

Discourse Connectives and Reference

Kateřina Rysov and Magdalna Rysov

Charles University, Prague, Czech Republic
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
[magdalena.rysova|rysova]@ufal.mff.cuni.cz

Abstract. In the present paper, we examine discourse connectives from the perspective of reference (i.e. a presence of an anaphoric element). We introduce a division of connectives into: i) connectives without an inherent (internal) reference (e.g. *and, but, or, if, however, so*), and ii) connectives with an inherent (internal) reference that is either optional (e.g. *as a result* vs. *as a result of this*), or obligatory – cf. already grammaticalized connectives (e.g. *thereafter, therefore* or *thereby*) vs. not yet grammaticalized connectives (e.g. *because of this* or *for this reason*). We apply this general division on Czech and German connectives and conduct a contrastive study on the parallel data of the corpus InterCorp 10. Specifically, we focus on the group of Czech connectives in the form of prepositional phrases with an obligatory inherent reference that do not have any fully grammaticalized form in Czech (like *krom toho*, lit. ‘‘except this’, ‘moreover’) and we search for their most frequent semantic counterparts in German. The results of our research demonstrate that the German counterparts of the selected connectives in Czech are mostly (in 72%) grammaticalized connectives containing a referential morpheme (e.g. *auerdem, deswegen, stattdessen, dagegen, demgegenber, daneben, infolgedessen*).

Keywords: Discourse Connectives, Reference, Anaphoric Connectives.

1 Introduction

Semantic discourse relations as well as coreference relations substantially participate in creating a coherent text. Both of them belong to the basic cohesive relations with an ability to form cohesive ties and chains (see Halliday and Hasan, 1976). Semantic discourse relations may be signaled explicitly by discourse connectives or they may be only implicit (details on borderlines between explicit and implicit discourse relations are given in Taboada, 2009). Coreference relations are realized very often through demonstrative and personal pronouns – a detailed description of coreference and anaphoric realizations in Czech is presented in Nedoluzhko (2011) or more recently in Ziknov et al. (2015).

It is interesting that many discourse connectives also contain an anaphoric element (like *therefore, thereby* etc.). In this way, semantic and coreference relations are mutually interconnected and investigation of their relationship is essential for text coher-

ence in general (the need for studying coherence through interplays is addressed e.g. by Hajičová, 2011 or Nedoluzhko and Hajičová, 2015).

In the present paper, we aim to examine the discourse connectives containing a referential (anaphoric) component. Specifically, we divide discourse connectives into several groups according to their ability to express anaphora in the surface structure and we present a contrastive analysis of Czech and German connectives containing an explicit anaphoric element like *kromě toho* – *außerdem* lit. “except this”, ‘moreover’, or *kvůli tomu* – *deswegen* “because of this”.

2 Discourse Connectives: Description and Delimitation

Generally, a discourse connective is defined as a predicate of a binary relation opening two positions for two text spans as its arguments and signaling a semantic or pragmatic relation between them (Prasad et al., 2008). Discourse connectives may be further divided into primary and secondary (Rysová and Rysová, 2014 and 2015), the groups differing in the degree of grammaticalization – cf. the grammaticalized primary connectives (e.g. *and*, *but*, *however*, *therefore*) and not yet fully grammaticalized secondary ones (e.g. *for this reason*, *on condition that*).

From the perspective of anaphora and discourse structure, a description of connectives is given in Webber et al. (2003) who distinguish between anaphoric connectives (mostly certain adverbials; picking up their external argument by means of anaphora resolution) and structural connectives (taking arguments qua the syntactic configuration they appear in). Anaphoric connectives in German were studied by Stede and Grishina (2016) who focused on the description of a group of German connectives containing a morpheme overtly referring backward (e.g. *demzufolge*). Anaphoric connectives in Czech are rather an unexplored topic – the first probe was carried out by Poláková et al. (2012) exploring a subgroup of these expressions in the form of a preposition and a demonstrative pronoun.

3 Discourse Connectives and Reference: General Overview

As mentioned above, a general property of discourse connectives is to connect two text units. Thus, if a discourse connective appears in a text, we assume that it somehow refers to the previous context, i.e. the presence of a connective implies the presence of the first discourse argument (see Halliday and Hasan, 1976). In this respect, discourse connectives and (co)reference relations are strongly inter-related – all discourse connectives may be viewed as implicitly referential (e.g. connectives like *but*, *and*, *or* do not contain any anaphoric element but still they signal a presence of the first discourse argument).

