The Centre and Periphery of Discourse Connectives

Magdaléna Rysová
Charles University in Prague
Faculty of Arts
magdalena.rysova@post.cz

Kateřina Rysová
Charles University in Prague
Faculty of Mathematics and Physics
rysova@ufal.mff.cuni.cz

Abstract

The paper tries to contribute to the general definition of discourse connectives. It examines connectives in broader sense, i.e. all language expressions that have an ability to express discourse relations within a text (e.g. both conjunctions like *but*, *and*, *or* and expressions like *the condition for this is*, *due to this situation* etc.). The paper tries to classify connectives from different perspectives and to divide them into several groups to specify their similarities and differences. We try to discuss various attributes an expression must have to be a connective. We understand discourse connectives as a set of expressions with a center and periphery and we focus here mainly on the periphery – i.e. on description of the secondary connectives (like *the reason is simple*, *this means that*... etc.) because it is not much investigated but a very current theme of discourse analysis.

1 Introduction

Discourse connectives are generally understood as explicit indicators of discourse relations within a text. However, there is not any shared and generally accepted definition of them. Therefore, various authors dealing with discourse studies try to give a list of connectives for the given language and to describe their common features.

In this paper, we want to contribute to this general discussion (as well as to the terminology issue) and to bring new perspectives from which we may look at discourse connectives. We also want to present general principles according to which we may draw boundaries among such a wide and heterogeneous group of expressions.

Our general observations are made on the basis of the large corpus study enabled by the annotated corpus Prague Dependency Treebank. Our research is carried out on Czech newspaper texts but we believe that our general principles may be used also for other languages.

2 Discourse Connectives – General Discussion

As said above, discourse connectives are hardly definable expressions, which is seen already in the fact that there are many different terms used for these expressions – cf. *discourse connectives* (Blakemore, 2002), *discourse operators* (Redeker, 1991), *discourse markers* (Schiffrin, 1987), *pragmatic connectives* (van Dijk, 1979) etc. We use the term *discourse connectives* following the Czech traditional terminology.

The variability in terminology points at the fact that discourse connectives are studied from different perspectives – e.g. from the syntactical, lexical, phonetic or pragmatic point of view. Since there is a chaos in terminology as well as in the definition of discourse connectives, we want to bring some new observations of this theme based on a large corpus study.

3 Discourse Connectives in the Prague Dependency Treebank

Our research on discourse connectives in Czech was carried out on the data of the Prague Dependency Treebank (PDT) – a manually annotated corpus of about 50 thousand sentences from newspaper texts containing, among others, annotation of discourse relations.

3.1 The First Annotation of Discourse Connectives in Czech

The first annotation of discourse relations in Czech was carried out in 2012. It was done on the data of the Prague Dependency Treebank 2.5 (Bejček et al., 2012) and was published independently as the

Prague Discourse Treebank 1.0 (Poláková et al., 2012).

The annotation was limited to explicit discourse connectives in narrow sense, i.e. connectives were understood only as expressions from selected parts of speech, especially conjunctions[1] (*ale* 'but', *nebo* 'or', *přesto* 'yet' etc.) and some types of particles (*jenom* 'only', *také* 'too' etc.). We will call these expressions **primary connectives**, as their primary function is to connect two units of a text and not to have some semantic role of a sentence element within the sentence.

Following the theory and terminology of the Pennsylvanian corpus Penn Discourse Treebank, the Prague Dependency Treebank understands connectives as expressions opening positions for two units of a text – in other words, connectives connect two textual pieces called arguments. During the first annotation of the primary connectives in PDT, there were annotated only such connectives whose arguments were verbal – i.e. represented mainly by two propositions or clauses – cf. an example from PDT:

(1) *Pro 600 zaměstnanců muselo nové vedení sehnat práci.*
 Proto se manažeři rozjeli za zakázkami nejen po republice, ale i do zahraničí.

'*The new leadership had to find a job for 600 employees.*
Therefore, the managers started to look for contracts not only around the country but also abroad.'

In the Example 1, there is a discourse connective *proto* '*therefore*' expressing a discourse relation of reason and result between two verbal (here propositional) arguments *the new leadership had to find...* and *the managers started to look...*

---

[1] There is often a discrepancy between the parts of speech like conjunctions, particles and adverbs. We define **conjunctions** (following the traditional Czech grammar) as synsemantic words with primary connecting function (like *but*, *or*, *therefore*, *however*, *and* etc.), structuring **particles** as synsemantic words expressing a relation of a speaker to the structure of a text (like *only*, *too* etc.) and **adverbs** as autosemantic words functioning as sentence elements expressing circumstances of events (like *subsequently*, *previously* etc.). Due to the often discrepancy of these parts of speech, the boundaries among connectives should not be stated strictly on the basis of the part-of-speech membership.

