Deep Syntactic Annotation: Tectogrammatical Representation and Beyond

Petr Sgall
Center for
Computational Linguistics
sgall@ufal.mff.cuni.cz

Jarmila Panevová
Institute of Formal
and Applied Linguistics
panevova@ufal.mff.cuni.cz

Eva Hajičová
Center for
Computational Linguistics
hajicova@ufal.mff.cuni.cz

Abstract

The requirements of the depth and precision of annotation vary for different intended uses of the corpus but it has been commonly accepted nowadays that the standard annotations of surface structure are only the first steps in a more ambitious research program, aiming at a creation of advanced resources for most different systems of natural language processing and for testing and further enrichment of linguistic and computational theories. Among the several possible directions in which we believe the standard annotation systems should go (and in some cases already attempt to go) beyond the POS tagging or shallow syntactic annotations, the following four are characterized in the present contribution: (i) predicate-argument representation of the underlying syntactic relations as basically corresponding to a rooted tree that can be univocally linearized, (ii) the inclusion of the information structure using very simple means (the left-to-right order of the nodes and three attribute values), (iii) relating this underlying structure (rendering the "linguistic meaning," i.e. the semantically relevant counterparts of the grammatical means of expression) to certain central aspects of referential semantics (reference assignment and coreferential relations), and (iv) handling of word sense disambiguation. The first three issues are documented in the present paper on the basis of our experience with the development of the structure and scenario of the Prague Dependency Treebank which provides for syntactico-semantic annotation of large text segments from the Czech National Corpus and which is based on a solid theoretical framework.

1 Introduction[1]

It has been commonly accepted within the computational linguistics community involved in corpus annotation that part-of-speech tagging and shallow syntactic annotation though very progressive, important and useful tasks at their time, are only the first steps in a more ambitious research program, aiming at a creation of advanced resources for most different systems of natural language processing and for testing and further enrichment of linguistic and computational theories. On the basis of our experience with the development and implementation of the annotation scheme of the Prague Dependency Treebank (PDT, (Hajič et al., 2001a), (Böhmová, 2004)), we would like to indicate four directions, in which we believe the standard annotation systems should be (and in some cases already attempt to be) extended to fulfill the present expectations. We believe that we can offer useful insights in three of these, namely

(i) an adequate and perspicuous way to represent the underlying syntactic relations as basically corresponding to a rooted tree that can be equivocally linearized,

(ii) with the inclusion of the information structure using very simple means (the left-to-right order of the nodes and two indexes), and

(iii) in relating this underlying structure (rendering the "linguistic meaning," i.e. the semantically relevant counterparts of the grammatical means of expression) to certain central issues of referential semantics (reference assignment and coreferential relations).

[1]The research reported on in this paper has been supported by the project of the Czech Ministry of Education LN00A063, and by the grants of the Grant Agency of the Czech Republic No. 405/03/0913 and No. 405/03/0377.

These insights have been elaborated into annotation guidelines which now are being used (and checked) on the basis of PDT, i.e. of syntactico-semantic annotations of large text segments from the Czech National Corpus, which allows for a reliable confirmation of the adequacy of the chosen theoretical framework and for its enrichment in individual details. The fourth dimension we have in mind is that of handling word-sense disambiguation, for which the material of our annotated texts, the PDT, serves as a starting point.

2 The Extensions

2.1 The Structure: Deep Dependency and Valency

The development of formal theories of grammar has documented that when going beyond the shallow grammatical structure toward some kind of functional or semantic structure, two notions become of fundamental importance: the notion of the head of the structure and the notion of valency (i.e. of the requirements that the heads impose on the structure they are heads of). To adduce just some reflections of this tendency from American linguistic scene, Fillmore's "case grammar" (with verb frames) and his FrameNet ((Fillmore et al., 2003)), Bresnan's and Kaplans's lexical functional grammar (with the distinction between the constituent and the functional structure and with an interesting classification of functions) and Starosta's "lexicase grammar" can serve as the earliest examples. To put it in terms of formal syntactic frameworks, the phrase structure models take on at least some traits of the dependency models of language; Robinson has shown that even though Fillmore leaves the issue of formal representation open, the phrase-structure based sentence structure he proposes can be easily and adequately transposed in terms of dependency grammar. Dependency account of sentence structure is deeply rooted in European linguistic tradition and it is no wonder then that formal descriptions originating in Europe are dependency-based (see Sgall, Kunze, Hellwig, Hudson, Mel'chuk). We understand it as crucial to use sentence representations "deep" enough to be adequate as an input to a procedure of semantic(-pragmatic) interpretation (i.e. representing function words and endings by indexes of node labels, restoring items which are deleted in the morphemic or phonemic forms of sentences and distinguishing tens of kinds of syntactic relations), rather than to be satisfied with some kind of "surface" syntax.

