treebank. The schema is similar to that of the Prague Dependency Treebank analytical-syntactic level with which TUT

shares the following basic design principles typical of the dependency paradigm:

- the sentence is represented by a tree where each node represents a word and each edge represents a dependency labelled by a grammatical relation which involves a *head* and a *dependent*,

- each single word and punctuation mark is represented by a single node, the so-called *amalgamated words*, which are words composed by lexical units that can occur both in compounds and alone, e.g. Verbs with clitic suffixes (*amarti* (to love-you) or Prepositions with Article (*dal* (from-the)), are split in more lexical units[7],

- since the constituent structure of the sentence is implicit in dependency trees, no phrases are annotated[8].

If the partial configurationality makes the dependency-based annotation more adequate for Italian, other features of this language should be well represented by exploiting a Negra-like format where the syntactic units are phrases rather than single words. For instance, in Italian, Nouns are in most cases introduced by the Article: the relation between Noun and Determiner is not very relevant in a dependency perspective, while it contributes to the definition of a clearer notion of NP in Italian than in languages poorer of Determiners like, e.g., Czech. The major motivation of a dependency-based schema is therefore theoretical and, in particular, to make explicit in the treebank annotation a variety of structures typical of the object language.

Moreover, in order to make explicit in cases of deletion and ellipsis the predicate argument structure, we annotate in the TUT null elements. These elements allow for the annotation of a variety of phenomena: from the "equi" deletion which affects the subject of infinitive Verb depending on a tensed Verb (e.g. *John(1) vuole T(1) andare a casa* - John(1) want to T(1) go home), to the various forms of gapping that can affect parts of the structure of the sentence (e.g. *John va(1) a casa e Mario T(1) al cinema* - John goes(1) home and Mario

---

[6]The data reported in the table refer to 1,200 annotated sentences where 1,092 verbal predicate argument structures involving Subject and at least one other Complement occur.

[7]Referring to the current TUT corpus, we see that around 7,7% words are amalgamated.

[8]If phrase structure is needed for a particular application, it is possible to automatically derive it from the dependency structure along with the surface word order.

```
                              dichiarava
                   ╱              │            ╲
         VERB,PREP        VERB,NOUN        VERB,DET+DEF
           RMOD             SUBJ               OBJ
           TIME            AGENT              THEME

            In              Sudja                il
              ╲                ╲                  ╲
      PREP,DET+DEF    NOUN,DET+DEF         DET+DEF,NOUN
          ARG          APPOSITION              ARG
           #             DENOM                  #

           quei             la              fallimento
             ╲                ╲
      DET+DEF,NOUN    DET+DEF,NOUN
          ARG             ARG
           #               #

          giorni          zingara
```
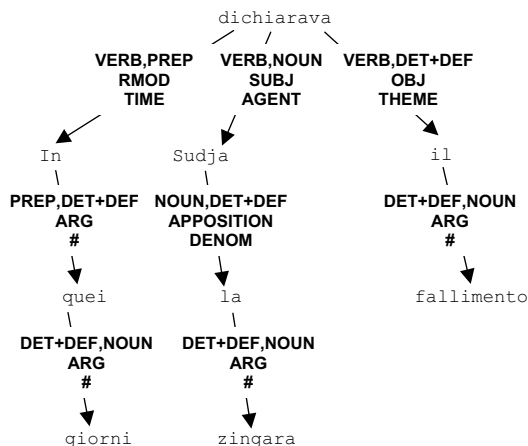
Figure 3: The TUT representation of *In quei giorni Sudja la zingara dichiarava il fallimento* (In those days Sudja the gipsy declared the bankruptcy).

T(1) to the cinema), to the pro-dropped subjects typical of Italian (as well as of Portuguese and Spanish), i.e. the subject of tensed Verbs which are not lexically realized in the sentence (e.g. *T Va a casa* - T goes home). For phenomena such as equi and gapping TUT implements co-indexed traces, while it implements non co-indexed traces for phenomena such as the pro-drop subject.

4.2 An ARS-based schema

In TUT the dependency relations form the skeleton of the trees and the ARS tripartite feature structures which are associated to these relations resolve the interface between the morphology, syntax and semantics. The ARS allows for some form of annotation also of relations where only parts of the features are specified. In TUT this has been useful for under-specifying relations both in automatic analysis of the treebank (i.e. we can automatically exclude the analysis of a specific component of the relations) and in the annotation process (i.e. when the annotator is not confident of a specific component of a relation, he/she can leave such a component void).

In the figure 3 we see a TUT tree.

