



Discours

Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics

24 | 2019

Varia

Multiple Signals of Coherence Relations

Debopam Das and Maite Taboada



Electronic version

URL: <http://journals.openedition.org/discours/10032>

DOI: 10.4000/discours.10032

ISSN: 1963-1723

Publisher:

Laboratoire LATTICE, Presses universitaires de Caen

Electronic reference

Debopam Das and Maite Taboada, « Multiple Signals of Coherence Relations », *Discours* [Online], 24 | 2019, Online since 30 October 2019, connection on 01 November 2019. URL : <http://journals.openedition.org/discours/10032> ; DOI : 10.4000/discours.10032

Licence CC BY-NC-ND



Revue de linguistique, psycholinguistique et informatique

<http://journals.openedition.org/discours/>

Multiple Signals of Coherence Relations

Debopam Das

Department of English and American Studies
Humboldt University of Berlin, Germany

Maite Taboada

Department of Linguistics
Simon Fraser University, Canada

.....
Debopam Das, Maite Taboada, «Multiple Signals of Coherence Relations», *Discours* [En ligne], 24 | 2019,
mis en ligne le 30 octobre 2019.

.....
URL : <http://journals.openedition.org/discours/10032>

.....
Titre du numéro : *Varia*
Coordination : Anne Le Draoulec & Josette Rebeyrolle

Date de réception de l'article : 07/01/2019
Date d'acceptation de l'article : 16/04/2019



Presses
universitaires
de Caen 

Multiple Signals of Coherence Relations

Debopam Das

Department of English and American Studies
Humboldt University of Berlin, Germany

Maite Taboada

Department of Linguistics
Simon Fraser University, Canada

.....

In this paper, we investigate the signalling of coherence relations when they are simultaneously indicated by more than one signal. In particular, we examine the co-occurrence of discourse markers and other relational signals when they are used together to mark a single relation. With the goal to identify the source of the usage of multiple signals, we postulate a two-fold hypothesis: the co-occurrence of discourse markers and other textual signals can result from the type of the discourse markers themselves, or it can be triggered by the semantics of the relations in question. We conduct a corpus study, examining instances of multiple signals (co-occurrence of discourse markers and other signals) in the RST Signalling Corpus (Das et al., 2015). We analyze discourse markers that appear as part of multiple signals and also relations that frequently employ multiple signals as their indicators. Our observations suggest that the signalling of relations by multiple signals is a complex phenomenon, since the co-occurrence of discourse markers and other textual signals appears to arise from multiple sources.

Keywords: coherence relations, signals, discourse markers, corpus annotation, Rhetorical Structure Theory (RST)

1. Introduction

- 1 One of the central topics of inquiry in the study of discourse coherence is how coherence relations (relations between propositions or idea units, such as *Cause*, *Evidence*, *List* or *Summary*) are signalled in text. Coherence relations are often signalled by discourse markers¹ (henceforth DMs) such as *and*, *because*, *however* and *while*, and relations are sometimes classified as *explicit relations* if they contain a DM (Knott & Dale, 1994; Martin, 1992; Meyer & Webber, 2013; Renkema, 2004; Taboada, 2009; Taboada & Mann, 2006; van der Vliet & Redeker, 2014; Versley, 2013). DMs are widely believed to be the most reliable signals of coherence relations. Accordingly, their role in discourse interpretation has been extensively studied both in (psycholinguistic) discourse processing (see Das [2014] for an overview) and computational discourse applications such as discourse parsing (Hernault et al., 2010; Lin et al., 2014), machine translation (Meyer et al., 2011), text summarization (Alonso i Alemany, 2005), or argumentation mining (Kirschner et al., 2015). This

1. In this paper, we define discourse markers as having the meaning of a two-place relation, and not representing elements like hedges, fillers or interjections, as in conversations. While the term “discourse connective” is deemed to be more appropriate, we prefer to use the term “discourse marker” in the spirit of the RST Signalling Corpus (Das et al., 2015), which we base our analyses on.

