Journal of Pragmatics 130 (2018) 16–32

Contents lists available at ScienceDirect

Journal of Pragmatics

journal homepage: www.elsevier.com/locate/pragma

Primary and secondary discourse connectives: Constraints and preferences

Magdaléna Rysová*, Kateřina Rysová

Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics, Malostranské náměstí 25, 118 00 Prague, Czech Republic

ARTICLE INFO

Article history:
Received 5 February 2017
Received in revised form 19 February 2018
Accepted 13 March 2018
Available online 16 April 2018

Keywords:
Discourse connectives
Primary connectives
Secondary connectives
Free connecting phrases
Prague Discourse Treebank
Czech

ABSTRACT

In this paper, we explore the linguistic factors that influence an author's choice of discourse connectives in the production of a coherent text. We focus on the competition between so-called primary connectives (grammaticalized and mostly one-word expressions such as *therefore*) and secondary connectives (not yet fully grammaticalized compositional discourse phrases such as *for this reason*). We attempt to describe the linguistic constraints on and preferences in connective selection. The analysis is based on manually annotated data from the Prague Discourse Treebank 2.0 (PDiT), which contains almost 50000 sentences from Czech newspaper texts. We demonstrate that discourse connectives are used in accordance with the economy principle in language, i.e. authors aim to achieve the maximal result with minimal effort. They most frequently choose short and semantically more generalized primary connectives. However, in cases where the discourse relations can be misunderstood, authors prefer more complex and specific structures.

© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

The transformation of mental content into a coherent text is a complex process whose essential feature is choice. The author chooses the most suitable way to express his or her thoughts under the influence of many different factors, both linguistic and extralinguistic. Hence, the text production process may be specific to each individual author as well as to each individual text type or genre.

In this paper, we focus on the linguistic factors which influence authors' choice of discourse connective expressions for particular contexts, including those which extend beyond sentence level, as in Example (1).

(1) *We have walked more than 25 kilometers through the woods in heavy rain.*
  a) ***Therefore**, I know that I can walk long distances in any weather.*
  b) ***Thanks to this**, I know that I can walk long distances in any weather.*
  c) ***Thanks to this trip**, I know that I can walk long distances in any weather.*

* Corresponding author.
E-mail addresses: magdalena.rysova@ufal.mff.cuni.cz (M. Rysová), rysova@ufal.mff.cuni.cz (K. Rysová).

https://doi.org/10.1016/j.pragma.2018.03.013
0378-2166/© 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

In Example (1), we may use any of the expressions in bold (i.e. *therefore, thanks to this, thanks to this trip*) to signal the discourse relation of result. Similarly, many different expressions may be used to signal other semantic types of relations. For example, the discourse relation of reason may be expressed through the conjunctions *because*, *since*, *as* and *for* as well through the multi-word phrase *the reason is that*, or the relation of condition may be signaled by expressions such as *if*, *the condition is*, *on condition that* or *under these conditions*. However, these expressions differ in many ways – lexically, syntactically, semantically as well as stylistically. Therefore, we can expect that they are not wholly interchangeable in 100% of contexts. This can be demonstrated by the following example featuring the inappropriate use of the connective *since*.

(2) *Peter is staying home.* ***The reason is that*** / *****Since*** *he is ill.*

In this paper, we describe the influences on the choice of discourse connectives in written texts. Our investigation involves two steps. First, we describe the factors which limit the set of candidate connectives due to structural restrictions. We discuss the contexts where the interchangeability of some connective types is not possible (cf. *the reason is that* vs. *since*), i.e. we focus on the description of linguistic constraints on discourse connectives across language levels (Section 5.1).

Second, we proceed from constraints to preferences in the use of discourse connectives in these texts. We demonstrate that although some uses of connectives are possible in certain contexts, they tend not to be preferred.[1] We focus on these preferences in Section 5.2 with the aim of describing the use of primary and secondary connectives in Czech.

2. Theoretical backgrounds

Prior to the analysis of the constraints on and preferences concerning discourse connectives, we discuss some essential theoretical issues related to text coherence with a focus on discourse connectives.

2.1. Text coherence and discourse theories

In the introductory part of their book, Halliday and Hasan (1976) claim that most people have the natural ability to determine whether a sequence of sentences is a coherent text or a random cluster of unrelated sentences. This suggests that there are objective characteristics of a text. In Halliday and Hasan's conception, a text is a hierarchical object constituted from smaller interconnected elements and "the interpretation of some element in the discourse is dependent on that of another" (Halliday and Hasan, 1976: 4). Discourse connectives are a class of language expressions whose function is to make this interpretation easier.

Halliday and Hasan introduce the concept of *a cohesive tie* for the relation between two text units and *a cohesive chain* for a higher cohesive piece of discourse formed by several cohesive ties. A text may be imagined as a net of cohesive relations of a different kind. Halliday and Hasan divide cohesion into grammatical and lexical. Within grammatical cohesion, they distinguish reference, substitution, ellipses and conjunction. Within lexical cohesion, they focus on reiteration and collocation. In our paper, we focus on text coherence represented by semantic relations (conjunction in Halliday and Hasan's terminology) that may be implicit or explicit (i.e. either expressed by discourse connectives or not) and whose description has been the subject of discourse analysis, see especially Harris (1952) as one of the first linguists oriented toward the complex study of discourse phenomena.

Growing interest in discourse analysis and the rise of corpus linguistics brought about the need to capture discourse relations in large corpora. The most important approaches to this problem include Rhetorical Structure Theory (RST, Mann and Thompson, 1988), Segmented Discourse Representation Theory (SDRT, Asher, 1993, later Asher and Lascarides, 2003) or the Penn Discourse Treebank project (PDTB, Prasad et al., 2008). These approaches reflect the difference between global and local discourse structure modeling – RST and SDRT belonging to the former by representing a text as an abstract structure and PDTB to the latter by analyzing discourse relations through discourse connectives (i.e. their lexical anchors).

These discourse theories led to the development of many annotated corpora and treebanks, see e.g. the RST Discourse Treebank (Carlson et al., 2001) based on Rhetorical Structure Theory, DISCOR (Reese et al., 2007) and ANNODIS (Afantenos et al., 2012) following Segmented Discourse Representation Theory or corpora inspired by the Penn Discourse Treebank annotation like the Potsdam Commentary Corpus (Stede and Neumann, 2014) or the French Discourse Tree Bank (Danlos et al., 2012).

2.2. Discourse connectives

Before analyzing constraints on and preferences in discourse connectives, it is necessary to delimit this class of expressions. Discourse connectives, along with other discourse markers like *oh*, *well* or *you know*, are part of a broader category of discourse relational devices (DRDs)[2] that participate in text coherence. While discourse connectives appear in both written

[1] Throughout this text, the term "preferred" is used to mean most frequent in the corpus data.

[2] This approach follows the practice of the European TextLink COST action (http://textlink.ii.metu.edu.tr) aimed at the inventarization, annotation and cross-linguistic analysis of DRDs.

and spoken language, discourse markers are typically observed in spoken language. DRDs (identified using various terms) have been studied extensively since the late 1970s.[3]

Very generally, discourse connectives are linguistic means whose function in a text is to signal semantic and rhetorical discourse relations such as condition, reason, result, conjunction, opposition or specification. However, approaches to connectives differ greatly. Individual authors use the term discourse connectives for differing classes of expressions, both more broadly and more narrowly conceived, and define them according to various language domains such as part-of-speech perspective, position in the sentence or prosody.

According to Zwicky (1985), discourse connectives often occur at the beginning of sentences, are separated prosodically from the surrounding context by intonation breaks or pauses, and are usually monomorphemic and syntactically isolated from the rest of the sentence. Schiffrin (1987) presents a broad category of discourse markers that do not easily fit into linguistic classes. She suggests that even paralinguistic features and non-verbal gestures can be discourse markers. In Schiffrin's approach, discourse markers in the form of linguistic means (i.e. connectives in our terminology) are delimited according to the following conditions: they have to be syntactically detachable from a sentence, commonly used in the initial position of an utterance, and must have a range of prosodic contours. Fischer (2006) delimits the most typical connectives (she uses the term discourse particles) as small, syntactically, semantically and often prosodically unintegrated, uninflected words. According to Urgelles-Coll (2010), discourse markers (connectives) are phonologically short and reduced words. They are not integrated syntactically and can be omitted from a sentence without affecting its grammaticality. Semantically, they do not usually affect the truth-conditions of the proposition in which they appear.

Most authors agree on the basic characteristics of typical discourse connectives, but disagree on the boundaries of their delimitation. Some authors (cf. Shloush, 1998, Hakulinen, 1998 or Maschler, 2000 who limit connectives to grammatical words) define connectives in a narrow sense, others in a broader sense, e.g. by also including verbal and noun phrases (see Schiffrin, 1987; Hansen, 1998 or Aijmer, 2002). However, there is no existing uniform and commonly shared definition of discourse connectives.

In the PDTB approach (Prasad et al., 2008), discourse connectives are described as predicates with two arguments (defined as abstract objects according to Asher, 1993). Discourse connectives open positions for argument 1 (Arg1) and argument 2 (Arg2), the latter being a host clause for the connective. In the PDTB approach, discourse connectives are distinguished from the alternative lexicalizations of connectives (AltLexes), i.e. multi-word expressions like *for one thing*, *one reason is*, *adding to that speculation*, *the increase was due mainly to*, *a consequence of their departure could be* (examples taken from Prasad et al., 2010). The PDTB approach has also inspired the discourse annotation of the Prague Discourse Treebank, whose data were used as a source of material for this paper (see Section 3).

2.2.1. Primary and secondary connectives

In our approach, we delimit discourse connectives following Rysová and Rysová (2014), who define connectives from the functional point of view, as expressions that i) signal one of the semantic types of discourse relations (such as reason, condition, opposition, specification etc.) between two text units, and at the same time, ii) are suitable for many contexts with the given type of relation, i.e. that are nearly context independent,[4] witness Example (3) with the discourse relation of result.