The first discourse annotation of the Prague Dependency Treebank includes both inter- and intra-sentential discourse relations and has been carried out partially manually and automatically[2]. The annotation of implicit discourse relations (i.e. without explicit connectives) and relations expressed by other means than primary connectives (e.g. by expressions like *that is the reason why*) has not been included here.

3.2 The Extended Annotation of Discourse Connectives in Czech

Apart from the annotation of primary connectives, we decided to annotate also discourse relations in Czech expressed by other means, i.e. by structures like *rozdílem bylo* 'the difference is', *to bylo způsobeno tím* 'this was caused by...', *jedinou podmínkou bylo* 'the only condition was' etc. – cf. an example from PDT where the expression *z tohoto důvodu* '*from this reason*' expresses a discourse relation of reason and result:

(2) *Jak vyplynulo z vyšetřování, oba muži si přepadení vymysleli.*
Z tohoto důvodu byli v těchto dnech z ČR vypovězeni.

'*The investigation revealed that the two men have lied about the attack.*
From this reason, they were expelled from the Czech Republic these days.'

The group of these connective structures is very wide and heterogeneous.

1) One subgroup of them are **open collocations** (grammatically free) containing mainly nouns (*příčina* 'cause', *důvod* 'reason', *podmínka* 'condition' etc.), verbs (*odůvodnit* 'to give reasons', *vysvětlit* 'to explain', *znamenat* 'to mean' etc.) and secondary prepositions[3] (*díky* '*thanks to*', *vzhledem k* '*with respect to*' etc.). Moreover, the individual connective "key words" occur in different structures – cf. the word *příčina* '*cause*' form structures like *příčinou bylo...* '*the cause was...*', *vidět příčinu v tom...* '*to see the cause*

---

[2] The automatic annotation has been checked by human annotators.
[3] The term *secondary prepositions* is used for prepositions that arose from another part of speech originally.

in...', *hledat příčinu v tom...* 'to seek the cause in...'.

2) Other types of these connective structures are **fixed phrases** (both grammatically and lexically restricted) that are fully frozen (like *o to více* 'what's more') or that enable only a slight modification – cf. *stručně/jednoduše/prostě řečeno* 'shortly/simply/generally speaking' etc. (more details to characteristics of these structures in Rysová, 2012).

As we can see, there is a wide range of structures that have a connecting discourse function within a text. Since they are not connectives from their nature (as conjunctions or structuring particles), but only in the form of certain collocations (whether free or fixed), we use for all of these connective structures a term **secondary connectives** (some authors use the term alternative lexicalizations of discourse connectives, shortly AltLexes – cf. Prasad et al., 2010).

This extended discourse annotation of secondary connectives in the Prague Dependency Treebank is manual, but the detection of some structures was done automatically (cf. Rysová and Mírovský, 2014). The annotation contains both inter- and intra-sentential discourse relations.

The aim of the next part is to compare and contrast these two annotations (i.e. of primary and secondary connectives) and then to draw some general observations that resulted from the practical data annotations.

4   Results and Evaluation

As said in section 3.1, the annotation of primary connectives contains only discourse relation between two verbal arguments (i.e. represented mostly by two propositions or clauses). Altogether, the Prague Dependency Treebank contains 20,255 of such expressions (measured on whole data).

The annotation of secondary connectives has been finished right now and we bring the first complex results of it (although we are aware that the numbers of tokens may slightly change, as the data are now being checked and corrected).

When preparing the annotation principles, we realized that it is not possible to strictly follow the principles stated for primary connectives. Secondary connectives form a very heterogeneous group of connective structures that behave differently than primary connectives in some cases. Therefore, we could adopt only a part of the old principles and we had to create some new for the specific structures. We will now present all these principles, i.e. which both overlap and differ from the principles for primary connectives.

Some of the secondary connectives are fully replaceable by the primary ones, as they may also connect two units of a text realized by verbal arguments – see Example 2 where the structure *z tohoto důvodu* 'from this reason' expresses a discourse relation between two propositions. It is here replaceable by the primary connective *proto* 'therefore' and the meaning remains practically the same. Altogether, the Prague Dependency Treebank contains 924 of such types of relations.

