Sentence Structure and Discourse Structure: Possible Parallels

Pavlína Jínová, Lucie Mladová, Jiří Mírovský
Charles University in Prague, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics (UFAL MFF CUNI)
Malostranské nám. 25, 118 00 Prague 1, Czech Republic
{jinova,mladova,mirovsky}@ufal.mff.cuni.cz

Abstract

The present contribution represents the first step in comparing the nature of syntactico-semantic relations present in the sentence structure to their equivalents in the discourse structure. The study is carried out on the basis of a Czech manually annotated material collected in the Prague Dependency Treebank (PDT). According to the semantic analysis of the underlying syntactic structure of a sentence (tectogrammatics) in the PDT, we distinguish various types of relations that can be expressed both within a single sentence (i.e. in a tree) and in a larger text, beyond the sentence boundary (between trees). We suggest that, on the one hand, each type of these relations preserves its semantic nature both within a sentence and in a larger text (i.e. a causal relation remains a causal relation) but, on the other hand, according to the semantic properties of the relations, their distribution in a sentence or between sentences is very diverse. In this study, this observation is analyzed for two cases (relations of condition and specification) and further supported by similar behaviour of the English data from the Penn Discourse Treebank.

1 Motivation and Background

Although the annotation in the Prague Dependency Treebank 2.0 (PDT, Hajič et al., 2006; Mikulová et al., 2005) in principle does not surpass the sentence boundaries, i.e. each sentence is represented by a single dependency tree structure, to a certain extent, the information about the context has always been one of its concerns. First, the context of every sentence is reflected in one attribute of the nodes in the syntactico-semantic (tectogrammatical) structure: the information structure of the sentence (Topic-Focus Articulation, TFA, cf. Sgall, Hajičová and Panevová, 1986; Hajičová, Partee and Sgall, 1998), second, some basic coreference relations are marked (especially the grammatical coreference and some types of the textual coreference). In recent years, the interest in analyzing the structure of discourse in a more complex way has increased, and also the PDT is being enriched with this type of information. After having annotated the anaphoric chains and also the so-called bridging relations (or the association anaphora, see Nedoluzhko et al., 2009), the annotation of semantic relations between text spans indicated by certain discourse markers is now in progress. This annotation has two linguistic resources: besides the Prague (syntactico-semantic) approach it is inspired also by the Penn Discourse Treebank 2.0 approach based on identifying discourse connectives and their arguments (Prasad et. al, 2007 and 2008).

One of the benefits of annotating discourse semantic relations on tectogrammatical trees is a possibility to exploit the syntactico-semantic information already captured in the corpus. This fact also enables us to compare the nature of relations expressed both within a single sentence (in a single tree) and in a larger text (between trees). Since the discourse annotation of the PDT is still a work in progress, it is premature to make some final conclusions in this respect. On the other hand, a majority of the corpus has already been processed and some tendencies are evident. In the present contribution we therefore want to introduce some observations about the nature of these corresponding relations and support them with our data.

The contribution is divided into three main parts. In Section 2, we describe some basic aspects of the Praguian approach to the syntactic structure (tectogrammatics); criteria according to which some relations from the tectogrammatics are considered to be discourse relations are introduced in Section 3; and in Section 4 a comparison of intra-sentential and inter-sen-

tential (discourse) relations is carried out on an example of two semantic relations from our manually annotated data.

