

Interplay of Coreference and Discourse Relations: Discourse Connectives with a Referential Component

Lucie Poláková, Pavlína Jínová, Jiří Mírovský

Charles University in Prague, Faculty of Mathematics and Physics

Institute of Formal and Applied Linguistics

Malostranské nám. 25, 118 00 Prague 1, Czech Republic

E-mail: {polakova, jinova, mirovsky}@ufal.mff.cuni.cz

Abstract

This contribution explores the subgroup of text structuring expressions with the form *preposition + demonstrative pronoun*, thus it is devoted to an aspect of the interaction of coreference relations and relations signaled by discourse connectives (DCs) in a text. The demonstrative pronoun typically signals a referential link to an antecedent, whereas the whole expression can, but does not have to, carry a discourse meaning in sense of discourse connectives. We describe the properties of these phrases/expressions with regard to their antecedents, their position among the text-structuring language means and their features typical for the “connective function” of them compared to their “non-connective function”. The analysis is carried out on Czech data from the approx. 50,000 sentences of the Prague Dependency Treebank 2.0, directly on the syntactic trees. We explore the characteristics of these phrases/expressions discovered during two projects: the manual annotation of coreference relations (Nedoluzhko et al. 2011) and discourse connectives, their scopes and meanings (Mladová et al. 2008).

Keywords: discourse connectives, coreference, manual annotation of large corpora

1. Motivation

In the linguistic research of phenomena “beyond the sentence boundary”, we typically ask questions, how, with what language means a coherent, well formulated text (discourse) is established. There are several different aspects contributing to the coherence of a text, several viewpoints of possible analysis, which, we believe, are closely related to each other. We can analyze the referential structure of a text, we can look for a rhetorical structure (as cf. Carlson and Marcu 2001), we can study speech acts and intentions (cf. Grosz et al. 1995), polarity and subjectivity (cf. Wiebe et al. 2004), information structure (Sgall et al. 1986) etc. The aim of this paper is to show one point of intersection of two aspects of discourse analysis – the interaction of coreference relations and relations signaled by discourse connectives (DCs).

Within the group of text structuring expressions (or discourse markers in the broad sense), there is a subgroup of expressions that have the form *preposition + demonstrative pronoun*. The gradual development of some of these forms from demonstratives to stable phrases with potentially connective meaning, the question of orthography (written as one or two units), a possible anaphoric link of the demonstrative part of the pattern to an antecedent, the nature of this antecedent – all this brings together interesting material from two different ways of discourse analysis. We believe that our observations are valid not only for Czech, but, in general, this tendency of historical language development is visible also in English and in other languages (e.g. Heine and Kuteva 2002).

The contribution consists of these steps: in Section 2, basic facts about discourse and coreference annotation in the Prague Dependency Treebank (PDT) are given,

Section 3 describes our motivation for the extension of the class of DCs, it brings a hypothesis for the identification of the connectives with a referential component among other expressions with the same form, it lists their types and describes their roles in discourse structuring. The content of Sections 4 and 5 is the analysis of the treebank data.

2. Discourse and Coreference Annotation in Prague Dependency Treebank

Our analysis is carried out on Czech data, taking advantage of the unique multilayer annotation of the approx. 50,000 sentences (3,165 journalistic documents) of the Prague Dependency Treebank 2.0 (PDT 2.0, Hajič et al. 2006). Recently, this language resource was enriched with few new layers of manual annotation concerning discourse phenomena: extended textual coreference annotation, some relations of bridging anaphora (Nedoluzhko et al. 2011) and annotation of discourse connectives, their scopes and meanings (Mladová et al. 2008). (At the time of its release, 2006, the PDT 2.0 already contained morphological information, syntactic trees with semantic roles assigned to the nodes, annotation of information structure (topic – focus articulation), grammatically bound coreference relations and pronominal textual coreference, see Mikulová et al. 2005). The recent annotation of discourse is carried out directly on the syntactic trees, which brings along many advantages. For instance, intra-sentential discourse relations have been already marked in the syntactic layer, elliptical constructions have been already resolved, and discourse connectives partially labeled (see Mladová et al. 2008).