(3) *It rained heavily in the afternoon.* ***Therefore*** / ***For this reason,*** *the trip was canceled.*
All restaurants were closed. ***Therefore*** / ***For this reason,*** *tourists were disappointed.*
The train was broken. ***Therefore*** / ***For this reason,*** *the people got in late.*

We also further divide connectives into primary and secondary (Rysová and Rysová, 2014, 2015). Primary connectives are grammaticalized single word units or non-compositional multi-word units (which can form correlative pairs like *either_or* or other complex forms like *so that*) that are lexically fixed, uninflected, and do not allow for internal modification. Examples of primary connectives are *and*, *but*, *because*, *when*, *if*, *however*, *therefore*, *so* or *thus*.

Secondary connectives are not yet fully grammaticalized, compositional structures that are lexically freer, i.e. they allow for lexical variation such as *for this / that / the given reason*, often inflected (*under this condition – under these conditions*) and can undergo internal modification (*the main / only / basic condition is*).[5] They contain so-called core units, i.e. semantically

[3] See van Dijk, 1979; Zwicky, 1985; Schiffrin, 1987; Fraser, 1990 or 1999; Fischer, 2006; Urgelles-Coll, 2010.

[4] This means that discourse connectives can be used in various contexts with the appropriate discourse relation. For example, the connective *therefore* is suitable for various contexts with the sense of result: *I am hot. Therefore, I cannot wear this jacket and stockings. / It is far away. Therefore, I will take a taxi*. Of course, the connective *therefore* is not appropriate for contexts with other relation types such as contrast*: I want to start eating healthy. However, / *Therefore, I don't like vegetables*.

[5] Modifiable connectives contain an expression (often of evaluative or modal nature) that further specifies/intensifies the discourse relation without changing its semantic type. It is necessary to distinguish between modified connectives and complex connectives. Complex connectives consist of two or more connective words that combine to form an expression of a single semantic type of discourse relation. Complex connectives occur in a single argument (*so that*) or they may form correlative pairs (*either_or*). All members of a complex connective participate in the expression of a single discourse relation, e.g. the expression *so that* signals the relation of purpose as a whole, and both parts of *either_or* express the relation of disjunctive alternative. On the other hand, in modified connectives, the modification (e.g. *main* in *the main condition is*) does not participate in the expression of the discourse relation of condition, but rather, merely modifies it by expressing the intensity of the relation. For more details see Rysová (2015).

strong words like *reason, condition* or *purpose*. The core words of multi-word secondary connectives are mainly nouns (like *explication, conclusion, result*), prepositions (like *due to, because of, on the basis of, thanks to* etc.) and verbs[6] (like *to precede, to conclude, to follow, to sum up*).

There are many "prototypical" primary and secondary connectives that conform to all of the mentioned criteria, cf. primary connectives like *and*, *but*, *or* and secondary connectives like *that is the reason why*, *the condition is* or *to come to a conclusion*. At the same time, not all primary and secondary connectives fulfil all of these criteria, e.g. the Czech primary connective *aby* "so that/in order to" may be inflected, cf. the forms *abych*, *abys*, *aby*, *abychom*, *abyste*, *aby* meaning "so that/in order to" with the relevant person as the subject of the following clause. Therefore, primary and secondary connectives do not constitute two closed and strictly separated classes. They are, rather, a scale of expressions differing in degree of grammaticalization, which means that some secondary connectives may eventually become primary due to language change and potential increases in grammaticalization.

The transformation of secondary connectives into primary ones is well documented by their historical origin. Present-day primary connectives often developed from similar structures that can be observed in contemporary secondary connectives. They are very frequently composed of a preposition and an anaphoric element, cf. the highly frequent English connective *therefore* (from the preposition *fore* and an anaphoric particle *there*), German *darum* "therefore" (preposition *um* and anaphoric particle *da*) or Czech *proto* "therefore" (preposition *pro* and anaphoric part *to*). More details on the origin of present-day primary connectives in relation to secondary ones are provided in Rysová (2017).

2.2.1.1. Variability of secondary connectives: lexical realizations and grammatical variants. Since secondary connectives are not fully grammaticalized structures, they exhibit a high degree of variation (in contrast to primary connectives) and it is sometimes difficult to decide whether two structures (e.g. *under this condition* vs. *under that condition*) are two separate secondary connectives or merely variants of the same connective. Therefore, further hierarchical categorization has to be done.

As presented above, secondary connectives contain so-called core units. We can observe that many secondary connectives with the same core unit are formed according to the same general scheme. For example, the structures *under this condition*, *under that condition* or *under the given condition* belong to the scheme "*under*_Pronoun/Adjective$_{\text{Anaphoric}}$_*condition*". We call these structures *lexical realizations* and delimit them as forms of secondary connectives containing slightly different lexical items (cf. *this* / *that* / *the given*).

In addition, the individual lexical realizations may appear in a text as several *grammatical variants*, differing in morphology (*under this condition* vs. *under these conditions*) or word order (e.g. in Czech *to je důvod, proč* "that is the reason why" vs. *je to důvod, proč* lit. "is that the reason why").[7] We thus propose the following hierarchy for secondary connectives: i) a general scheme of secondary connectives ("*under*_Pronoun/Adjective$_{\text{Anaphoric}}$_*condition*".), ii) its lexical realizations (*under this condition*, *under that condition* or *under the given condition*), and iii) the grammatical variants of these realizations (*under this condition* vs. *under these conditions*).

We will now return to the question stated above, i.e. which structures are actually connectives in their own right and which are merely connective variants. In our approach, all the lexical realizations of the same general scheme (with the same core unit) belong to a single secondary connective (i.e. we do not consider expressions like *under this condition*, *under that condition* or *under the given condition* to be separate secondary connectives). The basic representative form of a single secondary connective is then the most frequent lexical realization (presented in the basic grammatical form). In our case, the basic form is *under this condition*,[8] which also appears in other (less frequent) lexical realizations like *under that condition* or *under the given condition*. All of these realizations also have grammatical variants (*under these conditions*, *under those conditions*, *under the given conditions*). On the other hand, expressions such as *the condition is* or *on condition that* are separate secondary connectives because they have a different general scheme.

2.2.2. Free connecting phrases

Discourse connectives are not the only expressions used to signal discourse relations. This capacity is also a feature of free connecting phrases (the third term introduced by Rysová and Rysová, 2014, 2015), which differ from connectives (both primary and secondary) in that they are highly context-dependent. In other words, free connecting phrases can only be used in a few specific contexts whereas connectives may be used in almost any context with the relevant discourse relation (e.g. *due to this* is a secondary connective, whereas *the increase was due mainly to* or *a consequence of their departure could be* are free connecting phrases).[9] Therefore, only the expressions *therefore* and *thanks to this* from Example (1) may be considered connectives, whereas *thanks to this trip* is a free connecting phrase, witness Example (4).

(4) *I believe in myself.*
  a) ***Therefore / Thanks to this**, I am happy.*
  b) **__Thanks to this trip__, I am happy.*

[6] These verbs are called discourse verbs by Danlos (2006).

[7] The degree of grammatical variation differs across languages, e.g. it is especially high in inflected languages like Czech.

[8] The frequency of the mentioned secondary connectives with the core unit *condition* was measured in *IntelliText 2.6* (http://corpus.leeds.ac.uk/itweb/htdocs/Query.html).

[9] The PDTB category of AltLexes corresponds to the combined categories of secondary connectives and free connecting phrases.

Free connecting phrases always contain a referential component like *this weather* in *due to this weather* or *their departure* in *the consequence of their departure could be* and therefore must be always analyzed in context.

2.3. Constraints in discourse

As illustrated in Example (2), even connectives with a similar meaning are not wholly interchangeable in 100% of contexts (cf. e.g. *since* vs. *because* vs. *the reason is*). Therefore, we decided to investigate the constraints on discourse connectives. Our analysis uses Czech data, but we believe that our general conclusions may be applied (to a certain extent) to other languages as well.

Constraints on the selection of particular connectives have been studied for languages such as French (see Degand, 2000 or Degand and Fagard, 2012), English (see Soria, 2005) or Turkish (see Zeyrek et al., 2012). The authors usually select particular connectives or specific groups of connectives (cf. causal connectives vs. causal prepositions for French in Degand, 2000 or Degand and Fagard, 2012, the connectives *fakat*, *yoksa* and *ayrıca* for Turkish in Zeyrek et al., 2012) and examine the conditions under which authors select one of them in a particular context. Many authors are also currently focusing on the pragmatic aspects and (poly)functionality of discourse connectives and markers, i.e. they interpret particular markers in various contexts and combinations (see Fraser, 2015; Gonen et al., 2015; Fuentes-Rodríguez et al., 2016; Marmorstein, 2016; Smith-Christmas, 2016; Tanghe, 2016 or Thaler, 2016).

3. Data: Prague Discourse Treebank 2.0

This study is based on data from the Prague Discourse Treebank 2.0 (PDiT, Rysová et al., 2016). The PDiT is a multi-layer annotated corpus of Czech newspaper texts (3165 documents, 49431 sentences, 833195 tokens) containing discourse annotation built upon the data from the Prague Dependency Treebank (PDT, Hajič et al., 2006). The PDT combines annotations of three language layers at once: morphological, surface syntactic and deep syntactic (tectogrammatical). The sentences are represented by dependency trees, as in Fig. 1. The PDT also includes annotation of other phenomena such as information structure, pronominal and nominal coreference, bridging anaphora and multi-word expressions.