The above-mentioned development of formal frameworks toward an inclusion of valency in some way or another has found its reflection in the annotation scenarios that aimed at going beyond the shallow structure of sentences. An important support for annotation conceived in this way can be found in schemes that are based on an investigation of the subcategorization of lexical units that function as heads of complex structures, see. Fillmore's FRAMENET, the PropBank as a further stage of the development of the Penn Treebank (Palmer et al., 2001) and Levin's verb classes (Levin, 1993) on which the LCS Database (Dorr, 2001) is based. There are other systems working with some kind of "deep syntactic" annotation, e.g. the broadly conceived Italian project carried out in Pisa (N. Calzolari, A. Zampolli) or the Taiwanese project MARVS; another related framework is presented by the German project NEGRA, basically surface oriented, with which the newly produced subcorpus TIGER contains more information on lexical semantics. Most work that has already been carried out concerns subcategorization frames (valency) of verbs but this restriction is not necessary: not only verbs but also nouns or adjectives and adverbs may have their "frames" or "grids".

One of the first complex projects aimed at a deep (underlying) syntactic annotation of a large corpus is the already mentioned Prague Dependency Treebank (Hajič, 1998); it is designed as a complex annotation of Czech texts (taken from the Czech National Corpus); the underlying syntactic dependency relations (called functors) are captured in the *tectogrammatical tree structures* (TGTS); see (Hajičová, 2000). The set of functors comprises 53 valency types subclassified into (inner) participants (arguments) and (free) modifications (adjuncts). Some of the free modifications are further subcategorized into more subtle classes (constituting mainly the underlying counterparts, or meanings, of prepositions).

Each verb entry in the lexicon is assigned a valency frame specifying which type of participant or modification can be associated with the given verb; the valency frame also specifies which participant/modification is obligatory and which is optional with the given verb entry (in the underlying representations of sentences), which of them is deletable on the surface, which may or must function as a controller, and so on. Also nouns and adjectives have their valency frames.

The shape of TGTSs as well as the repertory and classification of the types of modifications of the verbs is based on the theoretical framework of the Functional Generative Description, developed by the Prague research team of theoretical and computational linguistics as an alternative to Chomskyan transformational grammar (Sgall et al., 1986). The first two arguments, though labeled by "semantically descriptive" tags ACT and PAT (Actor and Patient, respectively) correspond to the first and the second argument of a verb (cf. Tesnière's (Tesnière, 1959) first and second actant), the other three arguments of the verb being then differentiated (in accordance with semantic considerations) as ADDR(essee), ORIG(in) or EFF(ect); these five functors belong to the set of participants (arguments) and are distinguished from (free) modifications (adjuncts)

such as LOC(ative), several types of directional and temporal (e.g. TWHEN) modifications, APP(urtenance), R(e)STR(ictive attribute), DIFF(erence), PREC(eding cotext referred to), etc. on the basis of two basic operational criteria (Panevová, 1974), (Panevová, 1994):

(i) can the given type of modification modify in principle every verb?

(ii) can the given type of modification occur in the clause more than once?

If the answers to (i) and (ii) are yes, then the modification is an adjunct, if not, then we face an argument.

We assume that the cognitive roles can be determined on the basis of combinations of the functors with the lexical meanings of individual verbs (or other words), e.g. the Actor of *buy* is the buyer, that of *sell* is the seller, the Addressee and the Patient of *tell* are the experiencer and the object of the message, respectively.The valency dictionary created for and used during the annotation of the Prague Dependency Treebank, called PDT-VALLEX, is described in (Hajič et al., 2003). The relation between function and (morphological) form as used in the valency lexicon is described in (Hajič and Urešová, 2003).