During the data annotation, we found also such secondary connectives that allow nominalization of the second argument; in other words, they are followed not by a verbal clause but by a nominal phrase. See Example 3 from PDT:

(3) *Privatizované mlékárny se však zatím mezi sebou nedokázaly domluvit.*
Důsledkem je nekompromisní konkurenční <u>boj</u>, který tlačí ceny výrobků až takřka k nulové rentabilitě zpracovatelů.

'The privatized dairies have so far failed to agree among themselves.
<u>The consequence is</u> a rigorous <u>competition</u> that pushes up the price of products to almost zero profitability of processors.'

In this example, the connective structure *důsledkem je* 'the consequence is' is followed by a nominal phrase *boj* 'competition' (not by its verbal representation *that they rigorously compete*). These connectives have mostly a similar structure – *důvodem je* 'the reason is', *důsledkem je* 'the consequence is', *příčinou je* 'the cause is', *podmínkou je* 'the condition is' etc.

The difference between *the consequence is a rigorous competition* and *the consequence is that they rigorously compete* is only syntactic, not semantic so we decided to include these cases into our annotation. (However, we distinguish between annotation of verbal and nominal arguments technically, so they may be automatically detected for possible further investigation). It appeared that

nominalization of the second argument is a feature of written rather than spoken language.

This is the first case when the annotations of primary and secondary connectives differ, i.e. in the nature of discourse arguments. However, the restriction on verbal arguments for primary connectives is clear, as we cannot say, for example, *a proto boj '*therefore the competition'. In this respect, it is clearly visible the heterogeneity of secondary connectives and their bigger flexibility.

Altogether, the Prague Dependency Treebank contains 237 of these relations, i.e. relations expressed by secondary connectives followed by a nominalized argument.

4.1 The Universality Principle

During the annotation, we observed also another interesting phenomenon. In the data, the individual connective key words (like *reason*, *due to*, *because of*, *condition* etc.) occurred in different structures with respect to their connective status. We saw a difference between combinations like *kvůli tomu* '*because of this*', *kvůli této skutečnosti* '*because of this situation*', *kvůli tomuto nárůstu* '*because of this increase*' or *kvůli jejich pomoci* '*because of their help*'. All of these combinations containing the preposition *kvůli* '*because of*' refer to the preceding context and express a discourse relation of reason and result. However, we feel that there is a difference between them concerning the fact whether the structure is context dependent (as, for example, *because of this increase*) or not (as *because of this*). In other words, *because of this increase* may be used only in a limited set of contexts (in texts about *increasing*), but *because of this* is context independent – *this* is a deictic word so it may be embedded to any context and it will find there its semantic relations. Other words like *increase* or *help* do not have this ability, i.e. to adapt their meaning to context. Therefore, we decided to annotate these structures differently according to this contextual in/dependency that we called **universality principle**.

The universality principle evaluates connective structures from the fact whether they have a universal status of connectives, i.e. whether they function as indicators of certain discourse relation universally or occasionally. In other words, we tried to answer – if we have several different contexts with, e.g., the relation of reason and result

– whether the given connective structure (with an ability to express this type of relation) fits into each of them (and is therefore universal) or not.

In this respect, we evaluated *kvůli tomu* '*because of this*' as a universal secondary connective whereas *kvůli tomuto zvýšení* '*because of this increase*' as a non-universal connecting phrase.

We decided to state the boundary of connectives right here, i.e. according to the universal or non-universal status of expressions. In this respect, connectives function as universal indicators of given discourse relations (whether primary /like *and*, *but*, *or*/ or secondary /like *that is the reason why*, *due to this*, *the condition of this is*, *in spite of this* etc./). We decided that we will not include the non-universal structures in discourse connectives – because even though they express certain discourse relation, they contain too much other lexical items occurring in the connective structure only occasionally. Therefore, we will call these structures (like *because of this increase*, *the reason of his late arrival is* etc.) non-universal connecting phrases, not connectives.

In the Prague Dependency Treebank, we annotated 79 non-universal connecting phrases between two verbal arguments and 72 non-universal connecting phrases followed by nominal arguments. See the Table 1 depicting the annotation of secondary connectives and other connecting phrases in the Prague Dependency Treebank (where the abbreviation VP means verbal phrase and NP nominal phrase).

