

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

A Relation-based Schema for Treebank Annotation

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/6700> since 2018-03-01T23:23:03Z

Publisher:

Springer-Verlag

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

A relation-based schema for treebank annotation

Cristina Bosco and Vincenzo Lombardo

Dipartimento di Informatica, Università di Torino,
Corso Svizzera 185, 10149 Torino, Italy
{bosco, vincenzo}@di.unito.it
<http://www.di.unito.it/~tutreeb>

Abstract. This paper presents a relation-based schema for treebank annotation, and its application in the development of a corpus of Italian sentences. The annotation schema keeps arguments and modifiers distinct and allows for an accurate representation of predicate-argument structure and subcategorization. The accuracy strongly depends on methods adopted for defining the relations which are tripartite feature structures that consist of a morpho-syntactic, a functional and a semantic component. We presents empirical evidence for these tripartite structures by illustrating phenomena faced in the development of an Italian treebank.

1 Introduction

Currently, most NLP systems that operate in realistic settings use statistical methods trained on *treebanks*, very large corpora of syntactically annotated sentences. The major existing treebanks are the Penn Treebank for English ([18], [19], [1]), the Prague Dependency Treebank for Czech ([12], [1]), and the NEGRA-TIGER project for German ([4], [1], [3], [10]). Moreover, there is a worldwide effort for the development of such a resource (see [1] for reports about Spanish, Italian, French, Chinese, Polish, Japanese and Turkish).

Treebanks can be viewed as text repositories where the implicit linguistic information is made explicit through the process of annotation. Linguistic information consists of Part Of Speech tagging, including syntactic categories and possibly other features (gender, number, tense, ...), bracketing structure, and *syntactic dependencies*. Syntactic dependencies are a relevant part of the annotation, since the level of representation that many applications actually need is a representation of predicate-argument structure which provides a useful interface to a semantic or conceptual representation [23].

The relevance of the predicate-argument structure has been clearly demonstrated by [14], referring to the Machine Translation task, where the contribution of the predicate-argument structure to the quality of the translation output exceeds the contribution of syntactic structure and vocabulary coverage (this has been further discussed in [15]). The syntax-semantics interface has been useful in Question Answering approaches that match the semantic roles in the question with the semantic roles of the (documents which can contain the) answer (see, e.g. the TREC-8 system described in [16]). The use of semantic roles too provides a great contribution to solving a number of phrase attachment ambiguities

[6]. Considerable progress has been made in parsing by systems based on lexicalized probabilistic models that take into account syntactic dependencies (e.g. [9], [8]). Also, syntactic dependencies have been used successfully in NLP-based Information Extraction [27] and Information Retrieval [28].

All major approaches to treebank annotation include some forms of syntactic dependencies (see section 2), and the research for the enhancement of the annotation formats has led some projects to introduce further annotation levels, in propositional and pragmatic terms. However, there is not an agreement among the various proposals on the number and type of dependencies that are to be annotated in order to yield effective treebank-trained applications (see [13] for a comparison among several annotations of syntax-semantics information). Each approach comes with its own set of dependencies that is mostly guided by the actual data in the corpus at hand. Moreover, all the approaches miss to represent explicitly the interrelation among syntactic dependencies which is fundamental for the extraction of accurate linguistic knowledge and for the design of statistical models.

The goal of this paper is to systematize the representation of syntactic

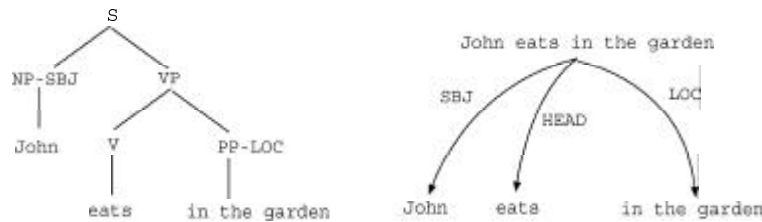


Fig. 1. An example in Penn-like format and a possible representation of the relational structure (*SBJ*=subject, *LOC*=location)

dependencies by providing a single-layered reference framework for treebank annotation. The reference framework relies on the notions of *grammatical relation* and *relational structure*, that have a long tradition in linguistics and are strongly related to the predicate-argument structure and the syntax-semantics interface (see figure 1 on the right).

