



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

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.*
- Therefore**, I know that I can walk long distances in any weather.
 - Thanks to this**, I know that I can walk long distances in any weather.
 - 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).

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.

Download English Version:

<https://daneshyari.com/en/article/7297302>

Download Persian Version:

<https://daneshyari.com/article/7297302>

[Daneshyari.com](https://daneshyari.com)