At the same time, within the discourse connectives, there is a narrower set of expressions containing a referential (anaphoric) element explicitly, cf. examples of secondary connectives like *because of this*, *after this*, *as a result of this*, *this is the reason why*, *under these conditions*, *for this reason* etc. However, also these expressions differ from each other, as some of them (e.g. *because of this*) contain the anaphoric

element obligatorily while some of them only optionally (cf. *as a result* vs. *as a result of this*). Concerning the presence of a referential (anaphoric) element, connectives may be thus divided into the following groups:

1) connectives **without an inherent (internal) reference** (e.g. *and, but, or, if, however, so*);

2) connectives **with an inherent (internal) reference** that is:

2a) **optional** (e.g. *as a result* vs. *as a result of this*);

2b) **obligatory**

- already grammaticalized connectives (e.g. *thereafter, therefore, thereby*);
- not yet grammaticalized connectives (e.g. *because of this, for this reason*).

The way of expressing reference in connectives may differ across languages, which is noticeable especially on semantic equivalents (cf. e.g. Czech *místo toho* vs. English *instead* vs. German *stattdessen*). For example, we cannot use prepositions without reference as discourse adverbs in Czech, which is possible in English, see Examples (1a) and (1b) from the parallel corpus InterCorp 10 (Rosen and Vavřín, 2017).

In the Czech example (1a), the discourse relation is expressed by the connective *místo toho* (lit. “instead of this”) that cannot be used without the anaphoric part *toho* “this”. Example (1a) without *toho* is ungrammatical (**Místo dál pochodovala...*). On the contrary, such usage of *instead* in the English version (1b) is fully functional.

(1a) Czech: *Ale i když měla pokušení koupit si dlouhé černé šaty, které viděla viset v butiku Betsey Johnsonové, nevešla ani dovnitř. **Místo toho** dál pochodovala jednou z uliček tam a druhou zase zpátky.*

(1b) English: *But even though she’s tempted by a long black dress she sees hanging on the far wall in Betsey Johnson, she doesn’t go inside. **Instead**, she continues trance-like up one street and down another.*

(1c) German: *Selbst als sie in einem Betsey-Johnson-Shop ein langes schwarzes Kleid hängen sieht, das sie reizen könnte, betritt sie den Laden nicht. **Stattdessen** schreitet sie wie in Trance eine Straße nach der anderen ab.*

Examples (1a) and (1c) illustrate that referential connectives may differ also in the degree of grammaticalization – whereas Czech *místo toho* (lit. “instead of this”) is not yet fully grammaticalized, its German anaphoric counterpart *stattdessen* is fully lexicalized as a one-word connective.

4 Referential Connectives in Czech and German

In our study, we focus on the group of Czech secondary connectives in the form of prepositional phrases with an obligatory inherent reference (representing the most frequent set of expressions in the group 2b in the scheme above) that do not have any fully grammaticalized form in Czech (like *kromě toho*, lit. “except this”, ‘moreover’). We select 10 most typical representatives of these referential connectives in Czech (listed in Table 1)¹ and we search for their most frequent semantic counterparts in German based on the parallel data of the corpus InterCorp 10 (using the Treq tool, see <http://treq.korpus.cz/>). The German counterparts were firstly found automatically in InterCorp 10 and then sorted out manually.

In the first step, we examine how often (in total numbers) these non-grammaticalized referential connectives in Czech are expressed as grammaticalized referential connectives in German (see Table 1), i.e. how many corpus occurrences correspond to the relation between Czech and German connectives demonstrated on Examples (1a) and (1c).

Table 1. Percentage of German counterparts of Czech connectives like *kromě toho* (lit. “except this”, ‘moreover’) in InterCorp.

Czech non-grammaticalized referential connectives	German grammaticalized referential equivalents		Other German equivalents	
	Occurrences in InterCorp	%	Occurrences in InterCorp	%
<i>kromě toho</i> “except this” ‘moreover’	5,671	71%	2,328	29%
<i>naproti tomu</i> “in contrast to this”	924	74%	324	26%
<i>místo toho</i> “instead of this”	708	84%	139	16%
<i>kvůli tomu</i> “because of this”	576	74%	205	26%
<i>navzdory tomu</i> “in contrast to this”	141	57%	106	43%
<i>díky tomu</i> “thanks to this”	156	71%	64	29%
<i>vedle toho</i> “besides this”	103	84%	19	16%
<i>oproti tomu</i> “in contrast to this”	109	94%	7	6%
<i>vzhledem k tomu</i> “with regard to this”	39	49%	40	51%
<i>na rozdíl od toho</i> “in contrast to this”	0	0%	9	100%
In Total	8,427	72%	3,241	28%

¹ The selection was based on the language material of the Prague Discourse Treebank 2.0 (PDiT 2.0; Rysová et al., 2016), a corpus containing manual discourse annotation of both primary and secondary connectives.

Table 1 demonstrates that non-grammaticalized referential connectives in Czech (like *naproti tomu*, *vedle toho*) are expressed as grammaticalized referential connectives in German (like *dagegen*, *daneben*) in 72%, i.e. German grammaticalized variants are the preferable ones in these cases.