3. Connectives with a Referential Component

A discourse connective in PDT is defined as a predicate of a binary relation which takes two text spans (mainly clauses or sentences) as its arguments. It connects these units to larger ones while signaling a semantic relation between them at the same time. Discourse connectives are in principle morphologically inflexible and they do not act as grammatical constituents of a sentence. Like modality markers, they are “above” or “outside” of the proposition. In general, they are represented by coordinating conjunctions (e.g. *and*, *but*), some subordinating conjunctions (e.g. *because*, *if*, *while*), some particles (e.g. *also*, *only*) and sentence adverbials (e.g. *afterwards*), and marginally also by some other parts-of-speech – mainly in cases of fixed compound connectives like *in other words* or *on the contrary*.

However, in course of the manual annotations it turned out necessary to extend the notion of a discourse connective to a group of (inflected) prepositional phrases that are not always fixed, and so not easy to distinguish. Our hypothesis was that the highly ambiguous pattern *preposition + demonstrative pronoun* can in a certain context act as a discourse connective: In some cases, it has the same functional characteristics as “classical”, basic connectives. In those and only those cases, first, it can be substituted with one of them (e.g. *kromě toho = apart from that*, lit. *besides that* can be in certain context substituted by *navíc = moreover*; *navzdory tomu = contrary to*, lit. *despite that* can be substituted by *přesto (yet, nevertheless)* and, second, the antecedent of the demonstrative pronoun should be an action, an event, not an entity. That means the antecedent would be a whole clause represented with a finite verb, not an NP. Once there is a corpus annotated for coreference relations available, this hypothesis is easy to prove.

Another reason for this extension of the DCs is the undeniable discourse function of certain prepositions in connection with nominalizations, such as *because of his arrival = because he arrived*. In some approaches to discourse segmentation, these prepositions are taken as regular discourse connectives (cf. Stede 2007, 167f). This principle is from the theoretical point of view legitimate but large annotation projects like the one in PDT have to limit themselves to only marking constructions with finite verbs (clauses) as discourse units.

In addition, the relevance of the connections *preposition + demonstrative pronoun* for marking discourse relations is evident from the form of some basic Czech DCs. Some common Czech discourse connectives, as well as many of their counterparts in English, originated from the composition of a demonstrative pronoun with a preposition (e.g. *proto (therefore, lit. for-that)*, *přesto (yet, lit. despite-of-that)*, *potom (then, lit. after-that)* etc.). The Czech linguistic literature mentions this fact (e.g. Mluvnice češtiny 1986) but this phenomenon has not been, at least as far as we know, studied in more detail up to now.

The tricky part about some connections of prepositions and demonstratives in Czech is the historical development of the forms and the change in their writing (separate x together) that goes not always together with the switch in their functions. Therefore, to see this phenomenon as a whole, we decided to analyze all expressions with the structure *preposition + demonstrative pronoun* that can have in a certain context the function of a DC. In this paper, we work with a classification of these forms introduced below in section 3.1.

3.1 Classes of the *Preposition + Demonstrative Pronoun* Forms

Basically, the connections of a preposition and a demonstrative pronoun with potential discourse function can be divided into three groups according to their written form:

1. **Type *PrepDem***: Connectives always written as one unit – bound form (e. g. *proto (therefore)*, *zatím (meanwhile, but)*). Their “divided forms” (e. g. *pro to (for that)*, *za tím (behind that)*) never function as a connective;
2. **Type *PrepDem/Prep_Dem***: Expressions sometimes written as one unit, sometimes as two units, but they can function as connectives in both forms (e. g. *přesto/přes to (yet/despite of that)*, *potom/ po tom (then/after that)*, *přitom/při tom (while, however/along with that)*);
3. **Type *Prep_Dem***: Expressions always written as two units – compound form, they sometimes do and sometimes do not function as connectives (e.g. *kvůli tomu (lit. due to that, sometimes with the meaning because)*, *vedle toho (lit. besides that, sometimes with the meaning in addition, moreover)*).

3.2 Coreference Links Establishing DCs

As we suggested at the beginning of this section, the structure *preposition + demonstrative pronoun* is closer to the connective meaning if the antecedent is a verbal node (a clause). In such a case, the potentially connective phrase can also be easily substituted with a classical connective (like *and*, *but*, *or*) with no (not even a small) shift in the meaning. The following two examples demonstrate two different ways a *preposition + demonstrative pronoun* structure takes part in creating discourse coherence:

(1) *Mövenpick provozuje několik desítek hotelů nejen v Evropě, ale i v Asii a Africe. **Kromě toho** je známý i jako obchodní a potravinářská firma.*

*Mövenpick operates dozens of hotels not only in Europe but also in Asia and Africa. **In addition**, (lit. *except that*) it is known also as a business and food company.*