Discourse annotation was carried out on top of the deep-syntactic trees – see Example (5) and Fig. 1 – and covers relations expressed by an explicit connective. The discourse annotation was published as the Prague Discourse Treebank. The first version (PDiT 1.0, Poláková et al., 2012a; described in Poláková et al., 2012b) covered discourse relations expressed by explicit connectives (delimited as conjunctions, adverbs, particles, some punctuation marks, some uses of pronouns and some types of idiomatic multi-word phrases). The newest version PDiT 2.0 reflects the division of connectives into primary and secondary. It contains a revision of the previous annotation (some types of explicit connectives, such as idiomatic multi-word

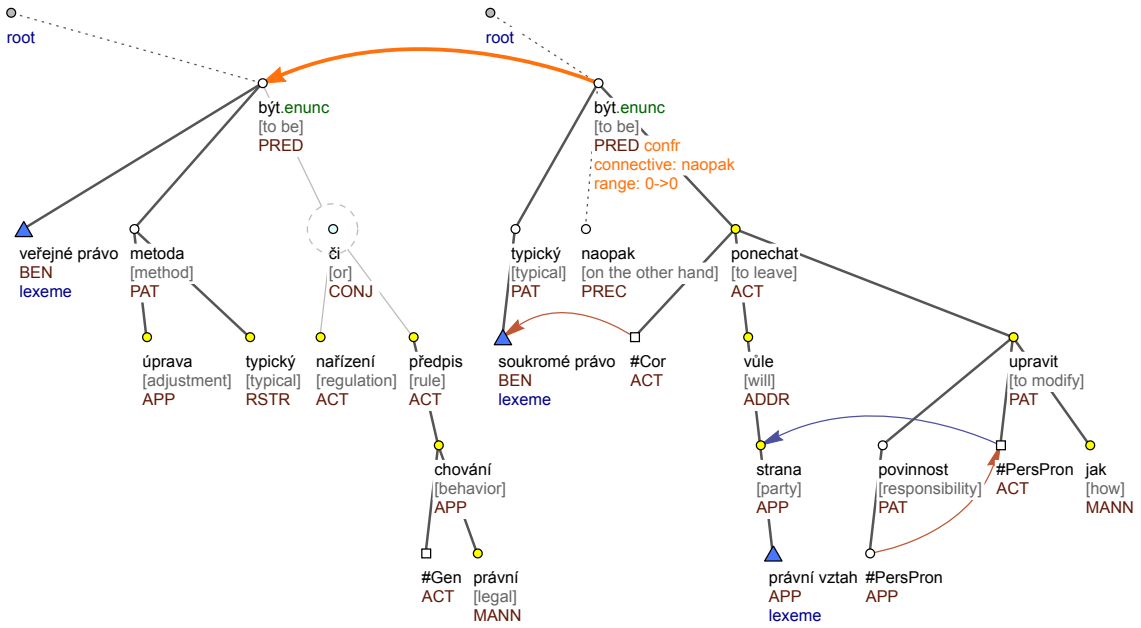

Fig. 1. Inter-sentential discourse relation in the PDiT 2.0 presented in Example (5).

phrases, were re-annotated as secondary connectives) and an annotation of a new set of secondary connectives (e.g. *for this reason*).[10]

The PDiT annotation contains both inter- and intra-sentential discourse relations. Each relation is also assigned a semantic type/sense. The set of semantic relations (the complete list of these can be found in Table 1) is inspired by the Penn Discourse Treebank 2.0 sense hierarchy (Prasad et al., 2008) and by the syntactico-semantic labels used for the representation of compound sentences on the deep-syntactic layer of the Prague Dependency Treebank. The sense hierarchy includes four major categories (contrast, expansion, contingency and temporal) that are further divided into 22 individual senses (examples for each sense are provided in Zikánová et al., 2015).

(5) *Pro veřejné právo jsou typickou metodou úpravy nařízení či předpis právního chování. Pro soukromé právo je* ***naopak*** *typické ponechat na vůli stran právního vztahu, jak své povinnosti upraví.*
"For public law, adjustments to the regulations or rules of legal behavior are a typical method. **On the other hand**, for private law, it is typical to leave to the will of the parties to the legal relationship how they will modify their responsibilities."

Fig. 1 represents two dependency trees from the PDiT 2.0 and demonstrates an inter-sentential discourse relation, see also Example (5). The discourse relation is represented by a thick orange arrow connecting the roots of the discourse arguments. The semantic type of the relation is represented by abbreviations, e.g. *confr* (*confrontation*). In this case, the discourse relation is signaled by the connective *naopak* "on the other hand". The annotation also captures the range of the discourse arguments, e.g., the symbol *0→0* in Fig. 1 means that the discourse relation holds only between the two sentences displayed. The newest version of the corpus, the Prague Discourse Treebank 2.0, was used as the source data for the analysis of discourse constraints and preferences presented in this paper.[11]

4. Methods

4.1. Methods used for analyzing discourse constraints

To analyze constraints on discourse connectives, we carried out an experiment using the PDiT data. Using a sample of selected connectives, we tested their suitability for various contexts and focused on their uses which the annotators evaluated as inappropriate. The contexts with the "inappropriate" connectives were divided into several groups (according to the particular type of constraint). The results of the linguistic analysis are provided in Section 5.1.

Example (6) illustrates an inter-sentential discourse relation of result expressed by the connective *proto* "therefore".

(6) *Pro 600 zaměstnanců muselo nové vedení sehnat práci.* ***Proto*** *se manžeři rozjeli za zakázkami nejen po republice, ale i do zahraničí.*
"The new management had to find a job for 600 employees. **Therefore**, the managers went looking for orders not only throughout the country, but also abroad."

At the same time, this type of relation in Czech may be signaled by a variety of other expressions like *a tak* lit. "and so"; *z tohoto důvodu* "for this reason"; *kvůli tomu* "because of this"; *to byl důvod, proč* "this was the reason why"; *díky tomu* "thanks to this" etc. Therefore, if an author wants to express this type of relation, all of these connectives are potential candidates.

We automatically extracted all contexts with the relation of result and all connectives expressing this type of relation in the PDiT. From among these, we selected the most frequent connectives – the four most frequent primary connectives: *proto* "therefore", *tedy* "thus", *takže* "therefore", *a tak* "therefore" lit. "and so" and the eight most frequent secondary connectives: *z tohoto důvodu* "for this reason", *to je důvod, proč* "this is the reason why", *kvůli tomu* "because of this", *díky tomu* "thanks to this", *z toho důvodu* "for this reason", *na základě čehož* lit. "on the basis of which", *z uvedeného důvodu* "for the given reason" and *vinou toho* "due to this" (with a negative connotation, lit. "by fault of this").

Table 1
List of senses annotated in the Prague Discourse Treebank.

Contrast	Confrontation	Opposition	Restrictive opposition	Pragmatic contrast	Concession	Correction	Gradation
Expansion	Conjunction	Conjunctive alternative	Disjunctive alternative	Instantiation	Specification	Equivalence	Generalization
Contingency	Reason–result	Pragmatic reason–result	Explication	Condition	Pragmatic condition	Purpose	
Temporal	Synchronous	Asynchronous					

[10] Discourse relations in the PDiT were not annotated according to a pre-defined list of connectives. The annotators were asked to search for connectives according to a general definition, illustrated by examples.

[11] The PDiT 2.0 can be downloaded as a single zip archive from the LINDAT-Clarin repository, see https://ufal.mff.cuni.cz/pdit2.0/data.

In the next step, we tested each of the selected connectives in 100 contexts of result and analyzed the contexts for which the given discourse connective was deemed not appropriate.[12] Of the 100 contexts selected for testing, 50 originally contained primary connectives and 50 contained secondary connectives. To illustrate, we can take the context of Example (6) with the discourse relation of result and substitute the original connective *proto* "therefore" with other selected connectives of result, see Example (7).

(7) *Pro 600 zaměstnanců muselo nové vedení sehnat práci.*
   a) ***Takže / A tak / Z tohoto důvodu / To je důvod, proč / Kvůli tomu / Díky tomu / Z toho důvodu / Z uvedeného důvodu*** *se manažeři rozjeli za zakázkami nejen po republice, ale i do zahraničí.*
   b) ****Tedy / *Na základě čehož / *Vinou toho*** *se manažeři rozjeli za zakázkami nejen po republice, ale i do zahraničí.*

   "The new management had to find a job for 600 employees.
   a) **Therefore / So / For this reason / This is the reason why / Because of this / Thanks to this / For that reason / For the given reason**, the managers went looking for orders not only throughout the country, but also abroad.
   b) ***Thus$_{\text{inappropriate position}}$ / *On the basis of which / *Due to this$_{\text{negative connotation}}$**, the managers went looking for orders not only throughout the country, but also abroad."

Example (7) demonstrates that not all of the selected Czech connectives of result suit this particular context. *Tedy* "thus" is not suitable because it is not being used in the first position of the sentence, *na základě čehož* lit. "on the basis of which" deviates from Czech grammatical structures (the appropriate use of this connective is intra-sentential), and *vinou toho* lit. "by fault of this" is pragmatically inappropriate, as it is bound to negative contexts. In our investigation, we focused on cases where the selected connectives did not fit into some of the tested contexts.[13] The linguistic analysis of all of these cases is described in Section 5.1.

4.2. Methods used for analyzing discourse preferences

For the analysis of preferences in the use of discourse connectives, we utilized a corpus data analysis method based on the use and frequency of connectives in the PDiT 2.0. We analyzed primary and secondary connectives in terms of their general frequency (Sections 5.2.1 and 5.2.2) and inter- vs. intra-sentential usage (Section 5.2.3).

5. Results and evaluation

5.1. Constraints on discourse connectives

As stated above, we examined all contexts where the use of the selected connectives was marked as inappropriate by the annotators in our experiment. The individual cases of such inappropriateness concern syntax, semantics, pragmatics, stylistics and salience. We will now deal with each level in turn.

5.1.1. Syntactic constraint: coordination vs. subordination

The use of some connectives is strongly limited by syntactic constraints in certain contexts. The connective *na základě čehož* lit. "on the basis of which" is highly limited by its boundedness to intra-sentential relations. Its inter-sentential usage is perceived as syntactically inappropriate by native speakers of Czech.

This restriction concerns subordinating connectives in general – they are bound to intra-sentential usage (for a detailed analysis of this, see Section 5.2.3). It is interesting that Czech subordinating connectives constitute together a quarter of all tokens of primary connectives in the PDiT 2.0., which indicates that this kind of constraint affects a relatively large segment of connective usages.