emergence of interest in DMs has also led to the production of an increasingly large number of text corpora annotated for DMs (most often accompanied by relational annotations) in different languages, based on different discourse frameworks; for instance, PDTB (Penn Discourse Treebank) annotations for: English (Prasad et al., 2008), Arabic (Al-Saif & Markert, 2010), Czech (Rysová et al., 2016), Hindi (Oza et al., 2009), Turkish (Zeyrek et al., 2010) and multilingual (Zeyrek et al., 2018); RST (Rhetorical Structure Theory) for: English (Das et al., 2015), Dutch (Redeker et al., 2012) and Spanish (Maziero et al., 2011); or SDRT (Segmented Discourse Representation Theory) for French (Afantenos et al., 2012).

- 2 DMs have traditionally been considered to be the most prototypical (sometimes the only type of) signals of coherence relations, and this is partly the reason why research on signalling in discourse has primarily been confined to the analysis of DMs alone. Nevertheless, studies on the signalling phenomenon, mostly in recent years, have demonstrated that the range of signalling goes well beyond DMs, and that coherence relations can be indicated by other textual signals, both in addition to or in the absence of DMs. A few notable corpora representing such initiatives include:

1. The PDTB corpus (Prasad et al., 2007) includes annotation of AltLex (alternative lexicalization) for relations which do not have an explicit discourse connective (and cannot be assigned an implicit one either²), but are signalled by indicative lexical expressions. In the following example from the PDTB 2.0 (Prasad et al., 2010: 1025), the causal relation between two text segments (content within square brackets) is indicated by the lexical expression “One reason is”.

- [1] [Now, GM appears to be stepping up the pace of its factory consolidation to get in shape for the 1990s.] [One reason is mounting competition from new Japanese car plants in the U.S. that are pouring out more than one million vehicles a year at costs lower than GM can match.]
[Relation: Contingency-Cause-Reason]

In the most recent version of the corpus (PDTB 3.0) (Prasad et al., 2018), the category of AltLex has been extended to include more varieties of lexico-grammatical constructions as relational signals. For instance, the Condition relation in Example [2] is conveyed through the use of auxiliary inversion (Webber et al., 2019: 10)³.

- [2] [... but would have climbed 0.6%,] [had it not been for the storm]
(file no: wsj-0573)

-
2. An implicit connective in the PDTB refers to an extra connective which is inserted to link (and thus considered to best signal the inferred relation between) two relational arguments that otherwise do not contain an explicit discourse connective.
 3. In PDTB 3.0, this is represented by a finer version of AltLex, called AltLexC, which records the position of the relevant lexico-grammatical signal within a sentence.

2. The Prague Discourse Treebank (Rysová et al., 2016) provides annotation of two types of discourse connectives, primary and secondary. Primary connectives are roughly comparable to (explicit) discourse connectives in the PDTB, and they represent single-word units or non-compositional multi-word units which have become grammaticalized expressions, such as *because* or *therefore*. Secondary connectives, on the other hand, are multi-word lexical expressions that are not (yet) fully grammaticalized, such as *the reason was* or *for this reason*. In Example [3], the Causal relation is signalled by the expression “This caused”, which comprises a subject and a verb, and is used with a compositional meaning.

[3] [Fred didn’t stop joking.] [This caused hilarity among his friends for the whole evening.]

(Danlos et al., 2018: 51)

3. The RST Signalling Corpus (Das et al., 2015) employs a much wider perspective of signalling, and it provides annotation for a large variety of textual signals, such as *reference*, *lexical*, *semantic*, *syntactic*, *graphical* and *genre-related* features, including DMs. A few examples from the corpus are as follows:

[4] [Mr. Palmero recommends Temple-Inland,] [explaining that it is “virtually the sole major paper company not undergoing a major capacity expansion”, and thus should be able to lower long-term debt substantially next year.]

(file no: wsj-0666)

[Relation: Explanation-argumentative; signal: lexical (indicative word: “explaining”)]

[5] [... Mr. Lawson resigned from his six-year post because of a policy squabble with other cabinet members.] [He was succeeded by John Major, who Friday expressed a desire for a firm pound and supported the relatively high British interest rates...]