	VP_VP	VP_NP	TOTAL
Universal Secondary Connectives	924	237	**1,161**
Non-universal Connecting Phrases	79	72	**151**
TOTAL	1,003	309	**1,312**

Table 1: Extended Discourse Annotation in PDT

The Table 1 demonstrates that PDT contains 1,161 occurrences of universal secondary connectives (i.e. 88 % within the total number) and 151 occurrences of non-universal connecting phrases (i.e. 12 %). So there are obvious fixing tendencies concerning the form and gaining a universal status of connectives.

5 General Observations – the Centre and Periphery of Discourse Connectives

In current stage, the discourse annotation in the Prague Dependency Treebank contains altogether 21,416 of discourse relations – 20,255 expressed by primary connectives and 1,161 by secondary connectives. We may see that relations of secondary connectives form 5 % of the total number. Therefore, the primary connectives may be viewed as a centre of all connecting expressions and the secondary as its periphery.

5.1 Primary Connectives

Based on the large data annotation, we would like to contribute to the general discussion on discourse connectives, especially on their definition. Although our research has been carried out on Czech, we believe that our statements may be used also for other languages.

As we discussed above, we understand connectives as a large and heterogeneous group of expressions with its center and periphery. The center is formed by expressions we called primary connectives.

The primary connectives are synsemantic words that do not function as sentence elements (i.e. they do not play a role of a subject, object, adverbial etc.), they are mostly one-word expressions, lexically frozen. Therefore, they mostly do not allow modification (i.e. it is not possible to say, e.g., *generally and*, *simply but* etc. with some exceptions like *simply*/*mainly*/*generally because*).

As they do not affect the syntax of the sentence, the primary connectives may be also omitted without any syntactical changes in most cases – see Example 1 from PDT where the primary connective *therefore* may be simply omitted from the sentence without any changes and the discourse relation is maintained (i.e. remains implicit).

We use the term primary connectives also because of the frequency. As said in the section 5, the primary connectives occur in 95 % of all discourse relations expressed explicitly (i.e. by some language expression, not implicitly).

5.2 Secondary Connectives

Secondary connectives are much more heterogeneous group than primary connectives – concerning the part-of-speech perspective as well as the syntactic, semantic and lexical point of view.

Secondary connectives occur in the sentence mainly as structures with some basic or key word – these words are from different parts of speech, the most numerous are nouns (like *cause*, *reason*, *condition*, *explication*, *justification*, *exception*, *contrast*, *difference* etc.), verbs (like *to give reasons*, *to explain*, *to mean*, *to be related to*, *to specify*, *to continue*, *to contrast*, *to precede*, *to follow* etc.) and secondary prepositions (like *because of*, *due to*, *despite*, *except for* etc.) .

Secondary connectives are structures containing autosemantic words (in contrast to synsemantic conjunctions or particles as primary connectives) and they are integrated (as a whole) into clause structure as sentence elements (e.g. *because of this* as an adverbial of reason) or they function as clause modifiers (like e.g. *shortly speaking*).

Some secondary connectives function as one sentence element (like *due to this fact* as an adverbial of reason) or their individual parts have a role of a sentence element on their own (cf. the structures like *this is the cause* where *this* is a subject, *is* is a predicate, *cause* is an object etc.). This is one of the phenomena in which the secondary connectives differ from the primary ones to a large and significant extent.

All secondary connectives also contain (implicitly or explicitly) the reference to the previous context. In this respect, the secondary connectives may be divided into three groups – they 1) may express the reference in the surface (like *the result /of this/ is*); 2) must express the reference in the surface due to valency (*this means that*...) or 3) cannot express the reference in the surface (e.g. it is impossible to say *this generally speaking*, although *speaking* indicates implicitly that it was spoken about something in the preceding context). For more details, see Rysová (2012 and 2014).

Another very interesting thing is that some secondary connectives may be even syntactically higher than the second argument of the discourse relation. There are examples like *I cannot go for a*

trip tomorrow. <u>*The reason is that*</u> *I am ill* etc. In these structures, the second discourse argument *I am ill* (syntactically a nominal content subordinate clause) is dependent on the connective structure *the reason is* (syntactically the main clause). This is a phenomenon that can never occur to primary connectives – in case of the primary connective (*I cannot go for a trip tomorrow* <u>*because*</u> *I am ill*), the second argument would be syntactically dependent on the first one. This is one of the phenomena making secondary connectives unique and original structures among the whole class of discourse connectives.

Most of the secondary connectives (apart from the lexically restricted phrases) are modifiable. So we may say *the main/only/first/important... reason is*.