These relational notions encode linguistic knowledge that is more proximate to semantics, and underlies syntax and morphology. Constraints on relations are expressed through subcategorization and valency, that inform about the lexical realization of a predicate, the number and type of arguments that a predicate requires, and the mapping from these syntactic arguments to a semantic representation [6]. The importance of the relational structure is testified by a number of formalisms: Relational Grammar [22] presents a relation-based sentence structure; Lexical Functional Grammar [5] introduces a relational structure, the f-structure, that is pivoted by the constituency structure; Categorical Grammar [17], forms a family of formalisms which represent the language in terms of func-

tions and arguments encoding the syntax-semantics interface.

This paper proposes a single-layered relation-based schema for the representation of syntactic dependencies in treebank annotation. The focus of the paper is the representation of grammatical relations as tripartite feature structures that consist of a morpho-syntactic, a functional, and a semantic component. The schema is applied to the annotation of a corpus of Italian sentences, the Turin University Treebank (TUT), and we present evidence for the tripartite feature structure from the corpus sentences.

The paper is organized as follows: after a review of the representation of the syntactic dependencies in the three major approaches to treebank annotation, we motivate and introduce the tripartite structure of grammatical relations, the relational structure and the TUT annotation schema. Finally we provide empirical evidence for the tripartite feature structure from the TUT corpus.

2 Related approaches

In this section we review the approaches to the annotation of syntactic dependencies in three known schemata, that have been successfully applied to the annotation of very large corpora.

The constituency-based Penn schema [18] is augmented with a limited number of “function” tags associated with constituents (see figure 1 on the left), and organized in three sets [19]: form/function tags, grammatical tags, and semantic role tags. Form/function tags assume a default morpho-syntactic function associated with each constituent type (form), and are used to mark those constituent labels for which the actual function is not the default function. For example, ADV (= adverbial) is a function tag that marks clauses and NPs that behave like adverbs, NOM (= nominal) marks non-NP that functions as NP. Grammatical tags refer to the relational structure in terms of syntactic functions: typical examples are SBJ (surface subject), LGS (logical subject in passives), DTV (dative object in unshifted dative constructions). Finally, semantic role tags are BNF (benefactive), DIR (direction), LOC (location), MNR (manner), Each constituent can be tagged with multiple tags, but never with two function tags from the same set.

The goal of a more accurate annotation of the predicate-argument structure has triggered a novel project, known as PropBank [21] [15]. The idea is to develop a new treebank of propositional structures associated with sentences.

Negra combines the annotation of constituents and relations with a mixed representation of both the bracketing structure and the syntactic dependencies, possibly spanning long distances. The bracketing structure consists of words and phrases; in order to conveniently represent German free word order without introducing a large number of traces, it allows discontinuous constituents. Syntactic dependencies between the bracketing units are represented by special nodes: HD (head) is a specific label that marks the head word of a phrase, several syntactic functions (DA = dative, JU = junctor, MNR = post-nominal modifier, MO = modifier, OC = clausal object, PD = predicative, SB = subject, SBP =

subject in passives, ...).

A specific enhancement with respect to Penn annotation is some distinction between arguments and modifiers and the identification of the head of a phrase. However, Negra annotation schema lacks semantic roles, which are the object of the TIGER project [3], [26].

The annotation schema of Prague descends directly from the functional generative approach to dependency syntax [24]. The schema consists of three separate levels, morphological, analytical (i.e. functional-syntactic) and tectogrammatical (i.e. semantic/pragmatic). The morphological level consists in information concerning a single word. The analytical level connects the sentence words (phrases are banned in the dependency approach) with grammatical relations (called analytical functions - e.g., Pred (predicate), Sb (subject), Obj (Object), Adv (adverbial), Atv (complement depending on non-Verb), AtvV (Verbal complement), Atr (attribute), AuxC (conjunction), AuxV (auxiliary "to be"), AuxK (punctuation) Coord (coordination node), ...). The tectogrammatical level is structurally based on dependency syntax, and represents the syntax-semantics/pragmatics interface, in order to reveal the topic-focus articulation of the sentence. Here there are about 40 functors that represent semantic roles (actor/bearer, addressee, benefactive, origin, effect, cause, manner, locative, ...).