2 Basic Aspects of the Underlying Syntactic Structure in the PDT Relevant for Discourse

There are three basic aspects of the syntactic structure already captured on the tectogrammatical layer in the PDT (see also Mladová et al., 2008) that are relevant for the discourse structure analysis: (i) the dependency edge between nodes filled with finite verbs (i.e. the relation between a subordinate clause and its governing clause), (ii) the coordination connecting finite-verbal nodes (i.e. the relation between coordinate clauses), and (iii) nodes with the label "reference to PREceding Context" (PREC) (i.e. the label for such expressions as *however, hence* and *so forth*). The subordinate and coordinate structures are classified according to their syntactico-semantic values and some of these values can be directly transferred to the discourse annotation (e.g. the semantic label of a subordinate clause "cause" corresponds in the vast majority of cases with its discourse counterpart). However, in other cases, the set of semantic values of the edges is not subcategorized enough for the discourse annotation and it needed to be classified in a more detailed way (e.g. the only semantic label for adversative meaning on the tectogrammatical layer was for the purpose of the discourse annotation divided into finer types of contrast, such as opposition, restrictive opposition and correction, cf. Zikánová, 2007). Moreover, one special type of relation – apposition – and the meaning indicated by expressions with the label PREC were not semantically interpreted at all on the tectogrammatical layer. The notion of apposition is descriptive, it stands for a syntactic structure with one syntactic position filled by two formally independent nominal or verbal phrases that are referentially at least partially identical (e.g. *he has only one obsession: he spends at least two hours a day playing computer games*). It follows that the notion of apposition is semantically too abstract for the purposes of the discourse annotation and so it was also subcategorized and re-labeled (see Figure 1 below in Section 4.2).

3 Discourse Annotation

3.1 Discourse Relevance of Intra-sentential Relations

From our point of view, there is a necessary condition for each syntactico-semantic relation (taken from the tectogrammatical analysis, Mikulová et al., 2005) to be considered a discourse relation: its possibility to relate two syntactically independent sentences. In other words, it must be possible in a natural language to relate two independent text spans with semantically exactly the same meaning, as there is on the syntactic level (often more plausibly) between the governing verbal node and its complement, dependent node; or, in a compound sentence, between the coordinate (verbal) clauses[1].

Another, milder requirement concerns the connective means of each relation. Whereas the transparency of the sentence semantics depends on the presence of subordinating connectives, which anchor the meaning (Mladová, 2009), we prefer to treat a syntactico-semantic relation as discourse-applicable, if we can find a corresponding means to the subordinating expression on the discourse level. In some cases, this is quite easy, such as in (1)[2]: in (1a), the discourse-semantic relation occurs between a subordinate clause and its governing clause, whereas in (1b) it relates two independent sentences.

(1)
(a) *[Arg1: **Protože** slovenská elita byla zklamána politickou volbou Slovenska,]*
[Arg2: většina kvalitních odborníků zůstala v Praze.]

*[Arg1: **Because** Slovak political elite was disappointed by the political choice of Slovakia,]*
[Arg2: the majority of skilled professionals remained in Prague.]

(b) *[Arg1: Slovenská elita byla zklamána politickou volbou Slovenska.]*
*[Arg2: **Proto** většina kvalitních odborníků zůstala v Praze.]*

[1] For the first phase of the discourse annotation, only clauses headed by a finite verb were taken to be discourse-level units. Nominalizations and other clause-like phenomena are to be explored for their discourse functions in the next phases of the project.

[2] Abbreviations Arg1 and Arg2 are used in examples for indication of the two text spans between which the discourse semantic relation occurs. Connectives are in bold.

[Arg1: Slovak political elite was disappointed by the polical choice of Slovakia.]

*[Arg2: **Therefore**, the majority of skilled professionals remained in Prague.]*

As for coordinated clauses, the situation is very simple. Coordinated clauses in a compound sentence always play the role of discourse arguments and their conjunction is a discourse-level connective.[3] This applies not only for structures connected by connectives such as *therefore, but, or* etc. but also when the coordinating connective is represented by a "mere" punctuation mark like a dash (see (2)) or a colon (see (3)). According to their semantics, these structures can be reformulated as two independent sentences (two trees) either by adjacency without any connective (the case of (2)) or by independent sentences linked with an explicit connective. In the case of (3), the connective *totiž* (in this context without any appropriate English equivalent, perhaps it can be roughly translated as "that is to say" or "as a matter of fact", depending on the context) can be used in the part after the colon.

Example (2) demonstrates a discourse semantic relation expressed (a) by a coordinative structure with a dash and (b) by two independent sentences.