In addition, some coordinating structures also demonstrate preference for intra-sentential usage, but are not syntactically bound to it like the subordinators are. The intra-sentential usage of coordinating expressions is described in Section 5.2.3.

5.1.2. Word order constraint

Another syntactic constraint concerns the position of connectives in the argument. During our experiment, we discovered that some of the tested connectives did not fit into several contexts due to word order restrictions. In other words, some connectives are bound to a certain position in a sentence, whereas others can move (to a certain extent). In some of the tested

[12] The annotators were asked to select one of two options: appropriate vs. inappropriate use of a connective. They were not asked to specify the kind or degree of inappropriateness, but they were instructed to make a brief comment in case of doubt and to specify the reasons for their hesitation in each particular case. The aim of the experiment was to collect all the inappropriate uses of connectives, which were then analyzed in detail.

[13] The experiment was conducted by three linguists with long-term experience in processing and analyzing discourse relations. Their inter-annotator agreement was 0.73 of Cohen's κ (annotator 1 – annotator 2), 0.75 of Cohen's κ (annotator 1 – annotator 3), and 0.64 of Cohen's κ (annotator 2 – annotator 3).

contexts, the original connective (e.g. *tedy* "thus") appeared inside the discourse argument and therefore, it was not possible to replace it with another (e.g. *takže* "so") that was bound to the initial position, see Example (8).

(8) a) *Další kritéria jsou naprosto nezajímavá, nebudu je **tedy** uvádět.*
"The other criteria are completely uninteresting. I will **thus** not list them here."

b) **Další kritéria jsou naprosto nezajímavá, nebudu je **takže** uvádět.*
"*The other criteria are completely uninteresting. I will **so** not list them here."

c) *Další kritéria jsou naprosto nezajímavá, **takže** je nebudu uvádět.*
"The other criteria are completely uninteresting, **so** I will not list them here."

In Czech, this kind of word order restriction is typical for subordinating conjunctions in general – in prototypical cases, they are bound to the first position in the argument. Similarly, basic coordinating conjunctions like *a* "and" or *nebo* "or" are typically bound to the position between arguments.[14] On the other hand, connectives that are adverbial in character (like *proto* "thus" or *díky tomu* "thanks to this") usually behave more freely in this way – many of them can occupy both the first and the second positions in the argument.

We presume that the rules for placing the individual connectives differ across languages, for example, if we consider the word order differences between synthetic and analytic languages, as in Example (9). The Czech secondary connective *kvůli tomu* "due to this" can be placed after the predicate, which is not possible in English. The placement of connectives in Czech is freer than in English, which has more fixed word order.

(9) Czech: *Zaměstnavatel po mně chce potvrzení o zdravotní způsobilosti. **Kvůli tomu** musím jít k lékaři. / Musím jít **kvůli tomu** k lékaři.*
English: *My employer needs my health certificate. **Due to this**, I must go to my doctor. / *I must go **due to this** to my doctor.*

However, some general tendencies may be shared, such as the position of basic conjunctive and disjunctive connectives (like *a* in Czech, *und* in German, *and* in English; *nebo* in Czech, *oder* in German, *or* in English) on the boundary between the two discourse arguments,[15] see Example (10).

(10) *Půjdu do supermarketu **a** koupím nějaká jablka. – I will go to the supermarket **and** I will buy some apples. – Ich gehe in den Supermarkt **und** kaufe ein paar Äpfel.*

Půjdu do supermarketu koupím **a nějaká jablka. – *I will go to the supermarket I will buy **and** some apples. – *Ich gehe in den Supermarkt kaufe **und** ein paar Äpfel.*

Information of this type could be used, for example, for automatic annotation and/or detection of discourse arguments.

5.1.3. Semantic constraint – meaning of connectives

Several uses of the tested connectives were deemed inappropriate due to their specific semantic characteristics. Even connectives expressing the same type of semantic discourse relations are not always absolute synonyms, i.e. they can only be replaced in some semantic contexts, see Examples (11) and (12). Some connectives thus express different semantic nuances of the given discourse relation than others.

(11) *Fotbalový zápas se protáhl o 15 minut, **a tak** / **kvůli tomu** někteří fanoušci nestihli pravidelný odjezd vlaku.*
"The football game ran over by 15 min, **so** / **because of this** some fans did not catch the regular train departure."

(12) *Mám žízeň, **a tak** / ?**kvůli tomu** se půjdu napít.*
"I'm thirsty, **so** / ?**because of this** I'll go get a drink of water."

Both Examples (11) and (12) express a relation of result. The connective *a tak* "so" is suitable for both of them, while use of the connective *kvůli tomu* "due to this" in Example (12) is questionable (it is deemed inappropriate by native speakers of Czech). It seems that even a relation of result has some subtypes which differ, for example, in the intensity of the relation.

It would be possible to introduce a new discourse relation type for each of the semantic nuances. However, this solution would not be effective, because establishing a new relation for each individual subtype would result in an enormous list of senses. In this respect, a certain degree of generalization is necessary, especially regarding semantics. However, it is also

[14] Similar observations are also presented in Zikánová et al. (2015) who focus on the sentence position of selected Czech connectives in detail.

[15] This concerns single connectives (like *and*, *or* etc.); the correlative pairs of connectives (like *either_or*) are specific in this way, as they do not occur together between arguments, but rather, each of their parts is placed in a different argument.

necessary to be aware of the subtle semantic differences among connectives in the same sense category. Some connectives have a broader sense, and thus a broader range of usage in texts, whereas the others are narrower in these respects.

The narrower sense is often expressed by secondary connectives, e.g. structures like *důvodem je* "the reason is" and *příčinou je* "the cause is" express the relation of reason and both can be replaced with the connective *protože* "because", but the semantics of the words *důvod* "reason" and *příčina* "cause" are not exactly the same. At the same time, most secondary connectives are internally modifiable. In their modified form, secondary connectives have a more specific sense, cf. *hlavním/možným/dobrým důvodem je* "the main/possible/good reason is", which limits their use in many contexts.

5.1.4. Pragmatic constraint

One connective revealed to be suitable for very few contexts was the expression *vinou toho* "due to this" (lit. "by fault of this"). The use of this connective is limited by the negative connotations of the preposition *vinou* "due to". This preposition arose from the noun *vina* "fault", i.e. the expression *vinou toho* "due to this" states what is at fault for or the negative cause of something, see Example (13).

(13) *Dostatečně jsme se nesoustředili.* ***Vinou toho*** *jsme inkasovali první branku.*
"We did not concentrate enough. **Due to this**$_{\text{negative connotation}}$, we gave up the first goal."

The use of this connective is thus limited to negative contexts, as it expresses the cause of something undesirable. At the same time, the evaluation of context (whether it is positive or negative) is dependent on the author's attitude toward the content. The author then selects the most appropriate connective accordingly. Most of the Czech connectives are neutral in this sense and may be used both in positive and negative contexts. Some connectives even lose their original connotations and become neutral, e.g. the expression *díky tomu* "thanks to this", which also appeared in negative contexts in the PDiT.

Whereas the connective *díky tomu* "thanks to this" may also occur in some contexts where the author originally used *vinou toho* "due to this", substitution in the reverse order does not work at all – see Example (14) with the PDiT context where we have substituted the original connective *díky tomu* "thanks to this" with *vinou toho* "due to this" (with negative connotations).

(14) *Máme určité kontaktní možnosti ve všech státech, odkud pocházeli zahraniční studenti v bývalém Československu.* ****Vinou toho*** *bychom tam mohli hledat uplatnění pro naše lidi.*
"We have many contact opportunities in all the countries where the foreign students came from during the former Czechoslovakia. ***Due to this**$_{\text{negative connotation}}$ we can look for job opportunities for our people in these countries."

The close binding of the connective *vinou toho* "due to this" to negative contexts thus impedes its more frequent usage in texts. There are not many connectives with similarly strong negative or positive connotations in Czech; however, when this phenomenon does appear, the use of the connective is highly limited.

5.1.5. Stylistic constraint

Another factor influencing the appropriateness of connectives concerns stylistics. Some connectives are considered more formal, e.g. *z uvedeného důvodu* "for the given reason" or *z tohoto důvodu* "for this reason" than others, e.g. the stylistically neutral *proto* "therefore" or the rather informal *a tak* lit. "and so". This aspect is reflected particularly in different language modes, functional styles or genres, e.g. in differences between spoken and written language, formal and informal language or newspaper texts, legal texts or fiction. The differences in these modes and styles can be observed on different language layers as well as in the discourse structuring and the use of coherence devices. For example, the use of multi-word structures such as *z tohoto důvodu* "for this reason" is more frequent in formal written texts than in informal or spoken ones (for more details see Rysová, 2015) where the usage of formal structures is often stylistically inappropriate, see Example (15).

(15) ?*Mami, ten čaj je horkej.* ***Z uvedeného důvodu*** *ho budu pít opatrně.*
"?Mom, the tea is hot. **For the given reason,** I'm drinking it carefully."

However, slight differences in style can also be observed within the PDiT corpus itself. The PDiT contains written newspaper texts of different genres, e.g. interviews with politicians, weather reports, theater reviews, culture-related pieces or sports commentaries. In this respect, some of the selected connectives of result were not appropriate for some PDiT contexts because they were stylistically marked – e.g. the annotators evaluated the connective *a tak* lit. "and so" as inappropriate for more formal contexts and connectives like *z tohoto důvodu* "for this reason" or *z uvedeného důvodu* "for the given reason" for informal ones.

5.1.6. Long distance constraint and salience

In the previous sections, we have described constraints associated with primary and secondary discourse connectives. However, as stated in Section 2.2.2, we are also interested in free connecting phrases, i.e. contextual expressions like *díky tomuto výletu* "thanks to this trip". After finishing our experiment with selected connectives, we thus posed the question of

whether there are also constraints on free connecting phrases. These phrases were annotated in the PDiT with a total number of 151 occurrences.[16] Their frequency is thus rather low.

