Explicit and Implicit Discourse Relations in the Prague Discourse Treebank

Šárka Zikánová$^{(\boxtimes)}$, Jiří Mírovský, and Pavlína Synková

Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics,
Charles University, Prague, Czechia
{zikanova,mirovsky,synkova}@ufal.mff.cuni.cz

Abstract. Coherence of a text is provided by various language means, including discourse connectives (coordinating and subordinating conjunctions, adverbs etc.). However, semantic relations between text segments can be deduced without an explicit discourse connective, too (the so called implicit discourse relations, cf. *He missed his train. 0 He had to take a taxi.*). In our paper, we introduce a corpus of Czech annotated for implicit discourse relations (*Enriched Discourse Annotation of Prague Discourse Treebank Subset 1.0*) and we analyze some of the factors influencing the explicitness/implicitness of discourse relations, such as the text genre, semantic type of the discourse relation and the presence of negation in discourse arguments.

Keywords: Implicit discourse relations · Text genre · Negation

1 Some Explicit Questions About Implicitness

Text coherence is often provided by various language means, such as information structure, anaphoric chains, bridging (associative) anaphora or discourse relations. In our analysis, we deal with discourse relations: they can be either signaled by a discourse connective (typically coordinating and subordinating conjunctions, discourse adverbs etc.) or there is no connective device signaling the relation (so called implicit discourse relations, cf. Example 1).

(1)   He did not come. **0** He was ill.

Our general research question is, how can we understand a text if some signals of coherence (discourse markers) are omitted? How do we deduce the meaning of such a relation between text segments? To get more insight into this general research question, we have split it into several subtopics that we describe in this paper. First, we want to know how often implicit discourse relations occur in Czech in general, since it may be just a peripheral phenomenon. Second, we want to see some conditions or correlations of implicitness and explicitness, connected with other language phenomena. Is the implicitness of a discourse relation influenced by text genre [12,13]? Are there any semantic types of discourse relations

more prone to be implicit than other [5]? Or, in other words, are there any semantic types of discourse relations which must be expressed explicitly, as they are not deducible from the context? And, last but not least, can the implicitness/explicitness of discourse relations be influenced by such phenomenon as a sentence negation?

2   Data

To formulate hypotheses about the questions raised in Sect. 1 and to test them reliably, we needed manually annotated data. We had at our disposal the already published annotation of discourse relations in the Prague Discourse Treebank 2.0 (PDiT 2.0; [11]). It is a corpus of 50 thousand sentences of Czech journalistic texts from the 1990's, manually annotated on morphological, surface syntax and deep syntax (tectogrammatical) layer. Additional annotations (performed on top of the tectogrammatical layer) include coreference, bridging anaphora and discourse relations.

The annotation of discourse relations in the PDiT was inspired by the annotation scenario of the Penn Discourse Treebank 2.0 [8], which follows a lexically-grounded approach [15]: A discourse connective is a lexical anchor of a discourse relation that holds between two text spans called discourse arguments. The connective signals the sense of the discourse relation (Table 1 gives a list of possible senses). See Example 2 for a temporal discourse relation of *synchrony* expressed by the connective *when*. If the connective is absent (like in Example 1), the relation is called "implicit".

(2)   **When** I was young, the winters were much colder.

However, annotation of discourse relations in the PDiT only covers explicit relations. To be able to study also other types of cohesive means (and having only limited resources), we have selected a subcorpus (approx. 5%) from the PDiT and enriched[1] the original annotation of explicit discourse relations by the annotation of implicit relations, entity-based relations, question–answer relations and other discourse-structuring phenomena (see Table 2 for overall numbers of the relations). Our aim was to mark all local connections between discourse arguments and to present a text as a continuous chain of discourse segments, with the following possible types of connections:

– explicit discourse relations expressed by primary discourse connectives (expressions such as *because, if, but* etc.)
– explicit discourse relations expressed by secondary discourse connectives (e.g. *the reason is that*)
– implicit discourse relations (without a discourse connective)
– entity-based relations (relations based on the coreferential connections between discourse arguments)

---

[1] Using an adapted environment for annotation of discourse relations on top of the deep syntax (tectogrammatical) layer implemented in tree editor TrEd [3,4].