An illustration of this framework is presented in Fig. 1.

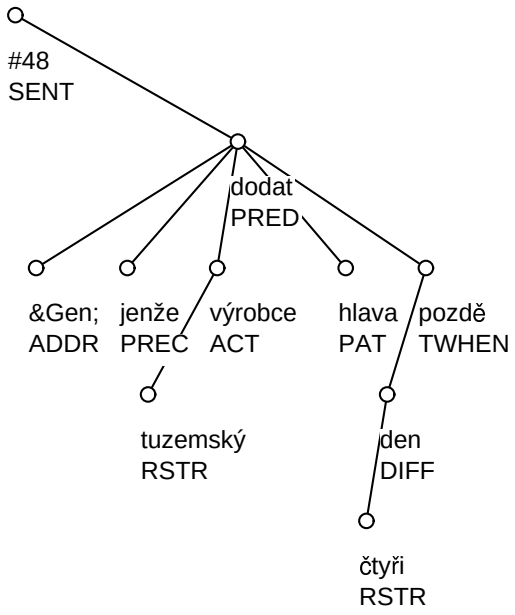

Figure 1: A simplified TGTS of the Czech sentence *Jenže tuzemský výrobce dostal hlavy o čtyři dny později. 'However, the domestic producer got the heads four days later.'*

Let us adduce further two examples in which the functors are written in capitals in the otherwise strongly simplified representations, where most of the adjuncts are understood as depending on nouns, whereas the other functors concern the syntactic relations to the verb. Let us note that with the verb *arrive* the above mentioned test determines the Directional as a (semantically) obligatory item that can be specified by the hearer according to the given context (basically, as here or there):

(1) Jane changed her house from a shabby cottage into a comfortable home.

(1') Jane.ACT changed her.APP house.PAT from-a-shabby.RSTR cottage.ORIG into-a-comfortable.RSTR home.EFF.

(2) Yesterday Jim arrived by car.

(2') Yesterday.TWHEN Jim.ACT arrived here.DIR3 by-car.MEANS.

A formulation of an annotation scenario based on well-specified subcategorization criteria helps to compare different schemes and to draw some conclusions from such a comparison. In (Hajičová and Kučerová, 2002) the authors attempt to investigate how different frameworks annotating some kind of deep (underlying) syntactic level (the LCS Data, PropBank and PDT) compare with each other (having in mind also a more practical application, namely a machine translation project the modules of which would be "machine-learned", using a procedure based on syntactically annotated parallel corpora). We are convinced that such a pilot study may also contribute to the discussions on a possibility/impossibility of formulating a "theory neutral" syntactic annotation scheme. The idea of a theory neutral annotation scenario seems to be an unrealistic goal: it is hardly possible to imagine a classification of such a complex subsystem of language as the syntactic relations are, without a well motivated theoretical background; moreover, the languages of the annotated texts are of different types, and the theoretical frameworks the authors of the schemes are used to work with differ in the "depth" or abstractness of the classification of the syntactic relations. However, the different annotation schemes seem to be translatable if the distinctions made in them are stated as explicitly as possible, with the use of operational criteria, and supported by larger sentential contexts. The third condition is made realistic by very large text corpora being available electronically; making the first two conditions a realistic goal is fully in the hands of the designers of the schemes.

2.2 Topic/Focus Articulation

Another aspect of the sentence structure that has to be taken into account when going beyond the shallow structure of sentences is the communicative function of the sentence, reflected in its information structure. As has

been convincingly argued for during decades of linguistic discussions (see studies by Rooth, Steedman, and several others, and esp. the argumentation in (Hajičová et al., 1998)), the information structure of the sentence (topic-focus articulation, TFA in the sequel) is semantically relevant and as such belongs to the semantic structure of the sentence. A typical declarative sentence expresses that its focus holds about its topic, and this articulation has its consequences for the truth conditions, especially for the differences between meaning proper, presuppositiona and allegations (see (Hajičová, 1993); (Hajičová et al., 1998)).