(file no: wsj-0693)

[Relation: Sequence; signal: semantic (lexical chain: “resigned” ~ “succeeded”)]

[6] [To compare temperatures over the past 10,000 years,] [researchers analyzed the changes in concentrations of two forms of oxygen.]

(file no: wsj-0683)

[Relation: Purpose; signal: syntactic (infinitival clause)]

3 Furthermore, it has also been shown that on a great many occasions relations can simultaneously be indicated by multiple signals (Das, 2014; Das & Taboada, 2018b). In Example [7] from the RST Signalling Corpus (Das et al., 2015), the Circumstance relation between the text segments is signalled by three signals: (i) the DM “since”, (ii) the change of tense between two clauses (simple past → present perfect), and (iii) an indicative phrase (“last December”).

- [7] [Since Mexican President Carlos Salinas de Gortari took office last December,
[special agents have arrested more than 6,000 federal employees on charges ranging
from extortion to tax evasion.]
(file no: wsj-1154)
[Relation: Circumstance]

4 The most recent version of the PDTB corpus (3.0) (Prasad et al., 2018) also acknowledges the co-occurrence of multiple signals, and it provides annotation of relations with both an explicit connective and AltLex. In Example [8], the connective is “but”, and the AltLex is “so did”.

- [8] [Admittedly last season’s runaway hit, “Steel Magnolias”, helped a lot,] [but so did
cost cutting and other measures insisted on by the board.]
(file no: wsj-0819)
(Webber et al., 2019: 14)

5 In this paper, we investigate the signalling of coherence relations by multiple signals, especially when relations are accompanied by both a DM and some other signal(s), as is shown in Examples [7] and [8]. We examine what necessitates the use of multiple signals, particularly a “DM + other signal(s)” combination, for relation marking. This perceived need to use multiple signals is taken mostly from the point of view of the speaker/writer: what makes them feel that the relation to be expressed needs more than one signal? This is a question that, ultimately, needs to be answered with psycholinguistic methods. Here, we provide a possible corpus linguistics explanation.

6 We address this signalling phenomenon from two directions: the perspective of DMs (and signalling in general) and from that of relation semantics (connectivity between discourse segments). The former perspective examines whether the need to employ a signal in addition to a DM stems from the presence of the DM itself. This is because DMs appear to vary with respect to the relations they prototypically signal (*if* as a signal for a Condition relation, or *since* as a signal for a Causal relation) and also with respect to the degree of being ambiguous (*while* can signal both a Temporal and a Comparison relation). On the other hand, the latter perspective examines whether the use of multiple signals arises from the relations themselves, since relations are considered to vary with respect to parameters such as intentionality (Mann & Thompson, 1988) or strength (Sanders et al., 2018; Sanders et al., 1992).

7 In order to address these questions, we examine what DMs frequently occur with other signals, and also what relations involve signalling by other signals in addition to DMs. Based on a corpus analysis of multiple signals in a discourse-annotated corpus, we analyze the functions of co-occurring DMs and other signals, and also examine the semantics of coherence relations that frequently employ multiple devices as their signals.

8 This paper is organized as follows: in Section 2, we provide an overview of the notion of signalling of coherence relations, particularly by signals other than DMs.

In this section, we also describe the RST Signalling Corpus (Das et al., 2015), which we use to examine the signalling phenomenon. Section 3 states our research objectives. In Section 4, we present our corpus findings about the co-occurrence of DMs and other textual signals. In Section 5, we discuss the findings, and reflect on the implications of those findings. Finally, Section 6 summarizes the paper, and presents a few directions for future research.

2. Expanding the notion of signalling, by looking beyond discourse markers

9 The rationale behind extending the notion of signalling of coherence relations beyond DMs rests on two premises. Firstly, coherence relations mostly occur without DMs. The scarcity of DMs has consistently been attested by a number of corpus-based studies, some of which are listed below.