Table 1. Semantic types of discourse relations in the PDiT-EDA

CONTRAST	EXPANSION
Confrontation	Conjunction
Opposition	Conjunctive alternative
Restrictive opposition	Disjunctive alternative
Pragmatic contrast	Instantiation
Concession	Specification
Correction	Equivalence
Gradation	Generalization
	Empty relation
CONTINGENCY	TEMPORAL
Reason–result	Synchrony
Pragmatic reason–result	Precedence–succession
Explication	
Condition	
Pragmatic condition	
Purpose	

– questions (question–answer relation and also a relation between the previous
  context and the question)
– lists (e.g. *First, ... Second, ...*)
– coherence gaps (no relation to the preceding context can be found)
– specific parts of a text (author, location, heading, caption, etc.)
– attribution (relation between the author speech and the reported speech)
– macrostructure (relation between large segments of the text related to the
  text as a whole)

The data was annotated by two annotators with an overlap for inter-annotator
agreement measurements. Newly we measured agreement on the implicit rela-
tions; for numbers of agreement on several other types of relations, see a paper
about the underlying PDiT corpus [7]. The agreement on the recognition of the
presence of an implicit relation was 0.54 (F1-measure). Agreement on discourse
types (senses) of implicit discourse relations recognized by both annotators was
57.7% (agreement ratio), with Cohen's κ 47.4%.[2]

   The enriched subcorpus contains 2 592 sentences in 100 documents and covers
15 genres (see Table 4 for the list of genres). It is available to download under
the Creative Commons license from the LINDAT/CLARIN repository as the
Enriched Discourse Annotation of PDiT Subset 1.0 (PDiT-EDA 1.0; [17]).

---

[2] For a comparison, a measurement of inter-annotator agreement for implicit relations
   in the Turkish Discourse Treebank reports chance-corrected κ values of 0.52 for the
   class level, 0.43 for the type level and 0.34 for the subtype level [16]. The measure-
   ment at the subtype level corresponds to our measurement of agreement on discourse
   types.

3   Results

The first question of our analysis is, what is the distribution of explicit and implicit discourse relations in our data. As presented in Table 2, the occurrence of explicit and implicit discourse relations is comparable, implicit discourse relations are slightly more common than explicit ones. (In the table, relations expressed with secondary discourse connectives, so called AltLexes, such as *the reason is,* are counted as a subset of explicit discourse relations.)

Let us compare these results with the annotation in the Penn Discourse Treebank 3.0 (PDTB 3.0; [10]), as presented in [14, p. 5]. In the PDTB 3.0, the amount of explicit and implicit discourse relations is comparable (25 865 explicit discourse relations, together with AltLexes, 21 731 implicit discourse relations), but in contrast with the PDiT-EDA, the explicit discourse relations are prevalent. The proportion of implicit discourse relations in the PDiT-EDA is comparable with the results of the Penn Discourse Treebank 2.0, too, where they present an amount of 40% of the whole set of annotated discourse relations [9]. The occurrence of implicit discourse relations in the Prague Discourse Treebank data is slightly higher (it must be accounted that the number for the implicit discourse relations covers inter-sentential as well as intra-sentential ones).

Table 2. Overall numbers of relations in the PDiT-EDA 1.0

Total on 2 592 sentences	3 149
Explicit discourse relations	1 288
Implicit discourse relations	1 427
Entity-based relations	264
Lists	105
Questions	65

To test the validity of the annotation of implicit discourse relations in the PDiT-EDA 1.0, a part of the data was annotated in parallel (12 documents containing 233 sentences). The inter-annotator agreement was measured according to the three following values: (a) agreement on the presence of an implicit discourse relation (the same pair of a starting and target nodes in the tree is connected with a discourse arrow by both annotators, disregarding its semantic label); (b) agreement on a semantic type of a discourse relation (in cases where both annotators agree on the presence of a discourse relation); (c) Cohen's κ for the second type of agreement.

In Table 3, the inter-annotator agreement on the annotation of implicit discourse relations in the PDiT-EDA 1.0 is compared to annotation of other kinds of text relations in the data, namely explicit discourse relations, textual coreference (i.e. mostly inter-sentential coreference in cases where the coreference cannot be directly deduced by grammatical rules), and bridging anaphora (such as *a room – the door* etc.), as they were annotated in the PDiT and reported in [7].