However, there are obviously some contexts where the author preferred the use of a free connecting phrase to the use of a discourse connective, i.e. the use of the particular structure like *díky tomuto počasí* "thanks to this weather" vs. more general connectives like *proto* "therefore". Therefore, we further investigated whether these cases only concern the author's subjective preference, or whether there are contexts where he or she is forced to use a free connecting phrase due to objective constraints. We analyzed contexts from the PDiT containing the free connecting phrases, e.g. *díky této výhře* "thanks to this victory" or *kvůli tomuto počasí* "due to this weather", and discovered that sometimes, the author cannot use a discourse connective due to the distance between discourse arguments.

Generally, discourse relations hold between two text spans – discourse arguments, defined according to Asher (1993) as abstract objects expressing situations, states, events, and the like. Discourse relations may appear between two adjacent arguments or the arguments may be separated from each other by another stretch of text, e.g. by one sentence, a sequence of sentences or even by a whole paragraph. In the vast majority of cases, discourse relations in the PDiT hold between two adjacent sentences/clauses, whereas long-distance arguments are rather rare. However, when these long-distance arguments do appear, they often impede the use of a connective because it could cause the misinterpretation of the given relation and disrupt the coherence of the whole text, see Example (16).

(16) *Družstvo B vyhrálo nad družstvem C v utkání, které bylo vyrovnané až do poslední chvíle. O konečném výsledku rozhodlo až posledních deset hodů, když oba naši zadáci dokázali svůj výkon vygradovat, a tím urvat utkání na naši stranu. Zejména výkon kapitánky patří do kategorie top výkonů. S tímto výkonem byla zaslouženě nejlepším hráčem utkání. Družstvo B se* ***díky této výhře*** *posunulo na 8. místo krajského přeboru.*
"Team B defeated team C in a match that was balanced until the last moment. The final outcome was decided by the last ten throws when both our defenders upped their performance and pushed us into the lead. The performance of the team captain in particular was one of the top performances. With this performance, she was the best player of the match, and deservedly so. **Thanks to this victory**, team B moved up to 8th place in the regional league rankings."

In Example (16), there is a discourse relation of result between the two underlined arguments, signaled by the phrase *díky této výhře* "thanks to this victory". We can see that the two arguments, which tell the reader about the victory of a team and the result on its shift in the regional league rankings, are not adjacent, but there is a set of other sentences between them which inform the reader about the details of the match. This is the reason why the author was forced to use the free connecting phrase *díky této výhře* "thanks to this victory" as opposed to one of the more frequent connectives like *proto* "therefore" or *a tak* lit. "and so" to signal a relation of result. The use of these connectives could cause misunderstanding of the given discourse relations and the reader could be confused as to which argument the connectives refer, i.e. how deep into the text they reach. If we use a connective such as *proto* "therefore", one possible interpretation could be that the relation of result is between the two adjacent sentences[17] – cf. in English *With this performance, she was the best player of the match, and deservedly so. Therefore, team B moved up to 8th place in the regional league rankings*. However, the author is not saying that the team moved up because one of the players was the best, but because the team won the match. The free connecting phrase *díky této výhře* "thanks to this victory" is thus used for the easier interpretation of deeply embedded relations in the text and for the better orientation of the reader.

At the same time, the free connecting phrases contain highly explicit anaphora, i.e. a noun (or nominal phrase) with a strong conceptual meaning, like *victory*, *weather*, *concert* or *experience*. These expressions already appear in the previous context (or are deducible from it) and their usage has an important function in the sentence information structure (or topic-focus articulation) and so-called salience (see below). The highly explicit anaphora inside the free connecting phrases topicalizes and recalls certain object(s) in order to keep them activated in the reader's consciousness, which connectives with a purely connective function can never do.

More specifically, connectives such as *therefore* or *due to this* cannot activate or recall specific objects such as *weather* or *concert* in the reader's consciousness. The generality (or high degree of context independency) of connectives benefits from the fact that the readers already have all the necessary objects activated in their minds. However, if the activation of an object is lost and if this object is needed to understand the following parts of a text, it is not sufficient to use connectives, but rather, it is necessary to select free connecting phrases that have the ability to refresh the object activation.

The explicitness of connective expressions is related to the salience (examining the degree of activation of an object in the reader's consciousness – see Hajičová et al., 2003, 2004) and relevance theory (Sperber and Wilson, 1986). In line with communicative expectations, a well-structured text should contain a balance of explicitness and implicitness, and the author should express discourse relations using the relevant language means so that the text is maximally coherent. Excessive implicitness, as well as the overuse and repetition of connective expressions, could be viewed as inappropriate and less comprehensible for the reader.

[16] Both secondary connectives and free connecting phrases were annotated systematically, i.e. based on the entire PDiT data.

[17] The readers subconsciously expect that the relation will hold between two adjacent sentences, because these short distance relations are the most commonly occurring ones in natural language.

5.2. Preference in discourse connectives

In the next step, we also analyzed the preferences in the use of discourse connectives. In many cases, a connective may be used in a particular context without disrupting text coherence, but this is not the preferred usage (preferred in the sense of its frequency in texts). In other words, such usage of a connective is not grammatically or semantically inappropriate, but it is rather rare in texts. Therefore, we extracted all discourse connectives in the PDiT (not only selected connectives of result) and examined the tendencies and preferences regarding their general frequency and intra- vs. inter-sentential usage.

5.2.1. Preference in connective types: principle of least effort

As we have described in Section 3, the PDiT contains annotation of discourse connectives categorized as primary or secondary, based in particular on the degree of their grammaticalization. For example, the expression *proto* "therefore" is taken as a primary connective because it is grammaticalized, whereas *kvůli tomu* "because of this" is perceived as a secondary connective because it is a non-grammaticalized prepositional phrase.[18]

The results of the annotation demonstrated that the PDiT contained 21416 discourse relations expressed using connectives – 20255 relations (i.e. 94.6%) were signaled by primary connectives and 1161 relations (i.e. 5.4%) by secondary connectives. The two groups of connectives thus differ greatly in frequency. Simply put, if the author has a choice, he or she prefers a primary connective to a secondary connective in the vast majority of cases.

One explanation for this may be that it is easier for authors to use shorter, mostly one-word and grammaticalized expressions than multi-word phrases that are very often inflected in Czech and may vary (cf. *důvod je* vs. *důvodem je* both meaning "the reason is" the former in nominative case, the latter in instrumental case). The preference for shorter and grammaticalized connectives conforms to the economy principle in language, specifically to the principle of least effort (Zipf, 1949), i.e. authors choose the easiest path toward the creation of a coherent text.

This idea is also supported by the fact that not even the individual primary connectives are used with the same or similar frequency – only some of them are highly frequent. The most frequent connective is *a* "and" that covers 27% of all tokens of discourse connectives in the PDiT (the PDiT contains 21416 tokens of all connectives, of which *a* "and" comprises 5766 tokens). Moreover, the three most frequent connectives (*a* "and", *však* "however" and *ale* "but") comprise 40% of all connective tokens (8554 out of 21416), the first five connectives (*a* "and", *však* "however", *ale* "but", *když* "when" and *protože* "because") comprise 45% (9653 out of 21416) and the first ten connectives (*a* "and", *však* "however", *ale* "but", *když* "when", *protože* "because", *totiž* lit. "that is", *pokud* "if", *proto* "therefore", *tedy* "so", *aby* "so that") comprise 54% (11508 out of 21416).

It is also interesting that the five most frequent connectives (*a* "and", *však* "however", *ale* "but", *když* "when" and *protože* "because") correspond semantically to the basic types of discourse relations (senses) annotated in the PDiT according to the PDTB style, namely expansion (*a* "and"), contrast (*však* "however" and *ale* "but"), temporal relations (*když* "when") and contingency (*protože* "because").[19] We may thus conclude that although authors have a variety of connectives at their disposal (the PDiT 2.0 contains 570 different forms of primary connectives and 400 forms, i.e. realizations of secondary connectives and their variants), they mostly use only ten of them. According to the PDiT data, the ten most frequent connectives cover more than a half of all cases of connective use.

5.2.2. Preferences of infrequent connectives: principle of maximal comprehensibility

In the previous subsection, we focused on the most frequent connectives. However, it is also important to look at this issue from the opposite perspective, and to pose the question of why the non-preferred connectives, the less frequent ones, exist in a language.

For example, the relation of conjunction has 195 connective forms in the PDiT data – the most frequent one is the connective *a* "and" (comprising 74% of all tokens of this relation type). Other connectives of conjunction include *také* "too", *což* "which", *dále* "further", or *rovněž* "too" (the frequency of which is dropping). Similarly, the relation of opposition is expressed mainly by the connective *však* "however" (covering 39% of all cases of this relation). However, the relation of opposition can be expressed by a total of 109 connective forms according to the PDiT data, e.g. by *ale* "but", *ovšem* "however", *sice_ale* "in fact_but", or *jenže* roughly meaning "but". At the same time, many connective forms appear only once in the PDiT, cf. 283 primary connective forms (50% out of 570), and 264 secondary connectives forms (66% out of 400) having only a single token.

The presence of these infrequent connectives in the corpus means that there exist contexts in which these connectives were preferred over the frequent ones and we need to ask why. One explanation concerns the concept of a partial synonymy, the second involves lexical diversity (or the need to avoid a lexical repetition).

Connectives belonging to the same sense type are expected to be synonyms. However, as we demonstrated in Section 5.1, they differ from one another pragmatically, stylistically, syntactically as well as slightly semantically and are therefore only

[18] The interesting thing is that both these connectives have a similar structure from the diachronic point of view. The connective *proto* "therefore" originally comes from the prepositional phrase *pro to* "for this", i.e. from the combination of a preposition and the demonstrative pronoun *to* "this" similarly to the present-day phrase *kvůli tomu* "because of this" or also *díky tomu* "thanks to this" or *navzdory tomu* "despite this"). *Proto* "therefore" gradually was grammaticalized into a one-word expression and lost its original lexical content.