Abstracting from the three approaches reviewed above, we can individuate three major aspects in syntactic dependencies annotation: a morpho-syntactic aspect, a functional-syntactic aspect, and a semantic-syntactic aspect. These three aspects are conveyed by the notion of grammatical relation, a feature structure where the interrelations among the various dependencies can be conveniently represented.

3 A relation-based schema for treebank annotation

In this section we introduce our relation-based approach to treebank annotation. Relations are represented as feature structures [25], partitioned in three components, morpho-syntactic, functional-syntactic, and semantic-syntactic. The set of all the relations in a sentence is the Augmented Relational Structure (ARS), that augments standard relational structure (in functional terms) with information about the morpho-syntactic realization of the relation and the mapping to a semantic role. The Augmented Relational Structure (ARS) is a directed acyclic graph (dag), where nodes are syntactic units of the sentence and edges are grammatical relations between the syntactic units. Syntactic units are strings of words of the sentence, possibly up to individual words. Now we introduce the three components, and then the ARS-based annotation schema.

3.1 The morpho-syntactic component

The morpho-syntactic component is useful in making explicit morpho-syntactic variants in the realization of one predicate-argument structure, that is expressions that differ for the morpho-syntactic features of at least one word (e.g.,

correre velocemente (to run quickly) is a morpho-syntactic variant of *correre in modo veloce* (to run in a fast way)). In marking the morpho-syntactic variants, we can account for the influences between the surface realizations of the several portions of a predicate-argument structure: this provides hints for the recognition of the same predicate-argument through several surface realizations [15]. In the

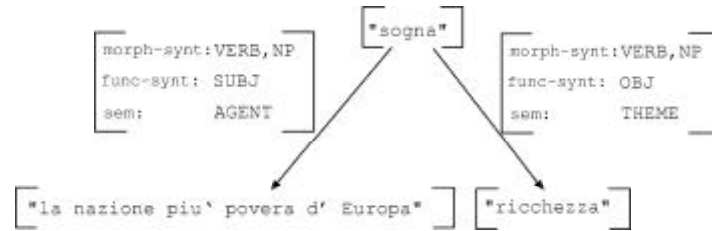


Fig. 2. A verbal realization of a predicate-argument structure: "La nazione più povera d'Europa sogna ricchezza" (The poorest nation of Europe dreams wealth)

figures 2 and 3 we find two morpho-syntactic variants of the predicate-argument structure "someone dreams of something". The fact that they only differ with respect to the morph-synt component immediately reveals they are two variants of the same predicative structure: when the head is verbal, the subject is a NP which agrees in number with the verb (**la nazione*[SING] *più povera d'Europa sognano*[PLUR]); when the head is nominal, the subject is realized by a PP introduced by "di" and agreement is not enforced (*i sogni*[PLUR] *di ricchezza della nazione*[SING] *più povera d'Europa*).

The morpho-syntactic component is also useful in cases when different morpho-



Fig. 3. A nominal realization of a predicate-argument structure: "I sogni di ricchezza della nazione più povera d'Europa" (The dreams of wealth of the poorest nation of Europe)

syntactic categories fulfil the same syntactic function within a large syntactic unit, though not part of a predicate-argument structure. For instance, both Noun Phrases and Prepositional Phrase (beyond Adverbs) can function as typical Adverbial Modifiers within a sentence, like in the sentence "Pat partì presto/quel

giorno/in quel giorno” (Pat departed early/that day/in that day). So, NP and PP are two morph-synt variants of Adverbials.

3.2 The functional-syntactic component

The functional-syntactic component is the core of the grammatical relations, and refers to those relations (like Subject and Object) that have been considered as purely syntactic in the literature. For example, in Relational Grammar [22] and Lexical Functional Grammar [5] there are three arguments syntactically marked, i.e. Subject, Object, Indirect Object. The other arguments are often called ‘indirect complements’ and distinguished on a semantic basis.