Table 3. Inter-annotator agreement in different types of the text annotation.

	Discourse implicit	Discourse explicit	Textual coreference	Bridging anaphora
F1 – presence of a relation	0.54	0.84	0.72	0.46
Agreement on types	0.58	0.77	0.90	0.92
Cohen's κ on types	0.47	0.71	0.73	0.89

3.1   Implicit Discourse Relations and Text Genres

Our data enables us to test the hypothesis that the proportion of implicitness and explicitness of discourse relations varies depending on text genre. We had supposed that in some text genres, the understanding is based rather on the recipient's genre-based expectations than on explicit lexical discourse signals. E.g., in a letter, an address and a date can be expected at the beginning of a text; a weather forecast provides the recipient with information about different aspects of the weather, structured as an additive chain without specific discourse markers. On the other hand, we assumed that narrative texts tend to express discourse relations rather explicitly, as they usually describe more complicated plots with unexpected nets of relations (additive, contrastive, temporal etc.).

Table 4 presents the distribution of implicit and explicit discourse relations among text genres in the PDiT-EDA. The relations between implicit and explicit discourse relations in different text genres were compared in Fig. 1, converted to percentages.

According to the Chi-square test, the differences in the distribution of implicit and explicit discourse relations among text genres are significant. Text genres with a typical prevalence of implicit discourse relations (weather, overview, invitation) describe simple events or more of them in a row, usually in a present or future tense, cf. Example 3 (overview):

(3)   (1) Jak na koncert Pink Floyd_HEADING

...

(9) Na pražské hlavní nádraží bude vypraven mimořádný rychlík z Bohumína (odjezd v 8.30 hod.), který zastavuje na hlavním nádraží v Ostravě, Ostravě-Svinově, Studénce, Suchdole nad Odrou, Hranicích na Moravě a Olomouci._IMPLICIT CONJUNCTION WITH (10)

(10) Další posilový rychlík (odjezd v 15.18 hod.) z Českých Budějovic zastavuje v Hluboké nad Vltavou-Zámostí, Veselí nad Lužnicí, Soběslavi, Táboře, Olbramovicích a Benešově u Prahy._IMPLICIT CONJUNCTION WITH (11)

(11) O další vozy budou rozšířeny pravidelné rychlíkové spoje Brno – Praha (odjezd ve 14.06 hod.) a Břeclav – Brno – Praha (odjezd ve 14.40 hod.)

(1) How to get to the Pink Floyd concert_HEADING

...

(9) A special express train from Bohumín (departure at 8.30 am) will be dispatched to Prague Main Station, which stops at the main railway station in Ostrava, Ostrava-Svinov, Studénka, Suchdol nad Odrou, Hranice na Moravě and Olomouc._IMPLICIT CONJUNCTION WITH (10)

(10) Another support train (departure at 15.18) from České Budějovice stops in Hluboká nad Vltavou-Zámostí, Veselí nad Lužnicí, Soběslav, Tábor, Olbramovice and Benešov u Prahy._IMPLICIT CONJUNCTION WITH (11)

(11) Regular express trains Brno – Prague (departure at 14.06) and Břeclav - Brno - Prague (departure at 2.40 pm) will be extended with additional wagons.

Table 4. Numbers of occurrences of implicit and explicit discourse relations among text genres in the PDiT-EDA.

Genre	Implicit	Explicit	Total
Sports news	126	108	234
Topical interview	112	121	233
Overview	111	36	147
Invitation	110	60	170
Reflective essay	107	106	213
Critical review	102	96	198
Letters from readers	101	126	227
Advice column	97	105	202
Weather forecast	96	13	109
Comment	92	105	197
News report	89	77	166
Readers' survey	89	81	170
Description	88	91	179
Collection	72	92	164
Personality-focused interview	35	71	106
Total	1 427	1 288	2 715

On the other hand, high explicitness of discourse relations is typical for other genres, such as personality-focused interview or letters from readers. In these cases, explicitness can be connected with a higher complexity of the situations/events described, typically argumentation in letters from readers or various inter-personal reactions and argumentation in dialogue personality-focused interviews (see Example 4, personality-focused interview).