(2)

(a) *[Arg1: Sparta přenechává volné pole konkurenci]*

[Arg2: – Látal odešel do Schalke 04, Hogen se Šmejkalem jsou ve Slavii, Poborský září na Žižkově.]

[Arg1: FC Sparta leaves the field open to competition]

[Arg2: – Látal left to Schalke 04, Hogen and Šmejkal are in Slavia, Poborský shines in FC Žižkov.]

(b) *[Arg1: Sparta přenechává volné pole konkurenci.]*

[Arg2: Látal odešel do Schalke 04, Hogen se Šmejkalem jsou ve Slavii, Poborský září na Žižkově.]

[Arg1: FC Sparta leaves the field open to competition.]

[Arg2: Látal left to Schalke 04, Hogen and Šmejkal are in Slavia, Poborský shines in FC Žižkov.]

[3] Coordinative connectives often connect also text spans larger than one sentence.

Example (3) illustrates the discourse semantic relation expressed (a) in a coordinative structure with a colon and (b) by two independent sentences:

(3)

(a) *[Arg1: Zdá se, že to byl šťastný krok]*

*[Arg2: **:** provinční rumunský časopis se vyhranil jako médium autorů kvalitní literatury z celé Evropy.]*

[Arg1: This step seems to have been lucky]

*[Arg2: **:** the provincial Romanian magazine crystallized into a platform of high quality literature from the whole Europe.]*

(b) *[Arg1: Zdá se, že to byl šťastný krok.]*

*[Arg2: Provinční rumunský časopis se **(totiž)** vyhranil jako médium autorů kvalitní literatury z celé Evropy.]*

[Arg1: This step seems to have been lucky.]

[Arg2: The provincial Romanian magazine crystallized into a platform of high quality literature from the whole Europe.]

Moreover, it turned out that this "punctuating" type of connecting discourse units is preferable in certain types of relations, see Section 4.2 below.

Third, in some cases, such a reformulation is not possible without a loss of the original meaning (as pointed out in Mladová et al., 2009) so that the syntactico-semantic relation does not hold inter-sententially.[4] Hence, subordinate clauses which can be expressed as independent pieces of discourse without having changed their meaning (and, as mentioned, also coordinate clauses) are considered discourse-level units connected with a discourse relation, others are not.

3.2 Basic Aspects of Discourse Annotation

In our approach to discourse we decided in the first phase to annotate only semantic relations between units (text spans) containing a finite

[4] Consider for example the following sentence (A) from Mladová et al., 2009. The syntactic form of the construction does not allow to express this type of relation by independent sentences (B).

(A) *The older the wine, the better it is. (Čím je víno starší, tím je lepší.)*

(B) **The older is the wine. The better it is. (*Čím je víno starší. Tím je lepší.)*

verb and indicated by an explicit connective.[5] The hierarchy of discourse sense labels was established on the basis of the tectogrammatical labels (see Mikulová et al., 2005) and the Penn hierarchy of sense tags (Miltsakaki et al., 2008). The original Penn division of the sense tags to four major categories is preserved: we differentiate temporal, contingency, contrast (comparison) and expansion relations.

In the following section, we show tendencies in the behaviour of two particular discourse relations observed during the annotation process in the PDT.

4 Two Semantic Relations Expressed both in a Sentence and in a Text

We have now at our disposal approximately 33,000 sentences of Czech texts annotated both for the underlying syntax (tectogrammatics) and for the discourse structure. We believe this builds a solid base for looking for certain tendencies in the behaviour of individual semantic relations. In the course of the development of the data annotation, we have built a hypothesis that there is a certain scale (though we do not yet present claims about determining its end points) that determines to what extent a language prefers a semantic relation to be expressed more likely within a single sentence or between sentences. In the following sections, we give examples of two relations that act very differently in this respect – condition and specification. These two relations, in our opinion, demonstrate two poles of the scale.