While the morpho-syntactic component marks the differences among morpho-syntactic variants of one predicative structure, the functional-syntactic component is useful in marking the similarity among them. This component identifies the subcategorized elements, that is it keeps apart arguments and modifiers in the predicative structures. For instance, the structures in figure 4 are differenti-

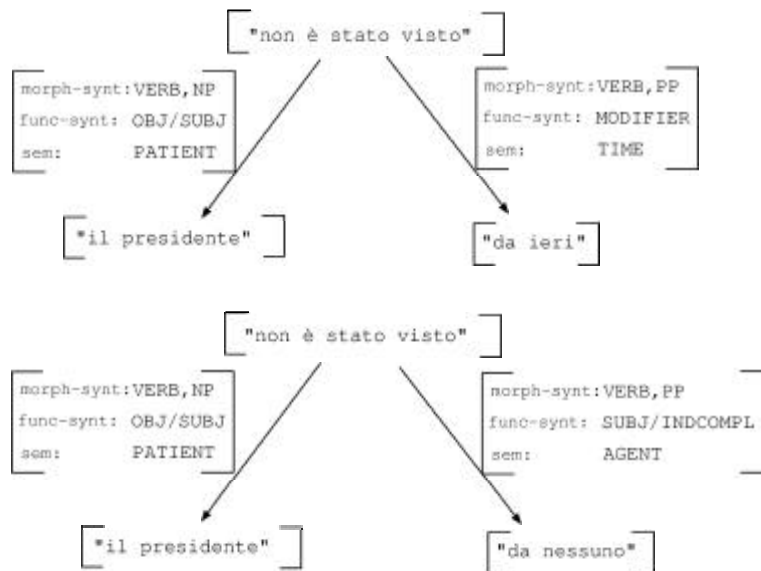


Fig. 4. *The argument-modifier distinction in the functional-syntactic component.*

ated with respect to the func-synt component: in *il presidente non è stato visto da ieri* (the president has not been seen since yesterday), the preposition *da* introduces a modifier, whilst in *il presidente non è stato visto da nessuno* (the president has not been seen by nobody (i.e. nobody has seen the president)) it introduces an argument of the verb, the agent complement¹).

¹ The X/Y notation refers to a relation X ‘trasformed’ into the relation Y. In the examples of fig. 4 the unit “il presidente” has been transformed from

The functions of the functional-syntactic component are organized in a hierarchy[2], as in various relational approaches to syntax (see, e.g., [11]). The hierarchical distribution of syntactic functions depends on criteria that define the degree of specification of a relation: the hierarchy proposed in this work distinguishes between Arguments and Modifiers. Arguments include Subject, Object, Indirect Complement, Predicative Complements (of the Subject and of the Object, respectively); modifiers are split into Restrictive Modifiers and Appositions. Syntactic functions also realize syntactically marked phenomena such as various forms of Coordination (correlative, comparative or adversative), Verbal Auxiliary functions (Tense, Progressive, Passive markers), Idiomatic utterance dependencies, Clausal Separators (through punctuation), Proper Name constructions. As it will be clear below, the hierarchy is useful when the annotator is not certain about some syntactic dependency and prefers to assign some less specific function.

3.3 The semantic component

The semantic component specifies the role of syntactic units in the syntax-semantics interface. It makes explicit the role that the participants play in the event described by the predicate (semantic or thematic roles such as Agent or Receiver), and supplementary information about the event (Location, Time, Manner, ...). Therefore it discriminates among different kinds of modifiers and arguments: for instance, in the following examples the functional-syntactic structure is the same, but the modifiers have different semantic roles (i.e. a measure of the amount of the wine or the material which the glass is made of, respectively).

un bicchiere di vino (a glass of wine)
un bicchiere di cristallo (a glass of crystal, i.e. a crystal glass)

The actual development of the semantic component in terms of the labels to be included in an annotation schema is a very hard task. 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. The approach pursued in this work has been to annotate very specific semantic roles only when they are immediately and neatly distinguishable. Attributes like Age (e.g., “un uomo di 50 anni” - *a 50 year old man*) or semantic roles like TransportMean (e.g., “George venne in autobus” - *George came by bus*) are easily assigned; the semantic role of “senza metodi pesanti” (*without violent methods*) in the example of fig. 5 is not easy to assign and is left unspecified.

Now we move to the annotation schema.