(4)   (1) Do té doby se aparát [Československé] konfederace [sportovních a tě-
lovýchovných svazů] vystěhuje?␣QUESTION
(2) "Nikdo z nás si nedovolí zbytečně zabírat nějaké prostory.␣OPPOS-
ITION TO (3), EXPRESSED BY **ovšem** IN (3)
(3) Nejdříve se **ovšem** musí rozdělit majetek konfederace.␣CONJU-
NCTION WITH (4), EXPRESSED BY **a** IN (4)
(4) **A** jestli v této budově tělovýchova zůstane i nadále, bude záležet na
tom, zda české svazy budou mít ve svých dosavadních kancelářích na Stra-
hově dost prostoru, **či**␣INTRA-SENTENTIAL DISJUNCTIVE ALTER-
NATIVE, EXPRESSED BY **či**␣ vzhledem ke svým novým rozšířeným
kompetencím budou mít zájem i o kanceláře tady Na Poříčí."

(1) Until then, the apparatus of the [Czechoslovak] confederation [of sport
associations] will move out?␣QUESTION
(2) "None of us will dare to occupy any space unnecessarily.␣OPPOS-
ITION TO (3), EXPRESSED BY **however** IN (3)
(3) **However**, the property of the confederation has to be divided
first.␣CONJUNCTION WITH (4), EXPRESSED BY **and** IN (4)
(4) **And** whether the sport association stays in the building will depend
on whether the Czech unions will have enough space in their existing
offices at Strahov, or␣INTRA-SENTENTIAL DISJUNCTIVE ALTER-
NATIVE, EXPRESSED BY **or**␣ due to their new extended competencies
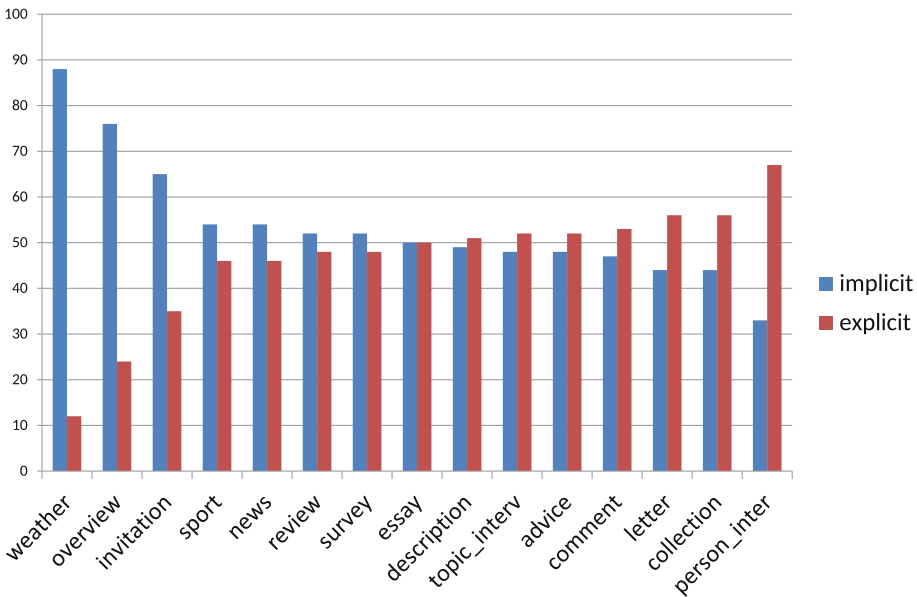will also be interested in offices here in Na Poříčí."

Fig. 1. Implicitness of discourse relations according to the text genre. The graph shows
relative frequencies of implicit (left column) and explicit relations (right column) in
individual genres.

3.2   Which Senses Are More Likely to Be Implicit?

The semantics of discourse relations differs in the specificity of the meaning: some relations have a very clear and narrow semantics with specific features which can be easily recognized (e.g. correction, with a typical discourse connective *not X - but rather Y*), whereas meaning of the others can be wide and rather free (conjunction).

Based on this observation, we assumed that discourse relations with more specific meaning would be more likely expressed by explicit devices while relations with "wider" semantics would be more often implicit.