In the next step, we analyse the individual German equivalents for the selected connectives in Czech in more detail, see Table 2.

The results of our research demonstrate that the selected connectives in Czech have diverse counterparts in German. In most cases (in 72%), these German counterparts are grammaticalized primary connectives containing a referential morpheme (e.g. *außerdem*, *deswegen*, *stattdessen*, *dagegen*, *demgegenüber*, *daneben*, *infolgedessen*). However, in some cases, a primary connective without an inherent reference is also used (e.g. *auch*, *obwohl*, *doch*). Some German counterparts in InterCorp are also non-grammaticalized secondary connectives, very often containing an explicit reference (cf. *abgesehen davon*, *hinzu kommt*, *ergänzend dazu*, *dessen ungeachtet*, *im Gegensatz dazu*, *angesichts dessen*).

Table 2. List of German counterparts of Czech connectives like *kromě toho* (lit. ‘except this’, ‘moreover’) in InterCorp.

Czech connectives	German equivalents (occurrences in InterCorp)
<i>kromě toho</i> lit. ‘except this’ ‘moreover’	<i>außerdem</i> (3,114), <i>darüber hinaus</i> (1,413), <i>auch</i> (1,059), <i>zudem</i> (595), <i>ferner</i> (589), <i>zusätzlich</i> (174), <i>überdies</i> (169), <i>des Weiteren</i> (149), <i>dazu</i> (149), <i>im Übrigen</i> (111), <i>weiterhin</i> (96), <i>abgesehen davon</i> (78), <i>daneben</i> (75), <i>hinzu kommt</i> (63), <i>ebenso</i> (53), <i>außer + NP</i> (49), <i>übrigens</i> (26), <i>nebenbei</i> (19), <i>desgleichen</i> (7), <i>weitere</i> (7), <i>ergänzend dazu</i> (4)
<i>naproti tomu</i> ‘in contrast to this’	<i>dagegen</i> (549), <i>hingegen</i> (300), <i>im Gegensatz dazu/hierzu</i> (238), <i>andererseits</i> (75), <i>demgegenüber</i> (69), <i>im Gegenteil</i> (11), <i>trotzdem</i> (6)
<i>místo toho</i> ‘instead of this’	<i>stattdessen</i> (689), <i>vielmehr</i> (98), <i>anstatt + NP</i> (41), <i>dagegen</i> (19)
<i>kvůli tomu</i> ‘because of this’	<i>wegen + NP</i> (199), <i>deswegen</i> (191), <i>deshalb</i> (142), <i>dafür</i> (119), <i>darüber</i> (64), <i>dazu</i> (33), <i>darum</i> (27), <i>aufgrund + NP</i> (6)
<i>navzdory tomu</i> ‘in contrast to this’	<i>trotzdem</i> (79), <i>dennoch</i> (62), <i>trotz + NP</i> (53), <i>doch</i> (19), <i>obwohl</i> (10), <i>trotz allem</i> (11), <i>allerdings</i> (9), <i>dessen ungeachtet</i> (4)
<i>díky tomu</i> ‘thanks to this’	<i>damit</i> (50), <i>dadurch</i> (47), <i>durch + NP</i> (38), <i>deshalb</i> (22), <i>dank + NP</i> (20), <i>infolgedessen</i> (17), <i>somit</i> (13), <i>deswegen</i> (7), <i>aufgrund + NP</i> (6)
<i>vedle toho</i> ‘besides this’	<i>daneben</i> (67), <i>außerdem</i> (25), <i>zudem</i> (11), <i>zusätzlich</i> (7), <i>nebenbei</i> (7), <i>andererseits</i> (5)
<i>oproti tomu</i> ‘in contrast to this’	<i>dagegen</i> (72), <i>hingegen</i> (28), <i>demgegenüber</i> (9), <i>im Gegensatz dazu/hierzu</i> (7)
<i>vzhledem k tomu</i> ‘with regard to this’	<i>daher</i> (39), <i>angesichts dessen</i> (34), <i>im Hinblick darauf</i> (4), <i>infolgedessen</i> (2)
<i>na rozdíl od toho</i> ‘in contrast to this’	<i>im Gegensatz dazu/hierzu</i> (9)

5 Conclusion

In our paper, we focused on the interaction of discourse connectives and (co)reference. We divided connectives into several general groups according to whether they contain an inherent reference (*and* vs. *therefore*), whether the reference is optional or obligatory (*as a result (of this)* vs. *because of this*) and whether the connectives with the obligatory reference are already grammaticalized (*therefore* vs. *for this reason*). This general description works for connectives across languages but languages differ in preferences of the individual groups. These differences are especially noticeable if they concern semantic equivalents.