[19] Of course, these connectives may also express other senses (e.g. *když* "when/if" may also be used to signal a relation of condition within the contingency category). However, our idea is that only the five most frequent connectives in Czech are sufficient to express all four basic sense categories.

partially synonymous. The existence of the infrequent connectives can be explained in the way that they can be used in specific contexts, as they can better reflect the nuances of discourse relations and thereby also better express the author's communicative intent.[20] The less frequent connectives can also be used in contexts requiring lexical diversity, i.e. where repeating the same connective over a relatively short stretch of text would be stylistically inappropriate and disruptive to the reader, as in Example (17) from the PDiT.

(17) *Máme určité kontaktní možnosti ve všech státech, odkud pocházeli zahraniční studenti v bývalém Československu.* ***Díky tomu*** *bychom tam mohli hledat uplatnění pro naše lidi, a naše licence je* ***proto*** *pojata dosti široce.*
"We have many contact opportunities in all the countries where the foreign students came from during the former Czechoslovakia. **Thanks to this**, we can look for job opportunities for our people in these countries, and **therefore**, our license is quite broad."

In Example (17), there are two discourse relations of result – the first one is expressed by the secondary connective *díky tomu* "thanks to this" and the second one by the primary connective *proto* "therefore". The repetition of the most frequent connective of result (*proto* "therefore") in both cases would cause discomfort for the reader and lower the overall text coherence, see Example (18).

(18) *?Máme určité kontaktní možnosti ve všech státech, odkud pocházeli zahraniční studenti v bývalém Československu.* ***Proto*** *bychom tam mohli hledat uplatnění pro naše lidi, a naše licence je* ***proto*** *pojata dosti široce.*
"?We have many contact opportunities in all the countries where the foreign students came from during the former Czechoslovakia. **Therefore**, we can look for job opportunities for our people in these countries, and **therefore**, our license is quite broad."

Therefore, the effort to use a broader repertoire of discourse connectives is likely caused by the author's effort to make the text as comprehensive as possible for the reader, which also corresponds to Grice's (1975) cooperative principle functioning between the author and the reader. This principle is in contrast with the principle of least effort (see Section 5.2.1). However, their coexistence is very natural in everyday communication.

5.2.3. Intra-sentential and inter-sentential connectives

The above-mentioned principles (the principles of least effort and maximal comprehensibility) are not related only to discourse connectives, but rather, they operate across the individual language layers. However, there are also specific preferences that are bound to discourse connectives. Within them, a significant opposition can be found in the intra- vs. inter-sentential use of connectives – i.e. some connectives preferably appear within a sentence, see Example (19), and others often extend beyond the sentence boundary, see Example (20).

(19) *Chtěli jsme ten film vidět,* ***ale*** *už byl vyprodaný.*
"We wanted to see the movie, **but** it was sold out."

(20) *Chtěli jsme ten film vidět. Byl* ***však*** *vyprodaný.*
"We wanted to see the movie. **However**, it was sold out."

Our investigation revealed that primary and secondary connectives differed significantly in this respect. Whereas primary connectives prefer intra-sentential relations in 70% of cases, secondary connectives tend toward inter-sentential relations in 63% of cases, see Table 2.

However, it is necessary to go further in the syntactic division of discourse connectives. For this purpose, we will first discuss the subtypes of primary connectives.

Table 2
Intra- vs. inter-sentential usage of primary and secondary connectives; the statistically significant difference (chi-square test) is marked with *** for significance level 0.001 (p-value $\leq$ 0.001).

Discourse connectives	Intra-sentential usage	Inter-sentential usage	Total
Primary***	70%	30%	20255
Secondary***	37%	63%	1161

[20] If the connectives of the same sense type were absolutely synonymous, they probably would not coexist side by side for such a long period of time because absolute synonymy tends to disappear from a natural language as it is not economic, see e.g. Cruse (2000).

Table 3
Intra- vs. inter-sentential usage of subordinating connectives; the statistically significant difference (chi-square test) is marked with *** for significance level 0.001 (p-value $\leq$ 0.001).

Most frequent subordinating connectives	Tokens in PDiT	Intra-sentential	Inter-sentential
		Tokens (%)	Tokens (%)
když "when" ***	574	574 (100%)	0 (0%)
protože "because" ***	525	518 (98%)	7 (2%)
pokud "if" ***	403	403 (100%)	0 (0%)
aby "so that" ***	305	304 (99%)	1 (1%)
-li "if" ***	248	248 (100%)	0 (0%)
zatímco "while" ***	204	203 (99%)	1 (1%)
i když "even though" ***	178	171 (96%)	7 (4%)
což "which" ***	174	170 (97%)	4 (3%)
takže roughly "so" ***	149	121 (81%)	28 (19%)
kdyby "if" ***	116	116 (100%)	0 (0%)
proto, že "because" ***	99	96 (96%)	3 (4%)
přestože "although" ***	98	98 (100%)	0 (0%)

Table 4
Intra- vs. inter-sentential usage of coordinating conjunctions and adverbs; the statistically significant difference (chi-square test) is marked with *** for significance level 0.001 (p-value $\leq$ 0.001).

Most frequent coordinating connectives	Tokens in PDiT	Intra-sentential	Inter-sentential
		Tokens (%)	Tokens (%)
Connectives preferring intra-sentential usage			
a "and" ***	5766	5428 (94%)	338 (6%)
ale "but" ***	1267	847 (66%)	420 (34%)
nebot' "for" ***	221	220 (99%)	1 (1%)
nebo "or" ***	191	167 (87%)	24 (13%)
sice_ale "in fact_but" ***	165	162 (98%)	3 (2%)
a tak "and so" ***	141	97 (68%)	44 (32%)
a to "namely" lit. "and this" ***	118	97 (82%)	21 (18%)
#neg_ale "not_but" ***	98	96 (97%)	2 (3%)
a proto "and therefore" ***	86	86 (100%)	0 (0%)
či "or" ***	86	86 (100%)	0 (0%)
avšak "but"	61	36 (59%)	25 (41%)
a pak "and then" ***	56	50 (89%)	6 (11%)
a také "and also" ***	51	46 (90%)	5 (10%)
nýbrž "but" ***	12	12 (100%)	0 (0%)
Connectives preferring inter-sentential usage			
však "however" ***	1521	266 (17%)	1255 (83%)
totiž lit. "that is" ***	460	24 (5%)	436 (95%)
proto "therefore" ***	380	41 (10%)	339 (90%)
tedy "so" ***	307	33 (10%)	274 (90%)
pak "then" ***	295	78 (26%)	217 (74%)
ovšem "however" ***	285	64 (22%)	221 (78%)
také "too" ***	234	9 (3%)	225 (97%)
navíc "moreover" ***	182	26 (14%)	156 (86%)
přitom "yet" ***	181	6 (3%)	175 (97%)
naopak "on the contrary" ***	152	29 (19%)	123 (81%)
dále "furthermore" ***	117	6 (5%)	111 (95%)
tak "so" ***	112	18 (16%)	94 (84%)
rovněž "too" ***	106	6 (5%)	100 (95%)
přesto "even though" ***	99	14 (14%)	85 (86%)
například "for example" ***	97	9 (9%)	88 (91%)
zároveň "too" ***	94	12 (12%)	82 (88%)

It is not surprising that subordinating primary connectives prefer intra-sentential usage (see Table 3 with the most frequent subordinating connectives, reaching almost 100% of intra-sentential preference).[21] Their inter-sentential usage was perceived as syntactically inappropriate (in most cases, see Section 5.1.1).

In the next step, we looked at the other types of primary connectives and divided them into two groups reflecting their preference in intra-vs. inter-sentential relations (see Table 4). Unlike subordinating connectives, many of them (like *ale* "but",

[21] The intra- and inter-sententiality of connectives on the Czech material has recently been studied by Jínová (2012) and by Zikánová et al. (2015). They point out that subordinating conjunctions are mostly used intra-sententially, whereas adverbs tend to occur inter-sententially. According to their findings, the intra- and inter-sentential proportions of coordinating conjunctions are highly dispersed. However, the description of coordinating conjunctions always depends on the part-of-speech concept being used.

a tak "and so", *však* "however" or *proto* "therefore") allow for both intra- and inter-sentential usage even though one usage prevails.

The group of connectives that prefer intra-sentential usage contains coordinating conjunctions such as *a* "and", *nebot'* "for", *nebo* "or" or *či* "or" and the group preferring inter-sentential relations consists of adverbs. Interestingly, when an adverbial connective combines with a conjunction (mostly with *a* "and"), its preference shifts from inter-sentential to intra-sentential (cf. expressions already merged into a one-word connective like ***a**le* "but", as well as thus far unmerged connectives like ***a** tak* lit. "and so", ***a** proto* "and therefore", ***a** pak* "and then" or ***a** také* "and also"). The tendency toward intra-sentential behavior is also exhibited by connectives involving negation (***ne**bo* "or", ***ne**bot'* "for", ***#neg**_ale* "not_but", ***ný**brž* "but") and by correlative pairs of connectives (*sice_ale* "in fact_but").

Traditionally, discourse connectives have been divided into subordinating and coordinating. However, our results demonstrate that this traditional division is insufficient. The opposition of inter-sentential vs. intra-sentential connectives should be reflected as well. The relationship between them can be described as follows.

The intra-sentential primary connectives are i) subordinating connectives like *když* "when", *protože* "because" or *pokud* "if", ii) coordinating connectives: basic ones like *a* "and" or *či* "or" (as well as connectives that have combined with them such as *ale* "but", *a proto* "and therefore"), connectives involving negation (***ne**bo* "or", ***ne**bot'* "for", ***#neg**_ale* "not_but", ***ný**brž* "but"), and forming correlative pairs (*sice_ale* "in fact_but"). The inter-sentential primary connectives are adverbs (*dále* "furthermore", *navíc* "moreover", *však* "however", *proto* "therefore" or *přitom* "yet").