Some secondary connectives have even a form of a separate sentence like *The reason is simple/easy*. etc. – see Example 4 from PDT:

(4) *S vašimi akciemi se musí obchodovat na burze, ale Wall Street vám nabízí cenu z RMS.*
<u>*Důvod je vcelku jednoduchý*</u>.
V RMS je cena většiny akcií nižší než na burze.

'*You must trade with your shares on the stock market, but the Wall Street offers you a price of RMS*.
<u>*The reason is quite simple*</u>.
In RMS, the price of most stocks is lower than on the stock market.'

In this respect, the secondary connectives demonstrate another big difference from the primary ones – they may stay alone, outside the discourse arguments, i.e. outside the two units of a text they connect. So it is interesting that some secondary connectives show a big deal of independency, as they may form syntactically and semantically complete textual units.

The secondary connectives in the form of whole separate sentences may be replaceable by primary connectives, but only in some cases – when the suitable primary connective allows the same modification as appears in the connective sentence – cf. *The reason is simple*. may be substituted by the modified primary connective *simply because*. In case of more complex modification, the substitution is only partial – some of the lexical

meaning is lost. Cf. the PDT example *Další důvod je složitější a je v podstatě filozofický* 'Another reason is more complex and in essence philosophical' that cannot be fully replaced by the primary connective, as we cannot say **philosophically because*. Therefore, the substitution by primary connectives is very limited in these cases.

The secondary connectives form altogether 5 % of all discourse relations (expressed explicitly) in the Prague Dependency Treebank, so their frequency in the texts is much lower than in case of the primary connectives. This could be another reason why to call them secondary. On the other hand, although they seem to be peripheral as a whole group because of their lower frequency, they enrich the discourse by various structures and behaviour the primary connectives can never do. Therefore, due to their idiosyncrasy in behaviour, they occupy a special and unique place in discourse and the term secondary does not mean less important – we established the opposition of primary and secondary connectives mainly due to their peculiarities and different behaviour.

For the structured difference between the primary and secondary connectives see Table 2.

5.3 Permeability of Borderline between Primary and Secondary Connectives

Within the secondary connectives, we may observe several subclasses of expressions being closer or farther to primary ones. Some of the secondary connectives may even cross the borderline and become primary, as we will demonstrate in this section.

One large group of the secondary connectives are structures containing prepositions (like *because of*, *due to*, *in spite of*, *despite*, *except for* etc.) that obligatory combine with some anaphoric autosemantic words to become discourse connectives – cf. it is impossible to say **due to, I did it* but only *due to* <u>*this*</u>, *I did it*.

Primary Connectives	Secondary Connectives
synsemantics	structures with autosemantic words
lexically frozen (grammaticalized)	open or fixed collocations (non-grammaticalized)
non-modifiable (with exceptions)	modifiable (with exceptions)
mainly one-word	mainly multiword
universal	universal
not sentence elements	sentence elements, clause modifiers or separate sentences
	convey anaphoric reference to the 1st argument
	uniqueness of some structures:
	a) syntactically higher than the 2nd argument
	b) form of a separate sentence
	c) nominalization of the 2nd argument

Table 2: Characteristic of Primary and Secondary Connectives

The secondary prepositions may combine with various nouns and pronouns (cf. examples like *because of this*, *because of this situation*, *because of this increase* or *because of their help*). Some of these variants are context dependent (as *because of their help*), some are universal (*because of this*).

The universal connecting structures are prepositions in combination especially with an anaphoric pronoun *this* and are very close to the primary connectives.

Primary connectives are lexically frozen, grammaticalized expressions that have also not been primary connectives from their origin. We will demonstrate this on examples of primary connectives that historically consisted of two words and later became grammaticalized as one-word connectives. A typical example is the primary connective *therefore* that arose from the connection of *there* and a preposition *fore* (an Old English and Middle English collateral form of the preposition *for*) meaning *in consequence of that*.

The same historical process may be seen in case of the foreign counterparts of *therefore* – like Czech connective *proto* (a connection of the preposition *pro* '*for*' and the pronoun *to* '*this*'), Dutch *daarfoor*, German *dafür* or Danish *derfor*.[4] So we may see that this process is not language specific but that it happened similarly in more languages. Since this process is generally common in language, there is a possibility that it might occur again.

Therefore, today's similar combinations of prepositions and anaphoric pronouns like *due to this*, *despite this* (in Czech *kvůli tomu*, *navzdory tomu*) etc. functioning as universal secondary connectives might be grammaticalized as well and might cross the borderline toward the primary connectives in the future.