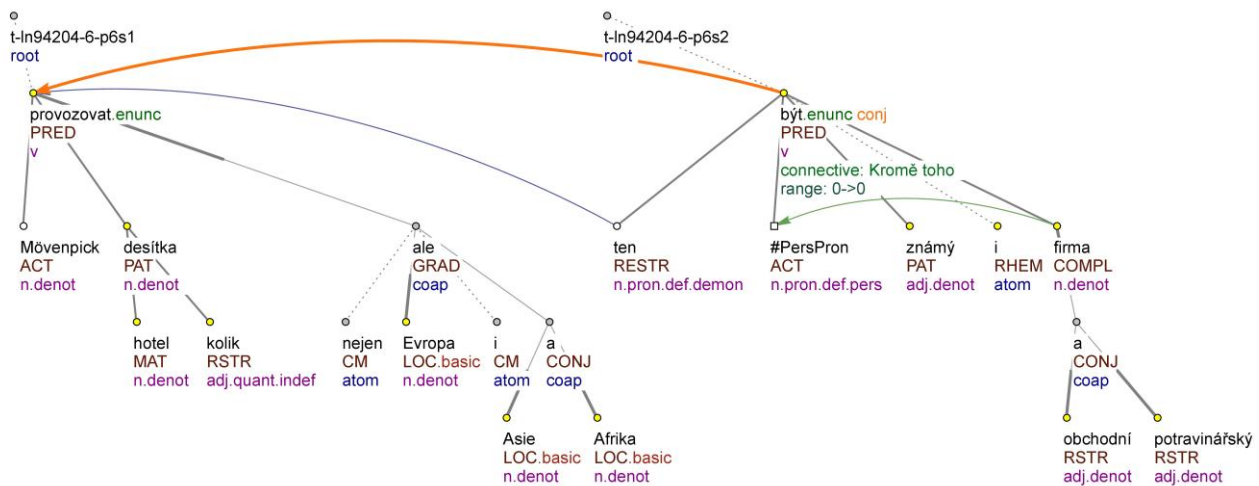


Figure 1: Example of a coreference link (blue arrow) that leads to a VP (*provozovat* = *operate*) – a discourse usage of the *preposition + demonstrative pronoun* construction. The discourse relation (conjunction) is marked with the orange arrow.

(2) *British Library vydala stručný katalog knih uvedené tematiky čítající přes šest set položek z majetku knihovny. K tomu lze na místě zakoupit dvě publikace o ruské avantgardní knize, vydané specialistkou Susan Comptonovou.*

The British Library has released a brief catalog of topic-related books containing over six hundred items from the library property. Along with that (lit. with that) you can purchase on-site two publications about the Russian avant-garde literature, published by the specialist Susan Compton.

In the example (1) in the PDT the coreferential link from the token *toho* (*that*), lemma = *ten*, leads to the node with the lemma *provozovat* (*operate*), which is a verb, see Figure 1. The event of operating dozens of hotels is connected with the event of being known for other activities. The structure *kromě toho* (lit. *except that, besides that*) can be easily substituted with a simple connective *navíc* (*moreover*) without a shift in the meaning. On the contrary in example (2), the antecedent for *tomu*¹ (*that*), lemma = *ten*, is a noun, the *catalog*. The sentence says that one can purchase several items, the catalog and other publications (together with the catalog). A substitution with a simple connective would cause a loss of the meaning “together with the catalog”. Rather, in this case, it can be substituted with a personal pronoun – *k němu* (*with it*). That implies unambiguously the non-connective reading, see Figure 2.

The presented analysis concerned in both cases the pattern *Prep_Dem* (compound forms). As for coreference regarding the first group, i.e. the type *PrepDem*, and the *PrepDem* forms in the second group (bound forms, see Section 3.1), the original reference carried by the

¹ *Toho* and *tomu* are two word forms of one Czech demonstrative pronoun – *to*, in their respective case. Other possible forms are *tom* and *tím*.

demonstrative pronoun is almost completely lost and Czech speakers, we dare to claim, do not normally feel the original structure of these expressions. The absence of coreference links in PDT in these cases reflects this fact. As a result, we only analyse coreferential links in the cases of compound forms. The bound forms are a part of the present analysis for the sake of completeness, as was mentioned above.