Further, we assumed that implicit discourse relations cannot cover the whole range of semantic categories introduced for explicit discourse relations, because some discourse connectives have specific lexical semantics which cannot be deduced from the context.

The distribution of implicitness and explicitness across semantic types of discourse relations is presented in Table 5 and in relative frequencies in Fig. 2. As

Table 5. Numbers of occurrences of implicit and explicit discourse relations according to the semantic type of the discourse relation in the PDiT-EDA.

|	Implicit	Explicit	Total
Conjunction	446	462	908
Specification	236	60	296
Empty relation	196	8	204
Reason–result	125	169	294
Explication	104	29	133
Opposition	67	178	245
Confrontation	66	31	97
Precedence–succession	55	47	102
Generalization	25	8	33
Instantiation	23	10	33
Equivalence	19	4	23
Gradation	12	22	34
Conjunctive alternative	12	8	20
Concession	11	44	55
Correction	10	15	25
Restrictive opposition	6	24	30
Synchrony	5	12	17
Condition	4	109	113
Purpose	3	26	29
Disjunctive alternative	1	12	13
Total	1 426	1 278	2 704
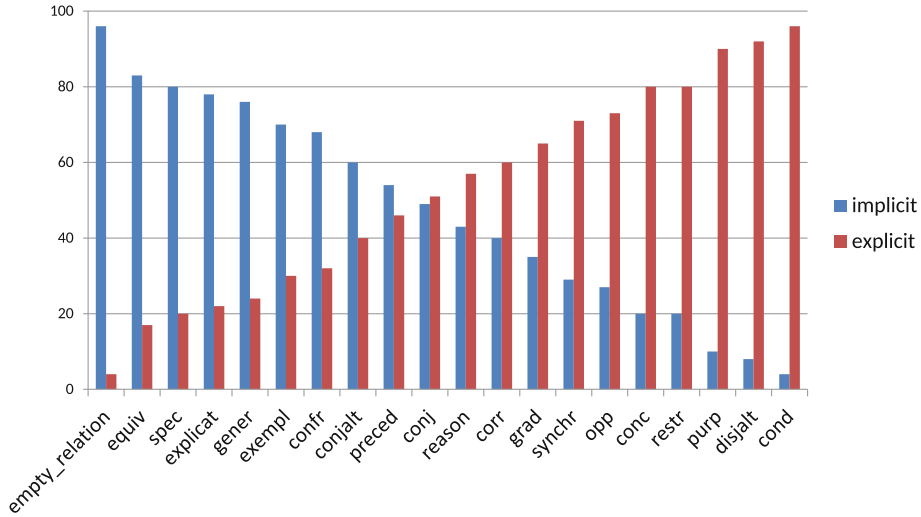
Fig. 2. Implicitness of discourse relations according to their semantic type. The graph shows relative frequencies of implicit (left column) and explicit relations (right column) for individual semantic types.

can be seen from these figures, there is a wide range of semantic types of discourse relations which can be realized implicitly. Contrary to our assumptions, implicit discourse relations cover almost all semantic categories of discourse relations.

Based on our annotation experience with explicit discourse relations, we did not expect specific groups of relations to be realized predominantly implicitly. However, there is a quite high occurrence of implicit *equivalence, specification, generalization* and *instantiation* in our data, i.e. relations from the general group of expansion. On the other hand, relations with a narrow semantics based on the links between events are rather explicit (*condition, purpose, concession*) [1].

There are interesting differences in implicitness/explicitness among relations which seem to be semantically close to each other. Within the class of temporal relations, the relation of *precedence–succession* is more often implicit, while *synchrony* tends to be expressed explicitly. Similarly, there is a big difference between two types of alternatives: *conjunctive alternative* is more often implicit, while *disjunctive alternative* is expressed with a discourse connective almost obligatorily.

3.3   Implicit Discourse Relations and Negation

We can assume that relations with an explicit discourse connective are easier to understand than implicit relations; there is no need to deduce the meaning of the relations from other linguistic signals or from world knowledge, the connective is a strong semantic signal. Intuitively, we assume that implicit discourse relations are more complex structures for the semantic decoding than the explicit ones.