Object to Subject because of passivization. For details see linguistic notes at <http://www.di.unito.it/~tutreeb/>

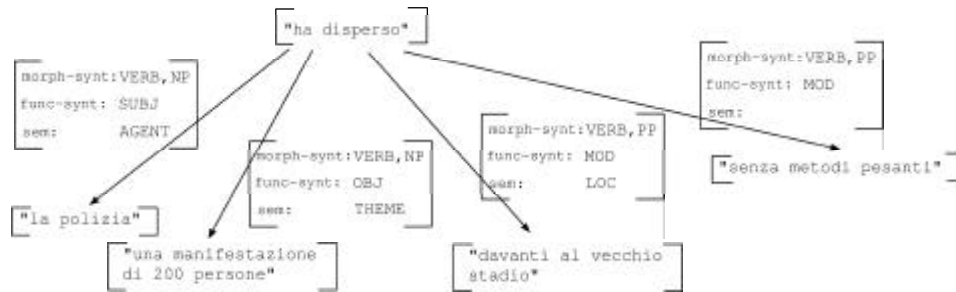


Fig. 5. A representation of the sentence "la polizia ha disperso una manifestazione di 200 persone davanti al vecchio stadio senza metodi violenti" (The police dispersed a demonstration of 200 people in front of the old stadium without violent methods)

3.4 The Augmented Relational Structure and the annotation schema

In this section we see how ARS has been used to develop a treebank annotation schema. First, we recall the major features of ARS and then we illustrate the implementation of the schema.

ARS is a single-layered structure that describes grammatical relations in terms of three separate components. This neat distinction of the components has not been an issue in the other treebank approaches described above, in the theoretical assumptions (Negra does not provide any partition of the functions), and in the annotation practice (Penn makes the theoretical distinction but annotation is practically restricted to a single function tag per dependency). In some approaches, the three components are viewed as distinct annotation layers (like in Prague). Even if the structure at the semantic level is not always isomorphic to the structure at the syntactic one (e.g., in quantifier scoping or coordination), in general, for phenomena that NLP applications take into account, we can represent semantic dependencies without going beyond the syntax-semantics interface. The reliability of such an approach is confirmed, for instance, by [23] which presents a syntax-semantic monostratal annotation, and by Prague too where both analytical and tectogrammatical layers share the same structure.

A consequence of a single-layered approach is that there are no problems of inter-layer alignment which must be solved in tasks involving more than one layer (e.g., PP-attachment in parsing), and which are usually hard to implement because of structural differences among independent levels. For instance, in Prague the inter-layers alignment is rather complex, because the number of nodes of the tectogrammatical (semantic) level is different from that at the analytical (syntactic) level. Finally, the tripartite structure of the relations in ARS guarantees that different components can be accessed and analyzed independently (like in [20]).

The ARS-based approach has been applied to the development of an Italian corpus, the Turin University Treebank (TUT, <http://www.di.unito.it/~tutreeb/>).

In the case of TUT, the implemented annotation schema is a dependency-based representation. In this schema the syntactic units are single words and the grammatical relations are dependency relations described as tripartite feature structures.

Given the ARS-based annotation schema, two are the issues at hand in the development of the treebank: what are the actual feature values in the three components that form the grammatical relations and how they scale when the corpus comes to a relevant size.

The choice of an adequate set of feature values for each component is a critical one. The use of a large variety of values provides a great accuracy and specialization in the description of relations, but contrasts with reasonable time in annotation, since annotators' task becomes very time-consuming for the assignment of the correct relation. This is due to two factors: first, the selection of the correct grammatical relation can be more difficult when the search space consists of a large number of competing labels; second, the specificity of relations can decrease the inter-annotator agreement. Finally, a great number of feature values requires for an application a very large set of training data, otherwise the problem of sparseness becomes relevant and the statistical models are not reliable. On the other hand, if we maintain the feature values at very general levels the annotation schema provides very few hints of the syntactic dependencies, and it is likely that statistical models trained on these data yield the same results in a large number of situations.

The solution in our approach is to adopt a variable degree of specificity, implemented in two ways: one is to drop some of the three components, the other is the hierarchical organization of the syntactic functions. Let us start from the first. All the relations annotated in TUT include the functional-syntactic component. This is the only component in the 23.1% of the relations, in case they represent purely syntactic dependencies like, e.g., idiomatic expressions or proper name compounds. The 35.2% of relations lacks a morpho-syntactic component. Only the 12.1% of relations include the semantic component. The motivations for such a low percentage of semantically annotated relations are: first, not all the relations carry semantic information (in fact, some relations are purely syntactic, see above, and typically the relations linking function words, like determiners or prepositions, do not carry relevant semantic contribution and in the Prague treebank are pruned at the semantic level); second, at this stage of the treebank development, the semantic component is systematically annotated only on modifiers and indirect complements (exclusively distinguished on a semantic basis), but not on terms (subject, object and indirect object)². The reference to ontological and/or lexico-semantic resources will further increase the consistency and accuracy of the semantic annotation ([21] and [20] use a database of predicate-argument structures).