The analysis was tested on the data. For all patterns *preposition + demonstrative pronoun* (fulfilling some practical conditions described further in the Section 4), the nature of the antecedents, mainly their PoS value, was observed and also the substitution test was performed. The results confirm our hypothesis. Many of the seemingly random prepositional phrases with a demonstrative have a coreferential link to a verbal node. Moreover, but this is just an observed tendency, these PPs often occur sentence-initially (or, in a compound sentence, clause-initially), as DCs normally do. These findings support the fact that in these cases coreference plays a substantial role in interpreting language expressions as text-structuring means. In the distribution tables in the Section 5, the exact proportions of functions of the relevant PPs are visible.

4. The Method

All phrases/expressions with the structure *preposition + demonstrative pronoun* “*to*” (*that/this*) were searched for within all data of the PDT 2.0 available for linguistic research (43 955 sentences). From the total amount of these expressions, those with the structure *preposition + demonstrative pronoun + hypotactic conjunction* were excluded (e.g. *přestože* (*although*, lit. *despite of that, that*), *proto, že* (*because*, lit. *for that, that*), *vzhledem k tomu, že* (*because*, lit. *due to that, that*), *místo toho, aby* (*instead of*, lit. *instead of that, that*)) because they represent a specific subgroup and they deserve to be studied separately.

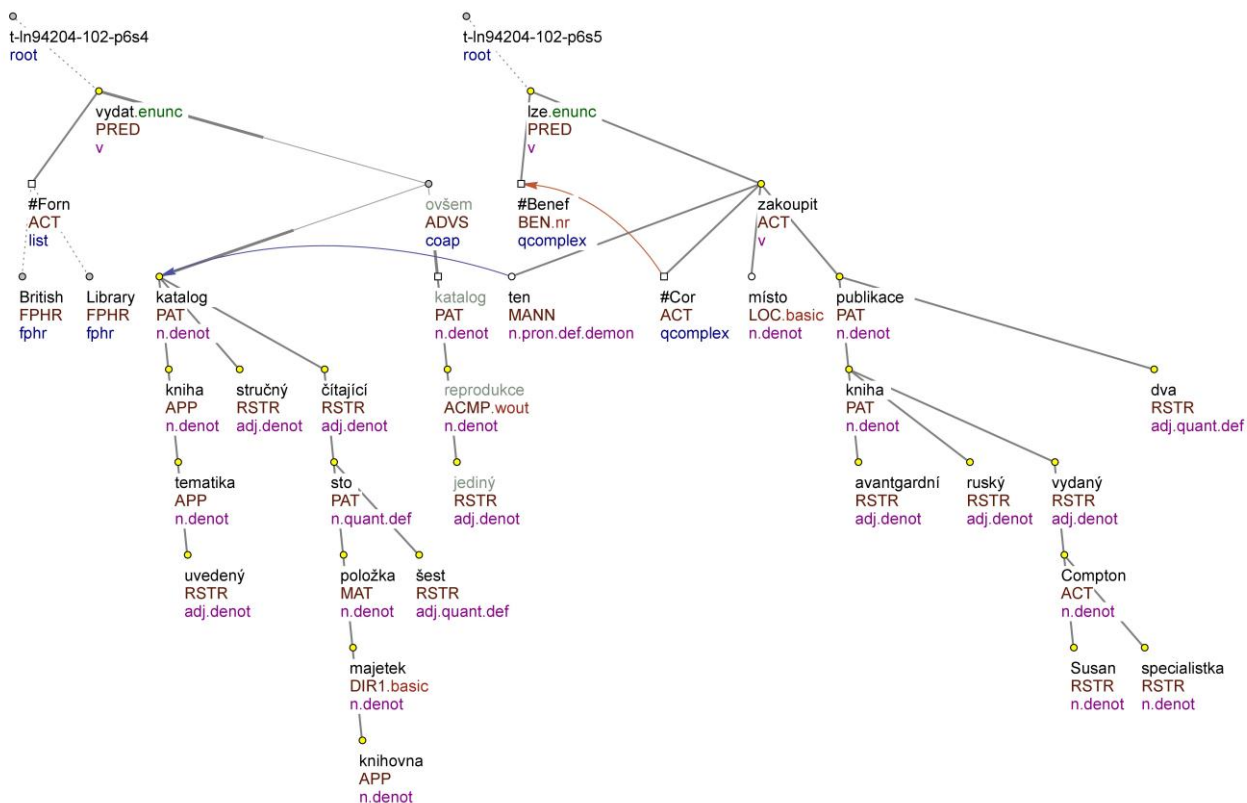


Figure 2: Example of a coreference link (blue arrow) that leads to an NP (head = *katalog*) – a non-discourse usage of the *preposition + demonstrative pronoun* construction

Also, as well as the English demonstrative pronoun *that*, the Czech demonstrative pronoun *to* (*that/this*) can be a head of a nominal phrase or it can serve as an adjective-like modifier of the noun (e.g. *pro to nové auto* (*for that new car*)). Only the first type of occurrences is relevant for our study and only these occurrences are listed in all tables.