4.1 The Case of Condition

Mladová et al. (2009) demonstrated that the semantic relation of condition, often expressed intra-sententialy, can be easily projected into an inter-sentential relation by using different language means (e.g. *if* + subordinate clause → *but* + modal verb in the second sentence), for clarification purposes we cite the example sentences below under (4):

(4)

(a) *[Arg 1: I will cook pancakes,] [Arg2: **if** you buy eggs.]*

(b) *[Arg 1: I will cook pancakes.] [Arg2: **But** you must buy eggs first.]*

[5] The only exception is the relation between a text span introducing a list structure (so-called hyper-theme) and the items of the list structure – (i) in our approach, they can be annotated also without any explicit connective, (ii) the hyper-theme needs not to be a verbal clause.

Nonetheless, our annotation indicates that in reality this type of a semantic relation strongly tends to be expressed within a sentence, as a relation between the main verb and its conditional modifier – a subordinate clause. The formulation of a conditional meaning in a language[6] seems to be closely associated with the occurrence of a (subordinating) connective such as *if* or *when* – in Czech mainly *pokud, zda, jestli(že)*. The overview of all possible syntactic forms of condition with their distribution in the 33 thousand sentences from the PDT is presented in Table 1:

Sentence/ Discourse	**Syntactic form of condition**	**Number of occurrences in the PDT sample**[7]
within one sentence (tree)	non-clausal modifier of the main predicate verb[8]	651
	dependent clause (clausal (= verbal) modifier of the main predicate verb)[9]	963
between sentences (trees)	between adjacent sentences[10]	7
	long-distant relation	0

Table 1. Distribution of various types of expressing conditional meaning in the PDT

Table 1 indicates that the usage of the inter-sentential relation of condition is quite rare.

[6] at least in languages like English or Czech

[7] 33,000 sentences of Czech journalistic texts

[8] Example (expression of condition in bold): *Kupující, který je získal za tisíc korun, je tedy např. může další den **v případě zájmu** prodat za 1 100 Kč. (A buyer who got them for 1000 Czech crowns can **in case of interest** sell them the next day for 1,100 Czech crowns.)*

[9] Example (expression of condition in bold): ***Pokud pracovník nemůže závazku z vážných důvodů dostát,** omluví se včas a navrhne jiné řešení. (**If an employee for serious reasons cannot meet the obligations,** he must apologize and suggest in good time a different solution.)*

[10] Example (expression of condition in bold): ***Posluchač musí přistoupit na pozici, že vše je dovoleno.** Potom se pobaví a také pochopí, že drama znázorňuje ztrátu reálné komunikace. (**The listener has to accept the position that everything is permitted.** Then he enjoys the drama and also understands that it symbolizes the loss of a real-life communication.)*

Moreover, the cases we found of such a use occur strictly between adjacent sentences, which means, the relation of condition applies neither for long distances nor between larger text units. All the cases of inter-sententially expressed conditional relations have a similar structure like the example in Table 1 (see footnote 10) – with the connective *potom/pak* (*then*) in the second argument. These findings imply that the nature of the given condition is very closely bound to the information in the second text span (the result of the condition). The best setting for relating these two pieces of information in communication is a clear delimitation of a sentence. Thus, we can state that in the repertoire of discourse-semantic relations, the condition relation tends to be one of the most condensed, the most syntax-bound.

To find out more about this matter, we compared the acquired numbers for Czech with those that were measured over the English data of the Penn Discourse Treebank (Prasad et al., 2007)[11]. The situation is quite similar – the absolute majority of the conditional relations was assigned to discourse connectives like *if* or *when* and their modifications (e.g. *at least when, especially if, even if, if and when, only when, particularly if, until, unless* etc.), which are all subordinate.[12] Hence, also for English holds that the conditional meaning tends to be expressed within a single sentence. Having discovered this symmetry, there arises an assumption that must be first verified in the course of a more detailed research, that, to a certain extent, this phenomenon is language-independent.