Secondary connective forms have been described in detail in Rysová (2012). Secondary connectives function as sentence elements (*kvůli tomu* "because of this") or sentence modifiers (*jednoduše řečeno* "simply speaking"). Very often, they function as adverbials (cf. *kvůli tomu* "because of this" is an adverbial of reason) and are thus similar to the adverbial primary connectives that had the same function historically (cf. the present-day connective *proto* "therefore" whose form was originally used as an adverbial of reason). Interestingly, both of these groups of connectives (secondary connectives and primary connectives with adverbial character) tend to be used inter-sententially.

In simple words, discourse connectives in the form of conjunctions (both subordinating and coordinating) are typically used intra-sententially, whereas they are used inter-sententially when in the form of adverbs and secondary adverbial phrases.

These findings correspond to the recent description of intra- and inter-sentential connectives by Danlos (2016), who delimits them on the basis of embeddability (intra-sentential connectives appear in discourse segments that can be embedded under a matrix clause, whereas inter-sentential connectives do not). She concludes that subordinating and coordinating conjunctions are typical intra-sentential connectives, and adverbs and adverbial prepositional phrases (such as *in summary*) are typical inter-sentential ones. We have supported these conclusions using a novel source of data (PDiT) with frequencies for the individual groups, and enriched them with the class of secondary connectives.

6. Conclusions

In this paper, we have introduced the constraints on and preferences in the use of discourse connectives in written Czech texts. We have demonstrated that although discourse connectives are suitable for many contexts with the relevant sense type (which we consider to be a crucial condition for classification as a discourse connective), they are not 100% context-independent. Based on our analysis, we draw the following conclusions.

I. Constraints

* **Syntax.** We have demonstrated that the use of subordinating connectives is bound to intra-sentential relations. Their inter-sentential usage was marked as inappropriate by the annotators.

* **Word order.** Discourse connectives differ in their word order tendencies. Subordinating conjunctions are typically restricted to the first position in the argument; similarly, basic coordinating conjunctions are typically bound to the position between arguments. On the other hand, most adverbial connectives (both primary and secondary) can occupy the first as well as the second position in the argument. The connective *takže* roughly "so" from our experiment is restricted to the first position in the second discourse argument, i.e. in between Arg1 and Arg2. The use of this connective in a different position was marked as inappropriate by the annotators.

* **Semantics**. The individual connectives (from the same semantic category according to the PDiT list of senses) differ in their semantic nuances. While the connectives may be used as contextual synonyms in some contexts (*The football game ran over by 15 min so / because of this some fans did not catch the regular train departure.*), in other cases they are not interchangeable (*I'm thirsty, so / ?because of this I'll go get a drink of water.*). Some connectives have thus a broader sense than the others.

The narrower sense is often expressed by modified connectives, which is especially a feature of the secondary ones. In their modified form, connectives have a more specific sense, cf. *hlavním/možným/dobrým důvodem je* "the main/possible/good reason is", which limits their use in many contexts.

* **Pragmatics.** Connectives with strong pragmatic connotations are greatly limited in their use. For example, the connective *vinou toho* lit. "by fault of this" is strongly bound to negative contexts. It is interesting that the connective *díky tomu* "thanks to this" can occasionally be used, on the other hand, in both positive and negative contexts according to the PDiT data. The pragmatic aspect of the connective *díky tomu* "thanks to this" is thus growing weaker.

* **Stylistics**. Connectives also differ stylistically – they can be neutral (*proto* "thus"), formal (*z uvedeného důvodu* "for the given reason") or informal (*teda* "thusly", *tak* "so"). Stylistically neutral connectives are used in a broad range of contexts. Formal and informal connectives are restricted to particular (stylistically appropriate) contexts. The mixing of formal connectives and informal contexts (and vice versa) was marked as inappropriate by the annotators (cf. ?*Mom, the tea is hot. For the given reason, I'm drinking it carefully.*).

* **Long distance constraint: Salience**. Most discourse relations occur between adjacent or not very distant arguments. However, the texts contain several cases of discourse relations extending over a number of sentences, i.e. in which Arg1 and Arg2 are separated by another, longer text. This long distance between discourse arguments impedes the use of a general connective, which could lead to the misinterpretation of the given relation and thus disrupt the coherence of the whole text. In these contexts, the author uses a free connecting phrase (e.g. *kvůli tomuto počasí* "due to this weather") rather than a contextually independent connective (either primary or secondary). In the case of long distance relations, the major function of the free connecting phrase (containing a highly explicit anaphora) is to topicalize and recall a certain object in order to keep it activated in the reader's consciousness. In these cases, it is necessary to recall the correct argument, not only to refer to it using a general connective. Our investigation revealed that the author's decision whether to use a contextually independent connective or a contextually dependent free connecting phrase is motivated mainly by the effort to avoid misunderstandings.

II. Preferences

* **Language economy: Principle of least effort**. The corpus data has demonstrated that discourse relations in written journalistic texts in Czech are expressed more frequently by primary connectives (95%) than by secondary ones (5%). Primary connectives are short, grammaticalized and lexically stable expressions, while secondary connectives are mostly longer, formally diverse and internally modifiable. For language users, it is thus probably easier and more convenient to use shorter and more stable expressions.

Our analysis has also revealed that although authors have a broad repertoire of connectives at their disposal (PDiT 2.0 contains 570 various forms of primary connectives and 400 forms of secondary connectives), they prefer to use only some of them, mainly the ten most frequent ones[22] that comprise more than half (54%) of all tokens of connectives in the PDiT.

At the same time, the five most frequent connectives cover the four basic sense categories annotated in the PDTB and subsequently in the PDiT: expansion: *a* "and"; contrast: *však* "however", *ale* "but"; temporal: *když* "when"; contingency: *protože* "because". This information can be used, for example, in teaching foreign languages. For the communicative needs of a foreign language learner, it is very effective to learn the most frequent discourse connectives.

* **Infrequent connectives: Principle of maximal comprehensibility.** Since a language has a wide repertoire of discourse connectives, of which only a small set is used at a high frequency, it is necessary to explain the existence of the infrequent ones. The repertoire of connectives is extensive for each sense for two particular reasons. First, there is a need for partial connective synonyms (differing syntactically, pragmatically, stylistically, and, in some cases, semantically) that can reflect all the subtle shades of discourse relations and thereby better express the author's communicative intent. The infrequent connectives are often very specific in some way and therefore, they fit well into specific (infrequent) contexts. Second, less frequent connectives enrich the lexicon and thus enable the author to avoid the inconvenient repetition of a single connective over a short stretch of text.

* **Intra-sentential connectives and inter-sentential connectives**. When dealing with the syntactic properties of connectives, it is necessary to divide them into subclasses. Our analysis has shown that the division of connectives into subordinating and coordinating groups is not enough – it is also necessary to divide them into intra-sentential (*I will meet my friend and we will go to the movies.*) and inter-sentential (*He is vegetarian. Therefore, he doesn't eat meat.*) classes.

Intra-sentential connectives, predictably, are typically subordinating conjunctions (like *protože* "because", *když* "when" or *jestli* "if"). In addition, intra-sentential usage is preferred by coordinating conjunctions. In general, the intra-sentential primary connectives can be described as i) subordinating connectives and ii) coordinating connectives: the basic coordinating conjunctions such as *a* "and" or *či* "or" (also in combination with adverbial connectives, either historically merged, such as *ale* "but", or not yet merged, like *a proto* "and therefore" or *a pak* "and then"), connectives involving negation (*nebo* "or", *nebot'* "for", *#neg_ale* "not_but", *nýbrž* "but"), and forming correlative pairs (*sice_ale* "in fact_but").

The inter-sentential primary connectives are adverbs (*dále* "furthermore", *navíc* "moreover", *však* "however", *proto* "therefore" or *přitom* "yet").

Secondary connectives often function as adverbials in a sentence (e.g. *z tohoto důvodu* "for this reason" or *kvůli tomu* "because of this" are adverbials of reason). In this respect, they are similar to the class of primary connectives with adverbial character that originally also had the same function in the sentence. At the same time, the connectives with adverbial character (both primary and secondary) tend toward inter-sentential usage.

Put simply, discourse connectives in the form of conjunctions (both subordinating and coordinating) are typically used intra-sententially, whereas they are typically used inter-sententially when in the form of adverbs and secondary adverbial phrases.

[22] *A* "and", *však* "however", *ale* "but", *když* "when", *protože* "because", *totiž* lit. "that is", *pokud* "if", *proto* "therefore", *tedy* "so", *aby* "so that".

Based on our investigation, we conclude that when selecting a proper connective, authors are influenced by three general factors: i) the conventions and grammatical rules of the given language, ii) the principle of least effort, and iii) the effort to avoid misunderstandings. All of these factors combine in the creation of the text so that the result is maximally coherent.

Acknowledgment

We acknowledge support from the Czech Science Foundation project no. GA17-06123S (Anaphoricity in Connectives: Lexical Description and Bilingual Corpus Analysis). This study has utilized the language resources distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (project LM2015071).

References

Afantenos, S.D., Asher, N., Benamara, F., et al., 2012. An empirical resource for discovering cognitive principles of discourse organization: the ANNODIS corpus. In: Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). European Language Resources Association (ELRA), Istanbul, pp. 2727–2734.

Aijmer, K., 2002. English discourse particles. Evidence from a corpus. In: Studies in Corpus Linguistics 10. John Benjamins, Amsterdam/Philadelphia.

Asher, N., 1993. Reference to Abstract Objects in Discourse. Kluwer Academic Publishers, Dordrecht.

Asher, N., Lascarides, A., 2003. Logics of Conversation. Cambridge University Press.

Carlson, L., Marcu, D., Okurowski, M.E., 2001. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In: Proceedings of the 2nd SIGDIAL Workshop on Discourse and Dialogue, Eurospeech.

Cruse, D.A., 2000. Meaning in Language: an Introduction to Semantics and Pragmatics. OUP, Oxford.

Danlos, L., 2016. Building Discourse Relational Device Lexicons. TextLing Training school, Valencia, Spain. https://hal.inria.fr/hal-01392824/document.

Danlos, L., 2006. Discourse verbs and discourse periphratic links. In: Second International Workshop on Constraints in Discourse, Maynoth, Irelande.