All the relations annotated in the tree include the morpho-syntactic component, formed by the morphological categories of the words involved in the relation separated by a comma, e.g. VERB,PREP for the relation linking the root of the sentence with its leftmost child

(*In*). Some relation involves a morpho-syntactic component where morphological categories are composed by more elements, e.g. DET+DEF (in DET+DEF,NOUN) for the relation linking *quei* with *giorni*. The elements of the morpho-syntactic component of TUT includes, in fact, 10 "primary" tags that represent morphological categories of words (e.g. DET for Determiner, NOUN for Noun, and VERB for Verb), and that can be augmented with 20 "secondary" tags (specific of the primary tags) which further describe them by showing specific features, e.g. DEF which specifies the definiteness of the Determiner or INF which specifies infiniteness of Verb. Valid values of the elements involved in TUT morpho-syntactic tags are 40.

By using the values of the functional-syntactic component, TUT distinguishes among a variety of dependency relations. In figure 3 we see the distinction between argument, e.g. the relation SUBJ linking the argument *Sudja* with the verbal root of the sentence *dichiarava*, and the relation RMOD which represents a restrictive modifier and links the verbal root *dichiarava* with *in quei giorni*. The dependents of Prepositions and determiners are annotated as argument too, according to arguments presented in (Hudson, 1990). Another distinction which is exploited in the annotation of the sentence is that between restrictive modifier (i.e. RMOD which links *dichiarava* with *in quei giorni*) and APPOSITION (i.e. non restrictive modifier linking *Sudja* with *la zingara*), which are modifiers that restrict the meaning of the head. Beyond these basic distinctions, TUT schema draws other distinctions among the functional-syntactic relations and includes a large set of tags for a total of 55 items, which are compounds of 27 primary and 23 secondary tags. These tags are organized in a hierarchy (Bosco, 2004) according to their different degree of specification. In the hierarchy of relations, Arguments (ARG) include Subject (SUBJ), Object (OBJ), Indirect Object (INDOBJ), Indirect Complement (INDCOMPL), Predicative Complements (of the Subject (PREDCOMPL+SUBJ) and of the Object (PREDCOMPL+OBJ)). The direct consequence of its hierarchical organization is the availability of another mechanisms of under-specification in the annotation or in the analysis of annotated data. In fact, by referring to the hierarchy we can both annotate and analyze relations at various degrees of specificity.

The semantic component discriminates among

different kinds of modifiers and oblique complements. The approach pursued in TUT has been to annotate very specific semantic roles only when they are immediately and neatly distinguishable. For instance, by referring to the prepositional dependents introduced by da^9 (from/by), we find the following six different values for the semantic component:

- REASONCAUSE, e.g., in *gli investitori sono impazziti dalle prospettive di guadagno* (the investors are crazy because of the perspectives of gain)
- SOURCE, e.g., in *traggono benefici dalla bonifica* ([they] gain benefit from the drainage)
- AGENT, e.g., *l'iniziativa è appoggiata dagli USA* (the venture is supported by USA)
- TIME, e.g., *dal '91 sono stati stanziati 450 miliardi* (since '91 has been allocated 450 billions)
- THEME, e.g., *ciò distingue l'Albania dallo Zaire* (that distinguishes the Albany from Zaire)
- LOC, which can be further specialized in LOC+FROM, e.g., in *da qui è partito l'assalto* (from here started the attack), LOC+IN, e.g., in *quello che succedeva dall'altra parte del mondo* (what happened in the other side of the world), LOC+METAPH, e.g., in *l'Albania è passata dal lancio dei sassi alle mitragliatrici* (the Albany has passed from the stone-throwing to the machineguns).

In figure 3 the semantic component has been annotated only in four relations, which respectively represent the temporal modifier *In quei giorni* of the main Verb *dichiarava*, the apposition *la zingara* of the Noun *Sudja*, and the arguments of the Verb, i.e. the subject *Sudja la zingara* which plays the semantic role AGENT and the object *il fallimento* which plays the semantic role THEME. In the other relations involved in this sentence a value for the semantic component cannot be identified[10], e.g. the argument of a Preposition or Determiner cannot be semantically specified as in the case of the verbal arguments.

5 Conclusions

The paper analyzes the annotation of the RS in the existing treebanks by referring to a notion of RS which is a generalization of dependency.

According to this notion, the RS includes types of linguistic knowledge which are different, but which have in common that they can be represented by a dependency paradigm rather than to a constituency-based one.

The paper identifies two major motivations for exploiting RS in treebank annotation: language-dependent motivations that have determined the use of dependency for the representation of treebanks of free word order languages, and task-dependent motivations that have determined a wider use of relations in treebanks.

In the second part of the paper, we show a system for the annotation of RS, i.e. the ARS, and how the ARS can be used for the annotation of a dependency-based treebank, the TUT whose schema augments classical dependency (functional-syntactic) relations with morphological and semantic knowledge according to the above mentioned notion of RS.