4.2 The Case of Specification

The semantic relation of specification occurs between two text spans when the second one describes something already expressed in the first one but in more detail. This relation corresponds on the sentential level in the PDT to the notion of apposition – the piece of information in the second span is not a new one, it only completes the information in the preceding context. In other words, when a specification relation is to be expressed intra-sententially, it fills a single syntactical position twice (see Figure 1) – first with a piece of information to some extent general, second with its details.

This relation has not been described in traditional Czech grammars[13] and therefore many instances of the specification relation are interpreted also as conjunction in the PDT. Specification applied intra-sententially is exemplified by (5)[14] (and also by Figure 1), an inter-sentential one is displayed in (6).

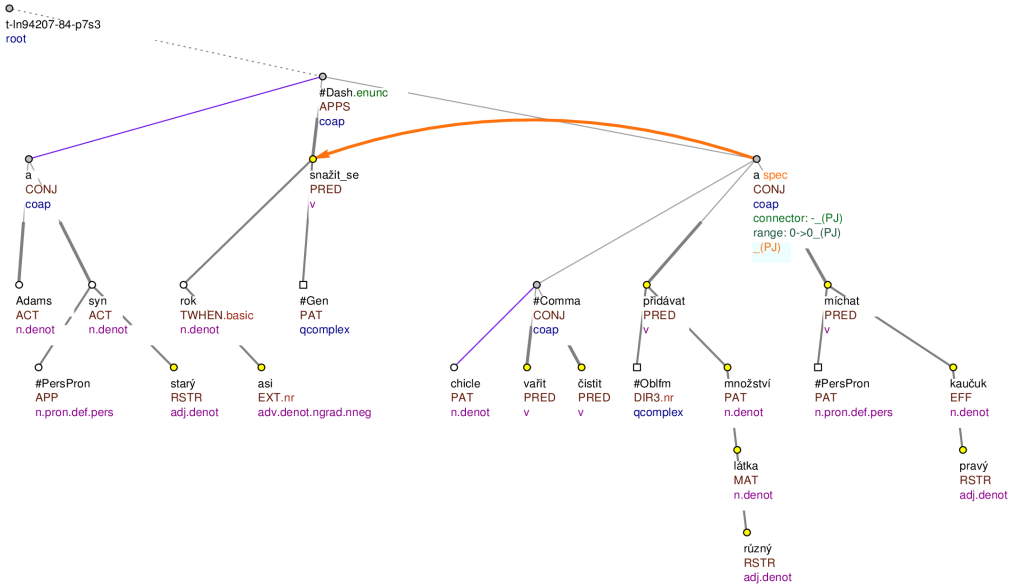

Figure 1. Apposition of two verbal nodes with the predicate function. (At the same time an example of intra-sentential specification (the bold arrow with the label spec). For the example sentence and its English translation see (5)).

[11] Approx. 49,000 sentences annotated for discourse structure.

[12] Exact distribution numbers for each connective see in Prasad et al. (2007).

[13] as they concern primarily the issues of sentence syntax and semantics in deeper insight

[14] Some necessary context is given in brackets.

(5)

[Arg1: Asi rok se Adams a jeho nejstarší syn snažili]

[Arg2: – chicle vařili, čistili, přidávali množství různých látek a míchali s pravým kaučukem.]

[Arg1: Adams and his eldest son struggled about a year]

[Arg2:– they cooked chicle, cleaned it, added to it a number of different substances and mixed it with genuine rubber.]

In (6) the semantic relation of specification occurs inter-sententially. The second sentence describes the details of the scoring and the bad start mentioned in the first sentence. This specification is indicated by the connective *totiž,* which does not have any English equivalent in this context (it can be perhaps loosely translated by "as a matter of fact").

(6)

[Arg1: Po dvou porážkách ve Frýdku-Místku a v Příbrami konečně zabral Havířov, ačkoliv premiéru na vlastním hřišti nezačal dobře.]