In this respect, we understand the borderline between primary and secondary connectives as being permeable, i.e. that some structures from the secondary connectives may undergo changes that would fix them to expressions with the primary connecting function.

6 Conclusion

In the paper, we introduced the annotations of discourse relations in the Prague Dependency Treebank, especially the annotation principles for expressions like *hlavní podmínkou je* '*the main condition is*', *to je důvod, proč* '*that is the reason why*' etc. signaling relations within a text.

On the large data analysis, we tried to contribute to the general discussion on discourse connectives and especially on their definition. We suggest a division of connectives on primary and secondary. Primary connectives are mainly one-word expressions, lexically frozen that are not integrated into the clause structure as sentence elements and whose primary function is to connect two pieces of a text. The primary connectives form 95 % of all explicitly expressed discourse relations in PDT and therefore we consider them the center of all connective expressions.

The secondary connectives function as connectives mainly in various structures or combinations, they may be integrated into clause structure as sentence elements (like *because of this*), function as sentence modifiers (like *simply*

---

[4] Other similar examples in English are, e.g., *thereafter*, *thereupon* etc.

speaking) or may even form a separate sentence (*the reason is simple*). All of them contain autosemantic words, most often nouns (*reason*, *cause*, *explanation*...), or verbs (*to explain*, *to result*, *to continue*...). The secondary connectives function as connectives universally (like *because of this*), which makes them closer to primary connectives (like *therefore*, *thereafter*). Other connecting structures are contextually dependent (like *because of this increase*). These non-universal phrases are on the very edge of the connecting elements and they have very little chance to be grammaticalized. Therefore, we do not count them among connectives.

Although the secondary connectives are not as frequent as the primary ones and in this respect, they could be viewed as the periphery within all connectives, they enriched the discourse annotation of Czech in PDT by 1,161 of new relations. Moreover, some of them behave differently than the primary connectives (e.g., they may form a separate sentence or stay syntactically higher than the second argument). Because of this idiosyncrasy, they have a unique place within other expressions structuring discourse.

Acknowledgments

References

Eduard Bejček et al. 2012. Prague Dependency Treebank 2.5 – a revisited version of PDT 2.0. In: *Proceedings of Coling 2012*, Bombay, India, pp. 231–246.

Diane Blakemore. 2002. *Relevance and Linguistic Meaning*. *The Semantics of Discourse Markers*. Cambridge, Cambridge University Press.

Teuen A. van Dijk. 1979. Pragmatic Connectives. In: *Journal of Pragmatics 3*. North-Holland Publishing Company, pp. 447–456.

Lucie Poláková et al. 2012. *Prague Discourse Treebank 1.0*. Data/software, ÚFAL MFF UK, Prague, Czech Republic.

Rashmi Prasad, Aravind Joshi, Bonnie Weber. 2010. Realization of Discourse Relations by Other Means: Alternative Lexicalizations. In: *Proceedings of Coling 2010*, Tsinghua University Press, Beijing, China, pp. 1023–1031.

Gisela Redeker. 1991. Linguistic markers of discourse structure. In: *Linguistics 29(6)*, pp. 1139-1172.

Magdaléna Rysová. 2012. Alternative Lexicalizations of Discourse Connectives in Czech. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, European Language Resources Association, Istanbul, Turkey, ISBN 978-2-9517408-7-7, pp. 2800–2807. WWW: http://lrec.elra.info/proceedings/lrec2012/pdf/420_Paper.pdf

Magdaléna Rysová. 2014. Verbs of Saying with a Textual Connecting Function in the Prague Discourse Treebank. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, European Language Resources Association, Reykjavik, Iceland, ISBN 978-2-9517408-8-4, pp. 930–935. WWW: http://www.lrec-conf.org/proceedings/lrec2014/pdf/79_Paper.pdf

Magdaléna Rysová, Jiří Mírovský. 2014. Use of Coreference in Automatic Searching for Multiword Discourse Markers in the Prague Dependency Treebank. In: *Proceedings of The 8th Linguistic Annotation Workshop (LAW-VIII)*. Dublin, Ireland, ISBN 978-1-941643-29-7, pp. 11–19. WWW: http://www.aclweb.org/anthology/W14-4902

Deborah Schiffrin. 1994. *Approaches to Discourse*. Malden (MA), Blackwell. ISNB 0-631-16623-8.