TFA often is understood to constitute a level of its own, but this is not necessary, and it would not be simple to determine the relationships between this level and the other layers of language structure. In the Functional Generative Description (Sgall et al., 1986), TFA is captured as one of the basic aspects of the underlying structure, namely as the left-to-right dimension of the dependency tree, working with the basic opposition of contextual boundness; the contextually bound (CB) nodes stand to the left of the non-bound (NB) nodes, with the verb as the root of the tree being either contextually bound or non-bound.

It should be noted that the opposition of NB/CB is the linguistically patterned counterpart of the cognitive (and pre-systemic) opposition of "given" and "new" information. Thus, e.g. in (3) the pronoun *him* (being NB), in fact constitutes the focus of the sentence.

(3) (We met a young pair.) My older companion recognized only HIM.

In the prototypical case, NB items belong to the focus of the sentence, and CB ones constitute its topic; secondary cases concern items which are embedded more deeply than to depend on the main verb of the sentence, cf. the position of *older* in (3), which may be understood as NB, although it belongs to the topic (being an adjunct of the CB noun companion).

In the tectogrammatical structures of the PDT annotation scenario, we work with three values of the TFA attribute, namely t (contextually bound node), c (contextually bound contrastive node) and f (contextually non-bound node). 20,000 sentences of the PDT have already been annotated in this way, and the consistency and agreement of the annotators is being evaluated. It seems to be a doable task to annotate and check the whole set of TGTSs (i.e. 55,000 sentences) by the end of 2004. This means that by that time the whole set of 55,000 sentences will be annotated (and checked for consistency) on both aspects of deep syntactic structure. An algorithm the input of which are the TGTSs with their TFA values and the output of which is the division of the whole sentence structure into the (global) topic and the (global) focus is being formulated.

2.3 Coreference

The inclusion into the annotation scheme of the two aspects mentioned above in Sect. 2.1 and 2.2, namely the deep syntactic relations and topic-focus articulation, considerably extends the scenario in a desirable way, toward a more complex representation of the meaning of the sentence. The third aspect, the account of coreferential relations, goes beyond linguistic meaning proper toward what can be called the sense of the utterance (Sgall, 1994). Two kinds of coreferential relations have to be distinguished: grammatical coreference (i.e. with verbs of control, with reflexive pronouns, with verbal complements and with relative pronouns) and textual (which may cross sentence boundaries), both endophoric and exophoric.

Several annotation schemes have been reported at recent conferences (ACL, LREC) that attempt at a representation of coreference relations in continuous texts. As an example of an attempt to integrate the treatment of anaphora into a complex deep syntactic scenario, we would like to present here a brief sketch of the scheme realized in the Prague Dependency Treebank. For the time being, we are concerned with coreference relations in their narrower sense, i.e. not covering the so-called bridging anaphora (for a possibility to cover also the latter phenomenon, see (Böhmová, 2004)).

In the Prague Dependency Treebank, coreference is understood as an asymmetrical binary relation between nodes of a TGTS (not necessarily the same TGTS), or, as the case may be, as a relation between a node and an entity that has no corresponding counterpart in the TGTS(s). The node from which the coreferential link leads, is called an anaphor, and the node, to which the link leads, is called an antecedent.

The present scenario of the PDT provides three coreferential attributes: *coref*, *cortype* and *corlemma*. The attribute *coref* contains the identifier of the antecedent; if there are more than one antecedents of one anaphor, the attribute *coref* includes a sequence of identifiers of the relevant antecedents; since every node of a TGTS has an identifier of its own it is a simple programming task to select the specific information on the antecedent. The attribute *cortype* includes the information on the type of coreference (the possible values are `gram` for grammatical and `text` for textual coreference), or a sequence of the types of coreference, where each element of *cortype* corresponds to an element of *coref*. The attribute *corlemma* is used for cases of a coreference between a node and an entity that has no corresponding counterpart in the TGTS(s): for the time being, there are two possible values of this attribute, namely `segm` in the case of a coreferential link to a whole segment of the preceding text (not just a sentence), and `exoph` in the case of an exophoric relation. Cases of reference difficult to be identified even if the situation is taken into account are marked by the

assignment of unsp as the lemma of the anaphor. This does not mean that a decision is to be made between two or more referents but that the reference cannot be fully specified even within a broader context.