According to the topic of this study, only those phrases/expressions that proved to be able to have a connective function are listed in the tables.

5. Distributions through the PDT 2.0

The results are summed up in three tables – each table for one group of the possible connectives according to the classification mentioned in Section 3.1. For the most frequent and/or linguistically most interesting instances, a more detailed characteristic follows. The first column of each table shows the total number of occurrences of each form/phrase in the PDT 2.0. The second column indicates the number of instances that function as DCs. For all the compound forms (written as two units) this means the number occurrences that have an anaphoric link to a verbal antecedent and that are at the same time replaceable by one of the basic connectives². Semantic types of discourse relations anchored by these connectives are given in the third column.³ For all

² And, just to complete, the bound forms (written as one unit) never have an anaphoric link.

³ For this task, the sense hierarchy developed for the discourse

compound forms/phrases (separately written) there is also the fourth column with the number of occurrences where the form has a link to an NP antecedent only (cf. example 2).

In all three tables, there is a slight divergence in numbers in the second and the third column. The difference between the total number of occurrences and those of a DC function follows from two facts: i) like any other DC, these expressions can also connect other units than verbal phrases (as e.g. in *kratší ale namáhavější cesta* = *shorter but more tiring way*). They are connectives but with no discourse function in the sense presented above (see Section 3)) and ii) some expressions (e.g. *předtím* (lit. *before that*)) are polysemous: in one meaning they function as discourse connectives (e.g. *předtím* in the meaning *before that*), in another they have other functions (e.g. *předtím* in the meaning *earlier*).

5.1 Group I. – Type *PrepDem*

The first group of possible discourse connectives with a referential component is represented by expressions

annotation is used (Mladová et al. 2009). We differentiate four major categories: temporal, contingency, contrast (comparison) and expansion relations. Their subtypes are: asynchronous, synchronous (temporal group); reason, pragmatic reason, condition, pragmatic condition, purpose, explication (contingency group); confrontation, restrictive opposition, concession, correction, opposition, pragmatic opposition, gradation (contrast group); conjunction, instantiation, specification, equivalence, generalization, conjunctive alternative, disjunctive alternative (expansion group).

always written as one unit.⁴ Example (3) demonstrates a connective use of one of these expressions – *proto* (*therefore*).

(3) *Výkonný výbor ČSSD nebyl na svém víkendovém zasedání usnášeníschopný, proto nemohl přijmout usnesení ke komunálním volbám.*

CSSD Executive Committee was not quorate at its weekend meeting therefore it could not adopt a resolution for the municipal elections.

When these expressions are written as two units, the meaning is strictly that of a regular PP and the connective function is never present (compare the example (4)).

(4) *Abychom všechny vzácné látky z medu uchovali, musíme med vhodně skladovat. Nejvhodnější jsou pro to skleněné nádoby a suché, chladné a tmavé místo při teplotě do 15 stupňů.*

To preserve all the precious substances of honey, honey must be stored properly. Glass jars and a dry, cool and dark place below 15 degrees are most appropriate for it.

One expression from this group is polysemous – *zatím* in the meaning *up to now/so far* cannot serve as a DC but in the meaning *meanwhile* and *although* it can. It should be added that the first meaning has an absolute majority in our data.

	total	DC function	types of relation (+ frequency)
zato (whereas, lit. behind that, for that)	41	38	14 confrontation 21 opposition 2 correction 1 gradation
nadto (moreover, lit. over that)	3	3	3 gradation
proto (therefore, lit. for that)	426	426	405 reason 15 explication 3 pragm. reason 2 correction 1 generalization
zatím (so far, meanwhile, although, lit. behind that)	410	1	1 concession

Table 1: Possible discourse connectives with a referential component always written as one unit

⁴ For the sake of completeness it should be added, that besides all connectives cited in the Table 1, there is also the expression *nato* (*after that*) with the bound structure *preposition + demonstrative pronoun* in Czech. Although this expression can in principal function as a discourse connective, it is in our data always connected with an expression of time, e.g. *krátce nato* (*shortly after that*). It was therefore decided to consider these phrases as temporal modifiers, not discourse connectives in our annotation.