References

A. Abeillé, editor. 2003. *Building and using syntactically annotated corpora*. Kluwer, Dordrecht.

A. Böhmova, J. Panevová, and P. Sgall. 1999. Syntactic tagging: procedure for the transition from analytic to the tectogrammatical treestructures. In *Proc. of 2nd Workshop on Text, speech and dialog*, pages 34–38.

A. Böhmová, J. Hajič, E. Hajičová, and B. Hladká. 2003. The Prague Dependency Treebank: A three level annotation scenario. In Abeillé (Abeillé, 2003), pages 103–127.

C. Bosco. 2004. *A grammatical relation system for treebank annotation*. Ph.D. thesis, University of Torino. http://www.di.unito.it/~bosco/.

T. Brants, W. Skut, and H. Uszkoreit. 2003. Syntactic annotation of a German newspaper corpus. In Abeillé (Abeillé, 2003), pages 73–87.

J. Bresnan, editor. 1982. *The mental representation of grammatical relations*. MIT Press, Cambridge.

S. Buchholz. 2002. Using grammatical relations, answer frequencies and the World Wide Web for TREC Question Answering. In *Proc. of TREC 2001*, pages 502–509.

M. Collins and S. Miller. 1997. Semantic tagging using a probabilistic context free grammar. In *Proc. of 6th Workshop on Very Large Corpora*.

---

[9]In 1,500 TUT sentences we find 591 occurrences of this Preposition.

[10]In figure 3, we marked the semantic component of these cases with ♯.

E. Hajičová and M. Ceplová. 2000. Deletions and their reconstruction in tectogrammatical syntactic tagging of very large corpora. In *Porc. of COLING 2000*, pages 278–284.

C. Han, B. Lavoie, M. Palmer, O. Rambow, R. Kittredge, T. Korelsky, N. Kim, and M. Kim. 2000. Handling structural divergences and recovering dropped arguments in a Korean/English machine translation system. In *Proc. of AMTA 2000*, pages 40–54.

D. Hindle and M. Rooth. 1991. Structural ambiguity and lexical relations. In *Proc. of ACL 91*, pages 229–236.

J. Hockenmaier. 2003. Parsing with generative models of predicate-argument structure. In *Proc. of ACL 2003*.

R. Hudson. 1990. *English Word Grammar*. Basil Blackwell, Oxford and Cambridge.

K.C. Litkowski. 1999. Question-answering using semantic relation triples. In *Proc. of TREC-8*, pages 349–356.

M. Marcus, B. Santorini, and M.A. Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19:313–330.

M. Marcus, G. Kim, M.A. Marcinkiewicz, R. MacIntyre, A. Bies, M. Ferguson, K. Katz, and B. Schasberger. 1994. The Penn Treebank: Annotating predicate argument structure. In *Proc. of HLT-94*.

A. Mazzei and V. Lombardo. 2004a. Building a large grammar for Italian. In *Proc. of LREC 2004*, pages 51–54.

A. Mazzei and V. Lombardo. 2004b. A comparative analysis of extracted grammars. In *Proc. of ECAI 2004*.

I.A. Mel'čuk. 1988. *Dependency Syntax: theory and practice*. SUNY, Albany.

S. Montemagni, F. Barsotti, M. Battista, and N. Calzolari. 2003. Building the Italian syntactic-semantic treebank. In Abeillé (Abeillé, 2003), pages 189–210.

S. Oepen, K. Toutanova, S. Shieber, C.D. Manning, D. Flickinger, and T. Brants. 2002. The LinGO Redwoods treebank: motivation and prliminary applications. In *Proc. of COLING 2002*, pages 1253–1257.

M. Palmer, J. Rosenzweig, and S. Cotton. 2001. Automatic predicate argument analysis of the Penn Treebank. In *Proc. of HLT 2001*.

D.M. Perlmutter. 1983. *Studies in Relational Grammar 1*. University of Chicago Press.

O. Rambow, C. Creswell, R. Szekely, H. Taber, and M. Walker. 2002. A dependency treebank for English. In *Proc. of LREC 2002*, pages 857–863.

L. Renzi, editor. 1988. *Grande grammatica italiana di consultazione, vol. I*. Il Mulino, Bologna.

W. Skut, T. Brants, B. Krenn, and H. Uszkoreit. 1998. A linguistically interpreted corpus of German in newspaper texts. In *Proc. of LREC 98*, pages 705–713.

O. Stock. 1989. Parsing with flexibility, dynamic strategies, and idioms in mind. *Computational Linguistics*, 15(1):1–17.

L. van der Beek, G. Bouma, R. Malouf, and G. van der Noord. 2002. The Alpino dependency treebank. In *Proc. of CLIN 2001*.