In our study, we further focused on referential connectives in Czech and German (and slightly in English) in parallel data of the corpus InterCorp 10. We demonstrated that there is a group of semantic equivalents of connectives with a similar structure in Czech, German and English that differ right in this referential aspect, cf. Czech *místo toho* belonging to the group of non-grammaticalized connectives with an obligatory reference, German anaphoric *stattdessen* that is already grammaticalized and English *instead* that is typically used without an explicit reference.

Based on the Czech-German analysis, we conclude that the German counterparts of Czech non-grammaticalized referential connectives (like *naproti tomu*, *vedle toho*) are mostly (in 72%) grammaticalized referential connectives (like *dagegen*, *daneben*). From this point of view, there is a stronger tendency to grammaticalization of referential connectives in German than in Czech.

Acknowledgements

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

1. Hajičová, E.: On interplay of information structure, anaphoric links and discourse relations. *Societas linguistica europaea, SLE 2011, 44th Annual Meeting*, Book of Abstracts, pp. 139–140. Copyright © Universidad de la Rioja, Center for Research in the Applications of Language, Logrono, Spain (2011).
2. Halliday, M. A. K., Hasan, R.: *Cohesion in English*. London: Longman (1976).
3. Nedoluzhko, A.: *Rozšířená textová koreference a asociální anafora (Koncepte anotace českých dat v Pražském závislostním korpusu)*. Copyright © Ústav formální a aplikované lingvistiky, Praha, Česká republika, ISBN 978-80-904571-2-6, 268 pp. (2011).
4. Nedoluzhko, A., Hajičová, E.: *Information Structure and Anaphoric Links – A Case Study and Probe*. Contributed talk, Corpus Linguistics 2015 (CL2015), Lancaster University, UK, Lancaster, UK (2015).

5. Poláková, L., Jínová, P., Mirovský, J.: Interplay of Coreference and Discourse Relations: Discourse Connectives with a Referential Component. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pp. 146-153. Copyright © European Language Resources Association, İstanbul, Turkey, ISBN 978-2-9517408-7-7 (2012).
6. Prasad, R., Dinesh, N., Lee, A., Mitsakaki, E., Robaldo, L., Joshi, A., Webber B.: The Penn Discourse Treebank 2.0. In: *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC '08)*, pp. 2961-2968. Marrakech, Morocco, ISBN 2-9517408-4-0 (2008).
WWW: <http://www.lrec-conf.org/proceedings/lrec2008/pdf/754_paper.pdf>.
7. Rosen, A., Vavřin, M.: *Czech National Corpus – InterCorp*, Institute of the Czech National Corpus, Prague. 5. 2. 2018. WWW: <<http://www.korpus.cz>>.
8. Rysová, M., Rysová, K.: Secondary Connectives in the Prague Dependency Treebank. In: *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pp. 291-299. Copyright © Uppsala University, Uppsala, Sweden, ISBN 978-91-637-8965-6 (2015).
9. Rysová, M., Rysová, K.: The Centre and Periphery of Discourse Connectives. In: *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing*, pp. 452-459. Copyright © Department of Linguistics, Faculty of Arts, Chulalongkorn University, Bangkok, Thailand, ISBN 978-616-551-887-1 (2014).
10. Rysová, M., Synková, P., Mirovský, J., Hajičová, E., Nedoluzhko, A., Ocelák, R., Pergler, J., Poláková, L., Pavlíková, V., Zdeňková, J., Zikánová, Š.: *Prague Discourse Treebank 2.0*. Data/software, ÚFAL MFF UK, Prague, Czech Republic, <http://hdl.handle.net/11234/1-1905> (2016).
11. Stede, M., Grishina, Y.: Anaphoricity in Connectives: A Case Study on German. In: *Proceedings of the Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2016), co-located with NAACL 2016*, pp. 41-46. San Diego, California, (2016). WWW: <<http://www.aclweb.org/anthology/W16-0706>>.
12. Taboada, M.: Implicit and explicit coherence relations. In Renkema, J. (ed.): *Discourse, of course*, pp. 125-138. Amsterdam: John Benjamins (2009).
13. *Treq – Translation Equivalent Database*, Institute of the Czech National Corpus, Prague. 5. 2. 2018. WWW: <<http://treq.korpus.cz/>>.
14. Webber, B., Stone, M., Joshi, A., Knott, A.: Anaphora and discourse structure. *Computational Linguistics* 29(4), 545-587 (2003).
WWW: <<https://dl.acm.org/citation.cfm?id=1105705>>.
15. Zikánová, Š., Hajičová, E., Hladká, B., Jínová, P., Mirovský, J., Nedoluzhko, A., Poláková, L., Rysová, K., Rysová, M., Václ, J.: *Discourse and Coherence. From the Sentence Structure to Relations in Text*. Copyright © ÚFAL, Praha, Czechia, ISBN 978-80-904571-8-8, 274 pp. (2015).