As already mentioned above, the reference carried by the original demonstrative pronoun is lost and Czech speakers do not normally feel the original structure of these expressions. Moreover, the expression *proto* (*therefore*) is a basic discourse connective for one of the most important semantic relations in a text – reason; *zato* (*whereas*) is one of the typical connectives for contrastive relations and *zatím* one of the basic time adverbs in Czech. In other words – they are not alternatives to some basic expressions with the same function, but they are basic expressions for some meanings themselves.

5.2 Group II. – Type *PrepDem/Prep_Dem*

The second group comprises possible discourse connectives with a referential component that can be in Czech written both as one or two units⁵ (but not quite arbitrarily so, the context is important). The persistence of the connective function in both forms is illustrated by the examples (5) and (6). Accidentally, we were able to find here a very similar context with precisely the same type of relation anchored by a bound and a compound form of the connective.

(5) *Jelcin označil za klíčový problém přípravy ústavy postavení subjektů federace a vyjádřil přitom přesvědčení, že návrh nového základního zákona obsahuje maximální možné řešení této otázky.*

Yeltsin called a status of subjects of federation a key issue in preparing the constitution and at the same time, (lit. during that) he expressed the conviction that the proposal for a new basic law contains the maximum possible resolution of this issue.

(6) *Tento jev (...) označil za normální a poukázal při tom na zdouhavé zavádění kompaktních disků na začátku osmdesátých let.*

He described this phenomenon as normal and at the same time, he pointed to the lengthy introduction of compact discs in the early eighties.

The differences between the total number of occurrences and the occurrences as DC follows from the same facts as in Table 1, polysemy plays a significant role in case of *předtím*, *mezitím* and *při tom*.

According to the data cited in Table 2, it is obvious that all these expressions occur in the PDT more frequently in a bound form than in a compound form. Some compound forms even do not occur in the PDT at all (they are well documented in the literature and also in the larger Czech

⁵ To see the expressions in Group II as a whole, in Table 2 we listed not only forms documented in PDT, but all possible forms of these expressions. All these forms can be found in literature (Mluvnice češtiny 2 1986, Mluvnice češtiny 3 1987, and in Czech National Corpus (ČNK, SYN 2012).

National Corpus⁶ (ČNK, SYN 2012). The fourth column of the Table 2 also shows that among all found compound forms there was no context in which the referential component would refer to an NP (differences between total number of occurrences and those with a DC function follow from the polysemy of the forms).

	total	DC function	types of relation (+ frequency)	non-DC function
přítom (in the same time, and, but, lit. during that)	230	192	76 conjunction 45 opposition 40 concession 8 specification 7 gradation 7 confrontation 3 synchronous 2 restr. opp. 2 pragm. opp. 1 correction 1 reason	X
při tom	11	7	4 synchronous 2 conjunction 1 gradation	0
potom (then, lit. after that)	72	55	42 precedence 6 condition 5 conjunction, 1 gener. 1 restr. opp.	
po tom	0	0	0	0
předtím (earlier, before, lit. before that)	49	14	14 precedence	X
před tím	0	0	0	0
mezitím (meantime, lit. between that)	30	9	7 synchronous 1 precedence 1 confrontation	X
mezi tím	1	1	1 synchronous	0
mimoto (besides, lit. beside that)	3	3	3 conjunction	X
mimo to	2	2	2 conjunction	0
přesto (nevertheless, lit. despite of that)	139	132	109 concession 20 opposition 1 restr. opp. 2 pragm. opp.	X
přes to	0	0	0	0

Table 2: Possible discourse connectives with a referential component written both as one unit and two units

Another fact deserving a commentary is the original temporal meaning of majority of these expressions (*potom*, *po tom*, *předtím*, *před tím*, *přítom*, *při tom*, *mezi tím*, *mezitím*). With the exception of the expression

⁶ Consider for example the context from the Czech National Corpus which instances the usage of compound form of the connective for concessive meaning (*přesto* (*nevertheless*)):

Pietro temně zařval: měl kouli v prsou. Přes to skočil po třetím...
Pietro yelled darkly: he had a bullet in his chest. Despite that, he jumped after the third man.

přítom, this meaning is almost always preserved in the usage of these expressions. Their other meanings result from the fact that temporal relations sometimes combine with relations from the contrast group or with conjunction, and these meanings are in some contexts stronger. The expression *přítom* loses its original temporal meaning in some contexts in favor to the conjunctive or adversative meaning. The compound form *při tom* (lit. *during that*) tends to some extent to the temporal meaning (synchronous). However, our data are too limited for reliable meaning comparison of bound form (*přítom*) and compound form (*při tom*).