A question arises then, whether this kind of complexity is in a certain interplay with other complexity features of discourse structure. Another such complexity feature could be e.g. sentence negation. A negative sentence requires one more cognitive operation of the recipient than an affirmative one. Is it possible that discourse arguments with sentence negation are connected rather explicitly than implicitly, in order to facilitate the understanding, or, in other words, to compensate the complexity of negation with the simplicity in the discourse structure.

We have observed that there is a relation between sentence negation and implicitness/explicitness in our data, see Table 6 and Fig. 3 with relative frequencies of implicit and explicit discourse relations depending on the presence of sentence negation in none, one or the other, or both arguments.

Table 6. Numbers of occurrences of implicit and explicit discourse relations according to presence of sentence negation in the two arguments in the PDiT-EDA (the arguments are referred to as start and target; "-" marks presence of negation).

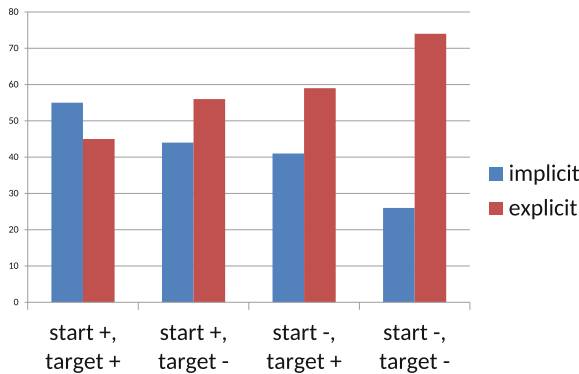| Negation	Implicit	Explicit	Total
start −, target −	10	29	39
start +, target +	1 234	1 008	2 242
start −, target +	89	130	219
start +, target −	94	121	215
Total	1 427	1 288	2 715

Fig. 3. Implicitness of discourse relations according to the sentence negation of the two arguments. The graph shows relative frequencies of implicit (left column) and explicit relations (right column) for various combinations of presence of sentence negation in the two arguments (the arguments are referred to as start and target; "-" marks presence of negation).

We find significant differences in implicitness/explicitness of discourse relations in structures with affirmative and negative discourse arguments, confirmed by a Chi-square test. Disregarding semantic types of discourse relations, we can say that the higher occurrence of negation correlates with the higher occurrence of explicit discourse relations. Thus, the complexity on the axis of affirmation and negation is compensated with the relative non-ambiguity of the meaning of the discourse relation.

4   Conclusion

The analysis of implicit and explicit discourse relations in Czech led to results and further observations on more levels then just one. First, we got a description of data presenting the distribution of implicitness and explicitness of discourse relations in the analyzed texts. Generally, the amount of implicit and explicit discourse relations is comparable, with a prevalent occurrence of implicit discourse relations. The implicitness/explicitness of discourse relations is connected with further features of texts and the surrounding language context. E.g. a text genre determines the measure of explicitness: the genres with easy and predictable (formal) structure describing simple events without argumentation tend to more implicit way of expressing discourse relations than text genres reporting about complex plots or containing argumentation and subjectivity.

Semantics of the discourse relation is another decisive feature for its implicitness/explicitness. Some discourse relations with a very specific, narrow semantics are predominantly explicit (purpose, condition, concession); these relations hold typically between events as wholes. On the other hand, many relations from the expansion group (exemplification, specification, equivalence, generalization) are usually implicit; it is characteristic for these relations that their underlying semantic concepts can be identified when connecting units like nominal phrases, too.

We tested the relation between implicitness of discourse relations and the presence of negation in the discourse arguments, too. It turned out that implicitness decreases with the presence of sentence negation. We understand this phenomenon as a mechanism compensating the complexity of a negative discourse argument for the recipient in the communication: the more complex the internal structure of the discourse arguments, the simpler orientation in the external structure between the arguments.

Second, more general insights and questions arise from this analysis of the relation between implicit and explicit discourse relations. The further we get from the surface forms, the lower inter-annotator agreement we get. Although some cases of disagreement always come from a vague annotation scenario and from annotators' mistakes, we dare to say – having certain experience in annotation, and in agreement with other works such as [2,6] – that with higher structures in the language, we may come to a limit of a reliable annotation.