In order to facilitate the task of the annotators and to make the resulting structures more transparent and telling, the coreference relations are captured by arrows leading from the anaphor to the antecedent and the types of coreference are distinguished by different colors of the arrows. There are certain notational devices used in cases when the antecedent is not within the co-text (exophoric coreference) or when the link should lead to a whole segment rather than to a particular node. If the anaphor corefers to more than a single node or to a subtree, the link leads to the closest preceding coreferring node (subtree). If there is a possibility to choose between a link to an antecedent or to a postcedent, the link always leads to the antecedent.

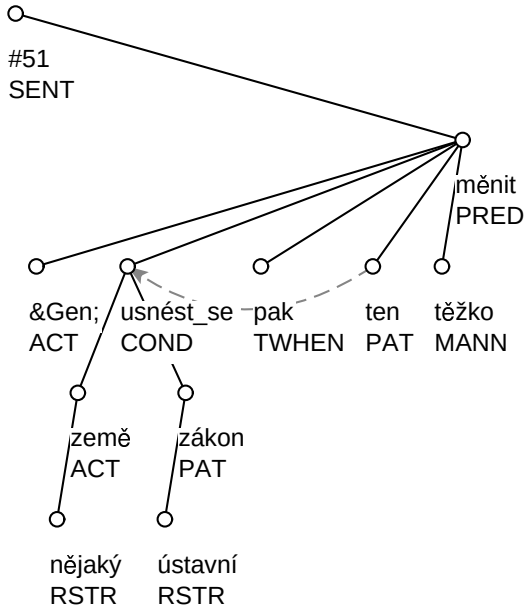

Figure 2: A TGTS of the sentence *Pokud se nějaká země usnese na ústavním zákonu, pak se to těžko mění. 'If a country accepts a constitution law, then this is difficult to change.*'

The manual annotation is made user-friendly by a special module within the TRED editor (Hajič et al., 2001b) which is being used for all three subareas of annotation. In the case of coreference, an automatic pre-selection of nodes relevant for annotation is used, making the process faster.

Until now, about 30,000 sentences have been annotated as for the above types of coreference relations. One of the advantages of a corpus-based study of a language phenomenon is that the researchers become aware of subtleties and nuances that are not apparent. For those who attempt at a corpus annotation, of course, it is necessary to collect a list of open questions which have a temporary solution but which should be studied more intensively and to a greater detail in the future.

Another issue the study of which is significant and can be facilitated by an availability of a semantically annotated corpus, is the question of a (finite) mechanism the listener (reader) can use to identify the referents. If the backbone of such a mechanism is seen in the hierarchy (partial ordering) of salience, then it can be understood that this hierarchy typically is modified by the flow of discourse in a way that was specified and illustrated by (Hajičová, 1993), (Hajičová et al., in prep). In the flow of a discourse, prototypically, a new discourse referent emerges as corresponding to a lexical occurrence that carries f; further occurrences carry t or c, their referents being primarily determined by their degrees of salience, although the difference between the lowest degrees of salience reduction, is not decisive. It appears to be possible to capture at least certain aspects of this hierarchy by some (still tentative) heuristic rules, which tie up the increase/decrease of salience with the position of the given item in the topic or in the focus of the given utterance. It should also be remarked that there are certain permanently salient referents, which may be referred to by items in topic (as "given" information) without having a referentially identical antecedent in the discourse. We denote them as carrying t or c, but perhaps it would be more adequate to consider them as being always able to be accommodated

(i) by the utterance itself, as especially the indexicals (*I*, *you*, *here*, *now*, *yesterday*,...),

(ii) by the given culture (*democracy*, *Paris*, *Shakespeare*, *don Quijote*,...), by universal human experience (*sun*, *sky*), or

(iii) by the general domain concerned (history, biology,...).