5.3 Group III. – Type *Prep_Dem*

The third table is devoted to possible DCs with a referential component that are always written as two units. As for number of its members, this group is the biggest one and it partly represents the group of expressions which would be called Altlex (alternative lexicalizations) in the approach of the Penn Discourse Treebank (Prasad, R. et al., 2010). With the exception of *místo toho* (*instead of that*), none of these expressions is the basic connective for any type of semantic relation. They arise possibly as results of the effort to refine the meanings of the basic connectives. In the contexts, where they are to be seen as DCs they can be easily replaced by some basic connectives. Consider e. g. the sentences under (7) and again the example (1). They illustrate the connective function of the phrases *kromě toho* (*besides, in addition*, lit. *besides that*), *k tomu* (*besides, lit. to it*). In both of the examples, these phrases can be easily replaced by basic connective *and*.

(7) *Obsluhuje střídavě dvě ordinace, k tomu vykonává posudkovou a návštěvní službu.*

He serves alternately in two surgeries, besides that, he performs an advisory and guest service.

A discrepancy in numbers in the second and the third column follows from the fact mentioned above. According to the data cited in the Table 3, except for the phrase *k tomu*, all listed expressions can be thought of as DCs with a considerable certainty. However, this fact may be closely dependent on the type of texts. As for the PDT, it may show a typical style of journalistic commentary text. Likewise, finding only a few examples where the referential element of the phrase refers to some other instance than a verb is in our opinion also related to text genre. Some of the structures from the Table 3 have also the typical local meaning (for example – *položil na stůl pero a papír a vedle toho sklenici. He put a pen and a piece of paper on the table and a glass next to it*), which was not documented in the PDT, possibly because of the genre of the treebank texts.

	total	DC function	types of relation + frequency	non-DC function
k tomu (in addition, lit. to that)	131	13	9 conjunction 3 gradation 1 reason	6
kromě toho (besides, in addition, lit. besides that)	39	36	31 conj. 5 gradation	3
naproti tomu (on the contrary, lit. opposite that)	21	21	20 confr. 1 opposition	0
oproti tomu (on the contrary, lit. opposite that)	4	4	4 confr.	0
vedle toho (in addition, lit. next to that)	2	2	2 conjunction	0
vzhledem k tomu (therefore, lit. due to that)	1	1	1 reason	0
místo toho (instead, lit. instead of that)	9	8	5 correction 1 opposition 1 pragm. opp. 1 confr.	1
navzdory tomu (nevertheless, lit. despite that)	0	0	0	0
spolu s tím (and also, lit. along with that)	1	1	1 conjunction	0
díky tomu (because, lit. thanks to that)	7	6	6 reason	1

Table 3: Possible discourse connectives with a referential component always written as two units

6. Conclusion

In this contribution, we discussed the potential function of the pattern *preposition + demonstrative pronoun* “to” (*that/this*) in Czech as a connective means in discourse structure. Summing up our analysis, we found out that the majority of the bound forms (written as one unit, e.g. *proto* (*therefore*), *potom* (*then*)) functions as a discourse connective of the basic kind, or/and it represents some of the basic temporal modifiers in a sentence. Furthermore, regarding the compound forms (written as two units, e.g. *kromě toho* (*besides that*)), a surprisingly high percentage of them has the function of a discourse connective (traditional grammars would treat all divided forms as regular syntactic parts of sentences). This is

distinguishable due to the referential link to the verbal antecedent and due to the substitution test (by a basic DC or, on the other hand, by a personal pronoun). Moreover, these PPs stand preferably in a sentence-initial (or clause-initial) position, which is also typical for DCs. However, our data is limited in its size and also in the character of the genre of the treebank texts. It would be suitable to check our results on bigger corpora.

Our findings should demonstrate that there are explicit interplays of different ways of discourse analysis. Our results show that coreference can play a substantial role in interpreting language expressions as text-structuring means. This can be useful for automatic analysis of text – with enough training data with coreference annotation available, one could automatically predict some of the occurrences of discourse connectives.