The second implementation of variable degree of specificity relies on the hierarchy of functions, in that the annotator provides the most specific relation

² A consistent semantic annotation of terms, e.g., as agent, patient, is a future task we will perform by using a database of subcategorization frames.

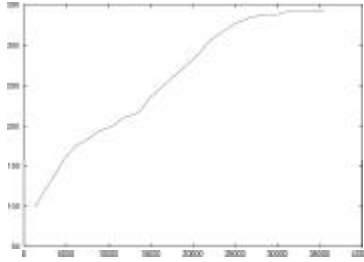


Fig. 6. *Grow of grammatical relations for the 1,200 sentences.*

he/she feels confident about. If specific annotations result in a problem of data sparseness, the hierarchy can provide a controlled mechanism of backing-off.

The hierarchical organization also allows the treatment of vagueness, because the high levels of the hierarchy offer broad categories of relations (e.g., the distinction between arguments and modifiers, or between restrictive and non-restrictive modifiers). The hierarchy is again a way of dealing with ambiguity, when the annotator is not able to select the correct relation. For instance, in the theoretical example *"Mario non l'ha ancora visto Gianni"* (*Mario has not seen Gianni yet/ Gianni has not seen Mario yet*), both *Mario* and *Gianni* can be subject or object [7]. In this case, the annotators can use the hierarchy in the bottom-top direction, from specific to generic levels to find a common parent of the involved relations. Other treebank annotation schemata offer different solutions: Prague assigns a second label or adds a special marker that denotes the uncertainty of the assignment; Penn uses a default anonymous constituent label X (when the bracketing is certain, but the label is uncertain) and the so-called *pseudo-attachment* (for structures that are globally ambiguous as attachment).

In order to see how this annotation schema scales over corpora of large sizes, we provide some data on the annotation of the TUT. The TUT consists of newspaper articles (about 60%), civil code (legal language - about 30%), miscellaneous (about 10%). Currently, we have annotated 1,200 sentences (about 35,000 words) with the current numbers of feature values: the morpho-syntactic component includes 40 items, the functional-syntactic component includes 71 items, the semantic-syntactic component includes 102 items. These 213 items combined into 343 valid feature structures (grammatical relations). This means that only few combinations are meaningful. All the three components seem to be very stabilized, and although it is preliminary to assert anything about the behavior of the same components in a very large corpus, the grow of the feature-values set have been very limited in the sentences from 950 to 1,200 (see fig. 6), and this is very promising for the future.

4 Conclusion

The paper presented a schema for treebank annotation based on an Augmented Relational Structure. The ARS is a representation of syntactic dependencies (aka grammatical relations) as tripartite feature structures, namely complex objects which take into account various interrelated informational domains, called components (morpho-syntactic, functional-syntactic and semantic-syntactic). The paper has illustrated the three components in detail, and has provided some quantitative data on the annotation of the Turin University Treebank.