Since every node in the PDT carries one of the TFA values (t, c or f) from which the appurtenance of the given item to the topic or focus of the whole sentence can be determined, it will be possible to use the PDT data and the above heuristics to start experiments with an automatic assignment of coreferential relations and check them against the data with the manual annotation of coreference.

2.4 Lexical Semantics

The design of the tectogrammatical representation is such that the nodes in the tectogrammatical tree structure rep-

resent (almost) only the autosemantic words found in the written or spoken utterance they represent. We believe that it is thus natural to start distinguishing word senses only at this level (and not on a lower level, such as surface syntax or linearized text).

Moreover, there is a close relation between valency and word senses. We hypothesize that with a suitable set of dependency relations (both inner participants and free modifications, see Sect. 2.1), there is only one valency frame per word sense (even though synonyms or near synonyms might have different valency frames). The opposite is not true: there can be several word senses with an identical valency frame.

Although in the detailed valency lexicon VALLEX (Lopatková, 2003), (Lopatková et al., 2003) an attempt has originally been made to link the valency frames to (Czech) EuroWordNet (Pala and Ševeček, 1999) senses to prove this point, this has been abandoned for the time being because of the idiosyncrasies in WordNet design, which does not allow to do so properly.

We thus proceed independently with word sense annotation based on the Czech version of WordNet. Currently, we have annotated 10,000 sentences with word senses, both nouns and verbs. We are assessing now further directions in annotation; due to low inter-annotator agreement, we will probably tend to annotate only over a preselected subset of the WordNet synsets. An approach to building semantic lexicons that is more related to our concept of meaning representation is being prepared in the meantime (Holub and Straňák, 2003).

3 Conclusions

Up to now, the framework has been checked on a large amount of running text segments from the Czech National Corpus (as for the valency classification 55,000 utterances, as for the topic-focus structure 20,000 ones). In several cases, it was found that a more detailed classification is needed (e.g. with the differentiation of the General Actor vs. Unspecified, cf. the difference between *One can cook well with this oven* and *At this pub they cook well*). However, it has been confirmed that good results can be achieved with the chosen classification of about 40 valency types and of 15 other grammatical attribute types (such as (Semantic) Number, Tense, Modalities, etc., but also different values of Location, such as those corresponding to the preferred functions of *in*, *at*, *on*, *under*, *over*, etc., or of Benefactive (positive vs. negative), and so on). It can be supposed that the core of language corresponds to underlying sentence structures and to their unmarked morphemic and phonemic counterparts. The marked layers have to be described by specific sets of rules, most of which concern irregularities of morphemics, including differences between the underlying order of nodes and the surface (morphemic) word order, especially in cases in which the latter does not directly meet the condition of projectivity (with no crossing of edges, cf. the discontinuous constituents of other frameworks).

The prototypical varieties of sentence structure can thus be characterized by projective rooted trees, which points to the possibility to describe the core of language structure on the basis of a maximally perspicuous pattern that comes close to patterns present in other domains (primitive logic, arithmetics, and so on) which are normally mastered by children. Structures of this kind are not only appropriate for computer implementation, but they also help understand the relative easiness of mastering the mother tongue, without a necessity to assume complex innate mechanisms specific for the language faculty.

References

Alena Böhmová. 2004. *Automatized Procedures in the Process of Annotation of PDT*. Ph.D. thesis, Charles University, Faculty of Mathemartics and Physics, Prague.

Bonnie Dorr. 2001. The LCS Database. `http://www.umiacs.umd.edu/ ~bonnie/ LCS_Database_Documentation.html`.

Charles J. Fillmore, Christopher R. Robinson, and Miriam R. L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography*, 16:235–250.

Eva Hajičová and Ivona Kučerová. 2002. Argument/valency structure in PropBank, LCS database and Prague Dependency Treebank: A comparative study. In *Proceedings of LREC*.

Jan Hajič and Zdeňka Urešová. 2003. Linguistic Annotation: from Links to Cross-Layer Lexicons. In *Proceedings of The Second Workshop on Treebanks and Linguistic Theories*, volume 9 of *Mathematical Modeling in Physics, Engineering and Cognitive Sciences*, pages 69–80. Växjö University Press, November 14–15, 2003.