References

1. Jínová, P., Poláková, L., Mírovský, J.: Sentence Structure and Discourse Structure (Possible Parallels), Linguistics Today, vol. 215, pp. 53–74. John Benjamins Publishing Company, Amsterdam (2014)
2. Mírovský, J., Hajičová, E.: What can linguists learn from some simple statistics on annotated treebanks. In: Henrich, V., Hinrichs, E., de Kok, D., Osenova, P., Przepiórkowski, A. (eds.) Proceedings of 13th International Workshop on Treebanks and Linguistic Theories (TLT13). pp. 279–284. University of Tübingen, University of Tübingen, Tübingen (2014)
3. Mírovský, J., Mladová, L., Žabokrtský, Z.: Annotation tool for discourse in PDT. In: Huang, C.R., Jurafsky, D. (eds.) Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010). vol. 1, pp. 9–12. Chinese Information Processing Society of China, Tsinghua University Press, Beijing (2010)
4. Pajas, P., Štěpánek, J.: Recent advances in a feature-rich framework for treebank annotation. In: Scott, D., Uszkoreit, H. (eds.) The 22nd International Conference on Computational Linguistics - Proceedings of the Conference. vol. 2, pp. 673–680. The Coling 2008 Organizing Committee, Manchester (2008)
5. Pitler, E., Louis, A., Nenkova, A.: Automatic sense prediction for implicit discourse relations in text. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, vol. 2, pp. 683–691. Association for Computational Linguistics (2009)
6. Poláková, L.: K možnostem korpusového zpracování nadvětných jevů [on the possibilities of a corpus-based approach to discourse phenomena]. Naše řeč 4-5/2014, pp. 241–258 (2014)
7. Poláková, L., Mírovský, J., Nedoluzhko, A., Jínová, P., Zikánová, Š., Hajičová, E.: Introducing the prague discourse treebank 1.0. In: Proceedings of the 6th International Joint Conference on Natural Language Processing, pp. 91–99. Asian Federation of Natural Language Processing, Asian Federation of Natural Language Processing, Nagoya (2013)
8. Prasad, R., et al.: Penn Discourse Treebank Version 2.0. Data/software (2008). lDC2008T05
9. Prasad, R., et al.: The Penn Discourse Treebank 2.0 Annotation Manual. Technical Report IRCS-08-01. Institute for Research in Cognitive Science, University of Pennsylvania (2007)
10. Prasad, R., Webber, B., Lee, A., Joshi, A.: Penn Discourse Treebank Version 3.0. Data/software (2019). lDC2019T05
11. Rysová, M., et al.: Prague discourse treebank 2.0. Data/Software. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University (2016). http://hdl.handle.net/11234/1-1905

12. Taboada, M., Brooke, J., Stede, M.: Genre-based paragraph classification for sentiment analysis. In: Healey, P., Pieraccini, R., Byron, D., Young, S., Purver, M. (eds.) Proceedings of the SIGDIAL 2009 Conference. The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue, pp. 62–70. Association for Computational Linguistics, Stroudsburg (2009)

13. Webber, B.: Genre distinctions for discourse in the Penn TreeBank. In: Su, K.Y., Su, J., Wiebe, J., Li, H. (eds.) Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. pp. 674–682. Association for Computational Linguistics, Suntec (2009)

14. Webber, B., Prasad, R., Lee, A., Joshi, A.: The Penn Discourse Treebank 3.0 Annotation Manual. Technical report, University of Edinburgh (2019)

15. Webber, B., Stone, M., Joshi, A., Knott, A.: Anaphora and discourse structure. Comput. Linguist. **29**(4), 545–587 (2003)

16. Zeyrek, D., Demirşahin, I., Çallı, A.B.S., Kurfalı, M.: Annotating implicit discourse relations in Turkish & the challenge of annotating corrective discourse relations. Oral presentation. In: IPrA Conference 2015, Antverp, Belgium (2016)

17. Zikánová, Š., Synková, P., Mírovský, J.: Enriched Discourse Annotation of PDiT Subset 1.0 (PDiT-EDA 1.0). Data/Software. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University (2018). http://hdl.handle.net/11234/1-2906