References

1. Abeillé, A. (ed.): Building and using syntactically annotated corpora. Kluwer, Dordrecht (2003)
2. Bosco, C.: Grammatical relation's system in treebank annotation. In: Miltsakaki, E., Monz, C., Ribeiro, A. (eds.): Proceedings of Student Research Workshop of ACL/EACL, Toulouse France (2001) 1–6
3. Brants, S., Dipper, S., Hansen, S., Lezius, W., Smith, G.: The TIGER Treebank. Proceedings of the Workshop on Treebanks and Linguistic Theories, Szopól Bulgaria (2002)
4. Brants, T., Skut, W., Uszkoreit, H.: Syntactic annotation of a German newspaper corpus. Proceedings of Treebanks workshop - Journées ATALA sur les corpus annotés pour la syntaxe, Paris (1999) 69–76
5. Bresnan, J. (ed.): The mental representation of grammatical relations. MIT Press, Cambridge, Mass (1982)
6. Briscoe, T.: From dictionary to corpus to self-organizing dictionary: learning valency associations in the face of variation and change. Proceedings of Corpus Linguistics 2001, Lancaster UK (2001) 79–89
7. Carrol, J., Briscoe, E., Sanfilippo, A.: Parser evaluation: a survey and a new proposal. Proceedings of LREC98, Granada, Spain (1998) 447–454
8. Charniak, E.: A maximum entropy inspired parser. Proceedings of the First NAACL, Seattle WA (2000) 132–139
9. Collins, M.: Head-driven statistical models for natural language parsing. Ph.D. thesis, University of Pennsylvania, <http://www.ai.mit.edu/people/mcollins/> (1999)
10. Dipper, S., Brants, T., Lezius, W., Plaehn, O., Smith, G.: The TIGER treebank. Proceedings of LINC'01, Leuven Belgium (2001)
11. N. M. Fraser, R. A. Hudson: Inheritance in Word Grammar, Computational Linguistics, Vol. 18, Num. 2 (1992) 133–158
12. Haijčová, E.: Dependency-based underlying-structure tagging of a very large Czech corpus. In: Kahane, S. (ed.): Traitement automatique de langues. Vol. 41, Num. 1 (2000) 57–78
13. Haijčová, E., Kučerová, I.: Argument/valency structure in Propbank, LCS database and Prague Dependency Treebank: a comparative study. Proceedings of LREC 2002, Las Palmas Spain (2002) 846–851
14. Han, C., Lavoie, B., Palmer, M., Rambow, O., Kittredge, R., Korelsky, T., Kim, N., Kim, M.: Handling Structural Divergences and Recovering Dropped Arguments in a Korean/English Machine Translation System. Proceedings of the Association for Machine Translation in the Americas 2000, LNAI Springer Verlag (2000)

15. Kinsbury, P., Palmer, M.: From TreeBank to PropBank. Proceedings of LREC 2002, Las Palmas Spain (2002)
16. Litkowski, K.C.: Question-Answering Using Semantic Relation Triples. Proceedings of TREC-8. NIST, Gaithersburg MD (1999) 349–356
17. MacGee Wood, M.: *Categorial grammar*. Routledge (1993)
18. Marcus, M.P., Santorini, B., Marcinkiewicz, M.A.: Building a Large Annotated Corpus of English. The Penn Treebank. *Computational Linguistics*, Vol. 19 (1993) 313–330
19. Marcus, M.P., Kim, G., Marcinkiewicz, M.A., MacIntyre, R., Bies, A., Ferguson, M., Katz, K., Schasberger, B.: The Penn Treebank: Annotating Predicate Argument Structure. Proceedings of The Human Language Technology Workshop, Morgan-Kaufmann, San Francisco (1994)
20. Montemagni, S., Barsotti, F., Battista, M., Calzolari, N., Corazzari, O., Lenci, A., Zampolli, A., Fanciulli, F., Massetani, M., Raffaelli, R., Basili, R., Pazienza, M.T., Saracino, D., Zanzotto, F., Mana, N., Pianesi, F., Delmonte, R.: Building the Italian Syntactic-Semantic Treebank. In: Abeill'e, A. (ed.): *Building and using parsing corpora*, Kluwer Dordrecht (2003) 189–210
21. Palmer, M., Dang, H.T., Rosenzweig, J.: Semantic tagging for the Penn Treebank. Proceedings of LREC 2000, Athens Greece (2000) 699–704
22. Perlmutter, D.M.: *Studies in Relational Grammar 1*. University of Chicago Press, Chicago (1983)
23. Rambow, O., Creswell, C., Szekely, R., Taber, H., Walker, M.: A dependency treebank for English. Proceedings of LREC 2000, Athens Greece (2000) 857–863
24. Sgall, P., Haijcová, E., Panevová, J.: *The meaning of the sentence in its pragmatic aspects*. Reidel Publishing Company (1986)
25. Shieber, S.M.: *An introduction to unification-based approaches to grammar*. CSLI, Stanford (1986)
26. Smith, G.: Encoding thematic roles via syntactic functions in a German treebank. Proceedings of workshop on syntactic annotation of electronic corpora. Tübingen, Germany (2002)
27. Vilain, M.: Inferential Information Extraction. In: Pazienza, M.T. (ed.): *Information Extraction*. LNAI 1714, Springer (1999) 95–119.
28. Vilares, J., Barcala, F.M., Alonso, M.A.: Using syntactic dependency-pairs to improve retrieval performance in Spanish. Proceedings of CICLing-2002 (2002) 381–390