In addition, we believe that our analysis can offer some interesting observations and may be also useful for a comparative study of these phenomena in different languages, since we are aware of similar processes of development of connective means from pronominals (and the role of anaphora) in other languages.

7. Acknowledgements

We gratefully acknowledge support from the Grant Agency of the Czech Republic (grants P406/12/0658 and P406/2010/0875). This work has been using language resources developed and/or stored and/or distributed by the LINDAT-Clarin project of the Ministry of Education of the Czech Republic (project LM2010013).

8. References

- Český národní korpus - SYN. Ústav Českého národního korpusu FF UK, Praha. Cit. 12.03.2012, available at WWW:<<http://www.korpus.cz>>.
- Carlson, L., Marcu, D. (2001). Discourse Tagging Reference Manual. University of Southern California Information Sciences Institute. Available at: <<http://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf>>.
- Grosz, B. J.; Joshi, A. K. and Weinstein, S.. (1995). Centering: A Framework for Modeling the Local Coherence of Discourse. In *Computational Linguistics*, 21(2), pp. 203–225.
- Hajič, J.; Hajičová, E.; Panevová, J.; Sgall, P.; Štěpánek, J.; Havelka, J. and Mikulová, M. (2006). *Prague Dependency Treebank 2.0*. Philadelphia: Linguistic Data Consortium.
- Heine, B., Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Mikulová, M.; Bémová, A.; Hajič, J.; Hajičová, E.; Havelka, J.; Kolářová-Řezníčková, V.; Kučová, L.; Lopatková, M.; Pajas, P.; Panevová, J.; Razimová, M.; Sgall, P., Štěpánek, J.; Uřešová, Z.; Veselá, K. and Žabokrtský, Z. (2005). *Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka*. Praha: UFAL MFF. Available at:

- <<http://ufal.mff.cuni.cz/pdt2.0/doc/manuals/cz/t-layer/html/index.html>>.
- Mladová, L., Zikánová, Š. and Hajičová, E. (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, current version available at: <http://ufal.mff.cuni.cz/discourse/>
- Mladová, L., Zikánová, Š., Bedřichová, Z. and Hajičová E. (2009). *Towards a Discourse Corpus of Czech*. In Proceedings of the fifth Corpus Linguistics Conference (CL 2009), Liverpool, Velká Británie, available at <http://ucrel.lancs.ac.uk/publications/cl2009/#papers>
- Nedoluzhko A.; Mirovský, J.; Hajičová, E.; Pergler, J. and Ocelák, R. (2011). *Extended Textual Coreference and Bridging Relations in PDT 2.0*. Data/software, ÚFAL MFF UK, Prague, Czech Republic. Available at: <<https://ufal-point.mff.cuni.cz/xmlui/handle/11858/00-097C-0000-0005-BCCF-3>>
- Petr, J. (ed.). (1986). *Mluvnice češtiny*. [Díl] 2, Tvarosloví. Praha: Academia.
- Petr, J. (ed.). (1987). *Mluvnice češtiny*. [Díl] 3, Skladba. Praha: Academia.
- Prasad, R; Dinesh, N.; Lee, A.; Miltsakaki, E.; Robaldo, L.; Joshi, A. and Webber, B.. (2007). *The Penn Discourse TreeBank 2.0 Annotation Manual*. Available at: <<http://www.seas.upenn.edu/~pdtb/PDTBAPI/pdtb-annotation-manual.pdf>>.
- Prasad, R.; Dinesh, N.; Lee, A.; Miltsakaki, E.; Robaldo, L.; Joshi, A. and Webber, B. (2008). The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, CD-ROM.
- Prasad, R.; Joshi, A. and Webber, B. (2010). Realization of Discourse Relation by Other Means: Alternative Lexicalizations. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, Beijing, China, pp. 1023–1031.
- Sgall, P.; Hajičová, E. and Panevová, J. (1986). *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*, Dordrecht: Reidel Publishing Company; Praha: Academia
- Stede, M. (2007). *Korpusgestützte Textanalyse: Grundzüge der Ebenen-orientierten Textlinguistik*. Gunter Narr Verlag Tübingen.
- Wiebe, J.; Wilson, T.; Bruce, R.; Bell, M. and Martin, M. (2004). Learning Subjective Language. *Computational Linguistics*, 30(3), pp. 277–308.