*[Arg2: Pardubice se **totiž** ujaly vedení Plockovou bombou ve 26. minutě, ale domácí otočili skóre třemi góly v rozpětí dvaceti minut na přelomu prvního a druhého poločasu.]*

[Arg1: Havířov finally scored after two defeats in Frýdek-Místek and in Příbram, although the premiere at its own field did not start well.]

*[Arg2: Pardubice (**totiž**) took lead in the 26th minute by Plock´s bomb but players from Havířov turned the score by three goals within twenty minutes at the turn of the first and the second halves.]*

The current part of the PDT annotated for discourse structure contains 339 occurrences of the specification relation. 244 of them are expressed within one tree, only 95 between trees (moreover, 60 cases from these 95 occurrences represent the relation between a hyper-theme and a list structure and as such they either stand without connectives (36 occurrences) or are indicated by a colon (24 occurrences)). The most common connectives are punctuation marks: a colon (151 occurrences) and a dash (57 occurrences). Not only there is just one "non-punctuating" connective associated primarily with this relation – the compound connective *a to (and that),* but its occurrence is also restricted to special structures with an elided verb. Other "non-punctuating" connectives associated with specification are rather typical for other relations (for results summary see Table 2). We have not found any subordinate structure to express the specification relation.

Sentence/ Discourse		**Specification indicated by**	**Number of occurrences in PDT sample**[15]
within one sentence (tree)		„non-punctuating" connective	78
		punctuation mark	166
between sentences (trees)	list structure	punctuation mark	24
		no surface connective	36
	other structure	punctuation mark	8
		„non-punctuating" connective	27

Table 2. The distribution of the specification relation in the PDT

The decision to annotate in the first phase only relations indicated by explicit connectives limited especially the number of captured inter-sentential specifications. However, the fact that specification is the second most frequent relation with an implicit connective in the Penn Discourse Treebank (PDTB, 2,471 occurrences (Prasad et al., 2007: 90)) but it has a very low frequency when represented by explicit connectives (108 occurrences, Prasad et al., 2007: 75) supports our observation that, also in the PDT, this relation is expressed very often without any explicit connective. And this comparison enables us to go even further. If we take into account the fact that punctuation marks are supposed to be implicit connectives in the PDTB (and therefore we can only include 105 occurrences of specification in the PDT for the purpose of the comparison), we can claim that the semantic relation of specification strongly tends to be expressed inter-sententially. Only

[15] 33,000 sentences of Czech journalistic texts

inter-sententially expressed specifications indicated by no surface connective can explain the evident discrepancy between our and the PDTB data (see also Table 3).

PDT sample		PDTB	
Specification indicated by	**Number of occurrences**	**Specification indicated by**	**Number of occurrences**
"non-punctuating" connective	105	explicit connective	108
punctuation mark	198	implicit connective	2,471
no surface connective (list structure)	36		
no surface connective in other structures	not included into annotation		

Table 3. Comparison of the distribution of the specification relation in the PDT and in the PDTB

To sum up, the specification relation is indicated preferably by punctuation marks or by the pure adjacency of sentences and the only means of its expression in one sentence is a coordinate structure. The comparison with the PDTB data supports our observation that this semantic relation is expressed primarily inter-sententially. These findings result, in our opinion, from the semantic nature of specification – the information in the second text span is not very closely bound to the information in the first text span, it only supplements the information that has already been given. Therefore, we can claim that the nature of specification is connected with the discourse structure rather than with the sentence structure.

5 Conclusion

We have demonstrated on two examples of discourse-semantic relations – condition and specification – that there are great differences in the nature of these relations, namely in their distribution in the discourse structure. Whereas the conditional meaning is expressed primarily within a single sentence and it is in an absolute majority of cases bound by a subordinate form of expression and a usage of hypotactic language means, for the meaning of specification it is rather the opposite: it prefers to be expressed between sentences, via adjacency and with no discourse connectives at all or just with punctuation marks as a colon or a dash.

The aim of this study was to demonstrate that semantic relations between discourse units are not on the same level, but, on the contrary, their nature is quite different according to their semantic properties. In this regard, we consider the two analyzed relations to represent two poles of a scale leading from the language means used in the sentential syntax to those used in the discourse composition.