Danlos, L., et al., 2012. Vers le FDTB: French Discourse Tree Bank. Actes de la conférence conjointe JEP-TALN-RECITAL. TALN. Association Francophone pour la Communication Parlée (AFCP) et Association pour le Traitement Automatique des Langues (ATALA), Grenoble, pp. 471–478.

Degand, L., 2000. Causal connectives or causal prepositions? Discursive constraints. J. Pragmat. 32 (6), 687–707.

Degand, L., Fagard, B., 2012. Competing connectives in the causal domain: French *car* and *parce que*. J. Pragmat. 44 (2), 154–168.

Fischer, K., 2006. Approaches to Discourse Particles. Elsevier, Amsterdam.

Fraser, B., 1990. An approach to discourse markers. J. Pragmat. 14, 383–395.

Fraser, B., 1999. What are discourse markers? J. Pragmat. 31 (7), 931–952. Elsevier.

Fraser, B., 2015. The combining of discourse markers – a beginning. J. Pragmat. 86, 48–53.

Fuentes-Rodríguez, C., Placencia, M.E., Palma-Fahey, M., 2016. Regional pragmatic variation in the use of the discourse marker *pues* in informal talk among university students in Quito (Ecuador), Santiago (Chile) and Seville (Spain). J. Pragmat. 97, 74–92.

Gonen, E., Livnat, Z., Amir, N., 2015. The discourse marker *axshav* ('now') in spontaneous spoken Hebrew: discursive and prosodic features. J. Pragmat. 89, 69–84.

Grice, H.P., 1975. Logic and conversation. Syntax and Semantics 3: Speech Acts. Academic Press, New York, pp. 41–58.

Hajič, J., Panevová, J., Hajičová, E., Sgall, P., Pajas, P., Štěpánek, J., Havelka, J., Mikulová, M., Žabokrtský, Z., Ševčíková-Razímová, M., Urešová, Z., 2006. Prague Dependency Treebank 2.0. Linguistic Data Consortium, Philadelphia, USA.

Hajičová, E., Havelka, J., Sgall, P., 2004. Topic and focus, anaphoric relations and degrees of salience. In: Prague Linguistic Circle Papers/Travaux du cercle linguistique de Prague N.S. John Benjamins, Amsterdam.

Hajičová, E., Sgall, P., Buráňová, E., 2003. Topic-focus articulation and degrees of salience in the Prague dependency treebank. In: Formal Approaches to Function in Grammar. In Honor of Eloise Jelinek Arizona. John Benjamins, Amsterdam/Philadelphia, pp. 165–177.

Hakulinen, A., 1998. The Use of Finnish *nyt* as a Discourse Particle. Discourse Markers. Description and Theory. John Benjamins Publishing Company, Amsterdam/Philadelphia, pp. 83–96.

Halliday, M.A.K., Hasan, R., 1976. Cohesion in English. Longman, London.

Hansen, M.-B.M., 1998. The Function of Discourse Particles. A Study with Special Reference to Spoken Standard French. John Benjamins, Amsterdam/Philadelphia.

Harris, Z.S., 1952. Discourse analysis. Language 28 (1), 1–30.

Jínová, P., 2012. Nejčastější konektivní prostředky kauzálního vztahu v Pražském závislostním korpusu (The Most Common Connective Means of Causal Relation in the Prague Dependency Treebank). Studie z aplikované lingvistiky/Stud. Appl. Ling. (SALi) 1, 35–52.

Mann, W.C., Thompson, S.A., 1988. Rhetorical structure theory: toward a functional theory of text organization. Text 8 (3), 243–281.

Marmorstein, M., 2016. Getting to the point: the discourse marker *yaʕni* (lit. "it means") in unplanned discourse in Cairene Arabic. J. Pragmat. 96, 60–79.

Maschler, Y., 2000. Discourse Markers in Bilingual Conversation. Kingston Press, Middlesex.

Poláková, L., Jínová, P., Zikánová, Š., Hajičová, E., Mírovský, J., Nedoluzhko, A., Rysová, M., Pavlíková, V., Zdeňková, J., Pergler, J., Ocelák, R., 2012a. Prague Discourse Treebank 1.0. Data/software. ÚFAL, MFF, Charles University, Prague, Czechia.

Poláková, L., Jínová, P., Zikánová, Š., Bedřichová, Z., Mírovský, J., Rysová, M., Zdeňková, J., Pavlíková, V., Hajičová, E., 2012b. Manual for Annotation of Discourse Relations in Prague Dependency Treebank. Technical report no. 2012/47. ÚFAL, MFF, Charles University, Prague, Czechia, pp. 1–83.

Prasad, R., Lee, A., Dinesh, N., Miltsakaki, E., Campion, G., Joshi, A., Webber, B., 2008. The Penn Discourse Treebank 2.0. In: Proceedings of the 6th International Conference on Language Resources and Evaluation. Marrackech, Morocco.

Prasad, R., Joshi, A., Webber, B., 2010. Realization of discourse relations by other means: alternative lexicalizations. In: Coling 2010: Posters (August 2010), pp. 1023–1031.

Reese, B., Hunter, J., Denis, P., Asher, N., Baldridge, J., 2007. Reference Manual for the Analysis and Annotation of Rhetorical Structure, Technical report. Department of Linguistics, The University of Texas, Texas, Austin.

Rysová, M., 2012. Alternative lexicalizations of discourse connectives in Czech. In: Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12). European Language Resources Association (ELRA), Istanbul, Turkey, pp. 2800–2807.

Rysová, M., 2015. Diskurzní konektory v češtině (Od centra k periferii) (Discourse Connectives in Czech (From Centre to Periphery)) (Ph.D. thesis). Charles University in Prague, Prague, Czechia.

Rysová, M., 2017. Discourse connectives: from historical origin to present-day development. In: Menzel, K., et al. (Eds.), New Perspectives on Cohesion and Coherence. Language Science, Berlin, Germany, pp. 11–35.

Rysová, M., Rysová, K., 2014. The centre and periphery of discourse connectives. In: Aroonmanakun, W., et al. (Eds.), Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing (PACLIC 28). Department of Linguistics, Faculty of Arts, Chulalongkorn University, Bangkok, Thailand, pp. 452–459.

Rysová, M., Rysová, K., 2015. Secondary connectives in the Prague dependency treebank. In: Hajičová, E., Nivre, J. (Eds.), Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015). Uppsala University, Uppsala, Sweden, pp. 291–299.

Rysová, M., Synková, P., Mírovský, J., Hajičová, E., Nedoluzhko, A., Ocelák, R., Pergler, J., Poláková, L., Pavlíková, V., Zdeňková, J., Zikánová, Š., 2016. Prague Discourse Treebank 2.0. Data/software. ÚFAL, MFF, Charles University, Prague, Czechia. http://hdl.handle.net/11234/1-1905.

Shloush, S., 1998. A Unified Account of Hebrew Bekicur in Short: Relevance Theory and Discourse Structure Considerations. Discourse Markers: Descriptions and Theory. John Benjamins Publishing Company, Amsterdam, Philadelphia, pp. 61–82.
Schiffrin, D., 1987. Discourse Markers. Cambridge University Press, Cambridge.
Smith-Christmas, C., 2016. Regression on the fused lect continuum? Discourse markers in Scottish Gaelic–English speech. J. Pragmat. 94, 64–75.
Soria, C., 2005. Constraints on the Use of Connectives in Discourse. Manuscrito no publicado. Istituto de Linguistica Computazionale (CNR), Pisa, Italy.
Sperber, D., Wilson, D., 1986. Relevance. Harvard University Press, Cambridge, Massachusetts.
Stede, M., Neumann, A., 2014. Potsdam commentary corpus 2.0: annotation for discourse research. In: Calzolari, N., et al. (Eds.), Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC'14). European Language Resources Association (ELRA), Reykjavik, pp. 925–929.
Tanghe, S., 2016. Position and polyfunctionality of discourse markers: the case of Spanish markers derived from motion verbs. J. Pragmat. 93, 16–31.
Thaler, V., 2016. Italian mica and its use in discourse: an interactional account. J. Pragmat. 103, 49–69.
Urgelles-Coll, M., 2010. Continuum Studies in Theoretical Linguistics: Syntax and Semantics of Discourse Markers. Continuum International Publishing, London.
van Dijk, T.A., 1979. Pragmatic connectives. J. Pragmat. 3, 447–456.
Zeyrek, D., Turan, Ü.D., Demirsahin, I., Çakici, R., 2012. Differential properties of three discourse connectives in Turkish: a corpus-based analysis of Fakat, Yoksa, Ayrıca. In: Constraints in Discourse 3, vol. 223. John Benjamins Publishing, pp. 183–206.
Zikánová, Š., Hajičová, E., Hladká, B., Jínová, P., Mírovský, J., Nedoluzhko, A., Poláková, L., Rysová, K., Rysová, M., Václ, J., 2015. Discourse and Coherence. From the Sentence Structure to Relations in Text. ÚFAL, MFF, Charles University, Prague, Czechia.
Zipf, G.K., 1949. Human Behavior and the Principle of Least Effort. Addison-Wesley Press, Cambridge (Mass.).
Zwicky, A.M., 1985. Clitics and particles. Language 61 (2), 283–305.

PhDr. Magdaléna Rysová, Ph.D. works as a senior research associate at the Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Prague. Her main research interest is discourse analysis with orientation on discourse connectives and their annotation in large corpora. She focuses especially on secondary discourse connectives and their inclusion in discourse lexicons. She is the main author of the Prague Discourse Treebank 2.0.

PhDr. Kateřina Rysová, Ph.D. is a senior research associate at the Institute of Formal and Applied Linguistics (Charles University, Prague). Her research topics are discourse analysis, sentence information structure / topic-focus articulation, word order, dependency syntax and corpus linguistics. She got the Bolzano Prize (2013) for the best Ph.D. thesis at the Charles University in Prague in the category of social science. She is the author of the book *On Word Order from the Communicative Point of View*. She is the leader of a research team developing the software application EVALD (Evaluator of Discourse) on automatic evaluation of text coherence in Czech.