Jan Hajič, Eva Hajičová, Petr Pajas, Jarmila Panevová, Petr Sgall, and Barbora Vidová-Hladká. 2001a. Prague Dependency Treebank 1.0 (Final Production Label). CDROM CAT: LDC2001T10, ISBN 1-58563-212-0.

Jan Hajič, Petr Pajas, and Barbora Hladká. 2001b. The Prague Dependency Treebank: Annotation Structure and Support. In *IRCS Workshop on Linguistic Databases*, pages 105–114, Philadelphia, PA, Dec. 11–13.

Jan Hajič, Alevtina Bémová, Petr Pajas, Jarmila Panevová, Veronika Řezníčková, and Zdeňka Urešová. 2003. PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation. In Joakim

Nivre and Erhard Hinrichs, editors, *2nd International Workshop on Treebanks and Linguistic Theories*, volume 9 of *Mathematical Modeling in Physics, Engineering and Cognitive Sciences*, pages 57–68. Växjö University Press, Växjö, Sweden, Nov. 14–15, 2003.

Jan Hajič. 1998. Building a Syntactically Annotated Corpus: The Prague Dependency Treebank. In Eva Hajičová, editor, *Issues of Valency and Meaning. Studies in Honor of Jarmila Panevová*, pages 12–19. Prague Karolinum, Charles University Press.

Eva Hajičová, Barbara Partee, and Petr Sgall. 1998. *Topic-focus articulation, tripartite structures, and semantic content*. Kluwer Academic Publishers, Amsterdam, Netherlands.

Eva Hajičová, Jiří Havelka, and Petr Sgall. in prep. Topic and Focus, Anaphoric Relations and Degrees of Salience. In *Prague Linguistic Circle Papers 5*. Amsterdam/Philadelphia: John Benjamins.

Eva Hajičová. 1993. *Issues of Sentence Structure and Discourse*. Charles University, Prague, Czech Republic.

Eva Hajičová. 2000. Dependency-Based Underlying-Structure Tagging of a Very Large Czech Corpus. In *Special issue of TAL journal, Grammaires de Dépendence / Dependency Grammars (ed. Sylvian Kahane)*, pages 57–78. Hermes.

Martin Holub and Pavel Straňák. 2003. Approaches to building semantic lexicons. In *WDS'03 Proceedings of Contributed Papers, Part I*, pages 173–178, Prague. MATFYZPRESS, Charles University.

Beth Levin. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. The University of Chicago Press.

Markéta Lopatková, Zdeněk Žabokrtský, Karolína Skwarska, and Václava Benešová. 2003. VALLEX 1.0. `http://ckl.mff.cuni.cz/zabokrtsky/vallex/1.0`.

Markéta Lopatková. 2003. Valency in the Prague Dependency Treebank: Building the Valency Lexicon. *Prague Bulletin of Mathematical Linguistics*, 79–80:in print.

Karel Pala and Pavel Ševeček. 1999. Czech wordnet. `http://www.fi.muni.cz/nlp/grants/ewn_cz.ps.en`.

Martha Palmer, J Rosenzweig, and S Cotton. 2001. Automatic Predicate Argument Analysis of the Penn TreeBank. In J Allan, editor, *Processdings of HLT 2001, First Int. Conference on Human Technology Research*. Morgan Kaufmann, San Francisco.

Jarmila Panevová. 1974. On verbal Frames in Functional Generative Description. *Prague Bulletin of Mathematical Linguistics*, 22:3–40.

Jarmila Panevová. 1994. Valency Frames and the Meaning of the Sentence. In Philip Luelsdorff, editor, *The Prague School of Structural and Functional Linguistics*, pages 223–243. John Benjamins, Amsterdam-Philadephia.

Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The meaning of the sentence and its semantic and pragmatic aspects*. Reidel, Dordrecht.

Petr Sgall. 1994. Meaning, Reference and Discourse Patterns. In Philip Luelsdorff, editor, *The Prague School of Structural and Functional Linguistics*, pages 277–309. John Benjamins, Amsterdam-Philadephia.

Lucien Tesnière. 1959. *Élements de Syntaxe Structurale*. Klincksieck, Paris.