Second, the analysis of Czech and English language data processed on the basis of a similar theoretical background indicates that the findings about the nature of these semantic relations are in both languages identical, and this analysis further leads to the assumption that this phenomenon might be, at least to a certain extent, language independent.

For further enhancement of our findings, studies in three directions would be ideal to follow: (i) an analysis of the distribution of other discourse-semantic relations, for instance those from the contrast group (as we assume they might stay somewhere in between), (ii) an analysis of the distribution of discourse semantic relations in various genres (our findings are based on journalistic texts), and (iii) a comparison with data from a third, preferably typologically different language.

Acknowledgements

The research reported in this contribution has been carried out under the grant projects of the Grant Agency of the Czech Republic (GA405/09/0729), the Center for Computational Linguistics (LC536) and the Grant Agency of Charles University in Prague (GAUK 103609). We would like to thank anonymous reviewers for their detailed and insightful comments. All mistakes are ours.

References

Hajič, Jan, Eva Hajičová, Jarmila Panevová, Petr Sgall, Jan Štěpánek, Jiří Havelka and Marie Mikulová. 2006. *Prague Dependency Treebank 2.0*. Philadelphia: Linguistic Data Consortium.

Hajičová, Eva, Barbara H. Partee and Petr Sgall. 1998. *Topic, focus articulation, tripartite struc-*

tures and semantic content. Dodrecht: Kluwer Academic Press.

Mikulová, Marie et al. 2005. *Annotation on the Tectogrammatical Layer in the Prague Dependency Treebank: Annotation Manual*. Prague: Universitas Carolina Pragensis.

Miltsakaki, Eleni, Livio Robaldo, Alan Lee and Aravind Joshi. 2008. *Sense annotation in the Penn Discourse Treebank.* In Computational Linguistics and Intelligent Text Processing, Lecture Notes in Computer Science, vol. 4919: pp. 275–286.

Mladová, Lucie, Šárka Zikánová and Eva Hajičová. 2008. *From Sentence to Discourse: Building an Annotation Scheme for Discourse Based on Prague Dependency Treebank.* In Proceedings of the 6th International Conference on Language Resources and Evaluation, CD-ROM.

Mladová, Lucie, Šárka Zikánová, Zuzanna Bedřichová and Eva Hajičová. 2009. *Towards a Discourse Corpus of Czech.* In Proceedings of the fifth Corpus LinguisticsConference, Liverpool, UK, in press.

Mladová, Lucie. 2009. *Annotation of Discourse Connectives for the PDT.* In WDS'09 Proceedings of Contributed Papers. Praha, Czechia.

Nedoluzhko, Anja, Jiří Mírovský and Petr Pajas. 2009. *The Coding Scheme for Annotating Extended Nominal Coreference and Bridging Anaphora in the Prague Dependency Treebank.* In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing, Suntec, Singapore.

Prasad, Rashmi, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi and Bonnie Webber. 2008. *The Penn Discourse Treebank 2.0.* In Proceedings of the 6th International Conference on Language Resources and Evaluation, CD-ROM.

Prasad, Rashmi et al. 2007. *The Penn Discourse TreeBank 2.0 Annotation Manual,* available at:http://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf.

Sgall, Petr, Eva Hajičová and Jarmila Panevová. 1986. *The Meaning of the Sentence and its Semantic and Pragmatic Aspects*. Praha: Academia.

Zikánová, Šárka. 2007. *Possibilities of Discourse Annotation in Prague Dependency Treebank (Based on the Penn Discourse Treebank Annotation).* Technical report. Institute of Formal and Applied Linguistics, Charles University, Prague.

Zikánová, Šárka, Lucie Mladová, Jiří Mírovský and Pavlína Jínová. 2010. *Typical Cases of Annotators' Disagreement in Discourse Annotations in Prague Dependency Treebank.* In Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010), Valletta, Malta. Pages range: 2002–2006.