1. In the largest corpus of discourse connectives, the PDTB 3.0 (Prasad et al., 2018), 54.07% of the relations (28,995 out of 53,628 annotated tokens) are shown to occur without (explicit) discourse connectives (Webber et al., 2019).
2. In smaller corpora (Taboada, 2006), 69% of the relations in conversation and 57% of the relations in newspaper articles contain no DMs.
3. In the RST Signalling Corpus (Das et al., 2015), only 18.21% of all annotated relations are signalled by DMs, while the remaining majority of relations (81.79%) have no DMs⁴.

10 Secondly, relations without DMs are interpreted well by humans (if not as equally as relations with DMs). Research on text processing suggests that connectedness in discourse is a psychological phenomenon, and coherence relations are cognitive entities that aid the readers and hearers to connect different parts of discourse, and thereby, to construct a cognitive representation of the textual information (Knott & Sanders, 1998; Sanders & Noordman, 2000; Sanders et al., 1992). Psycholinguistic studies show that coherence relations are interpreted and recognized while processing texts (Knott & Sanders, 1998; Mak & Sanders, 2013; Sanders & Spooren, 2007 and 2009; Sanders et al., 1992; Sanders et al., 1993), and they are recognized even when no DMs are present (Kamalski, 2007; Mulder, 2008; Mulder & Sanders, 2012; Sanders & Noordman, 2000).

4. The proportion of DMs varies across corpora, primarily because DMs (or comparable alternatives) are assigned different scopes in their definitions as to what lexical expressions are considered as such. For example, the PDTB corpus uses more flexible parameters in the connective definition, and includes as connectives very frequently occurring words such as *by*, *from*, *in* or *like*. In contrast, the RST Signalling Corpus employs a much stricter definition for DMs, and considers these expressions not as DMs, but rather as lexical signals (more specifically, indicative words).

- 11 The omnipresence of coherence relations without DMs in text and their successful interpretation by humans both suggest that if readers or hearers can understand a variety of relations, then there must be indicators which guide the interpretation process, beyond DMs. This essentially means that, if we believe that the interpretation of relations (or a large part of it) originates from the text itself, then the effect of signalling must also originate from the text, by means of different textual features.
- 12 The issue of signalling by other means has been dealt more successfully in computational linguistics. With the common goal of automatically identifying and characterizing coherence relations in unseen texts, a number of computational studies have employed, in addition to DMs, a wide array of linguistic or textual features. Some of these features include tense or mood (Scott & Souza, 1990); anaphora and deixis (Corston-Oliver, 1998); lexical chains (Marcu, 2000); punctuation and graphical markers (Dale, 1991a and b); textual layout (Bateman et al., 2001); noun phrase (NP) and verb phrase (VP) cues (Le Thanh, 2007); reference and discourse features (Theijssen, 2007; Theijssen et al., 2008); specific genre-related features (Cardoso et al., 2011; Maziero et al., 2011); collocations (Berzlánovich & Redeker, 2012), polarity, modality, and word-pairs (Pitler et al., 2009); coreference, givenness, and lexical features (Louis et al., 2010); word co-occurrences (Marcu & Echihiabi, 2002); noun and verb identity/class, argument structure (Lapata & Lascarides, 2004); or positional features, length features, and part-of-speech features (Sporleder & Lascarides, 2005 and 2008). For a summary of these, see Das (2014).
- 13 On the other hand, a few corpora have been developed for relational signals (in addition to, or other than DMs). Afantenos et al. (2012) presented ANNODIS, a corpus in French, which provides annotation of a wide range of textual signals in discourse such as punctuation, lexico-semantic patterns, layout and syntactic parallelism. Redeker et al. (2012) compiled a corpus of Dutch texts annotated with discourse structure and lexical cohesion. The cohesive devices representing lexical cohesion in the corpus include features such as lexical expressions indicative of certain relations, anaphoric chains and ellipsis. Duque (2014) developed a small corpus of 84 texts in Spanish, annotating only two relations, Cause and Result, from the RST Spanish Treebank (Maziero et al., 2011), with signalling information using numerous linguistic features, such as DMs, anaphors, non-finite verbs and genre structure. Hoek (2018) conducted an analysis of relational markers in a multilingual discourse-annotated parallel corpus (part of the Europarl-direct corpus [Cartoni et al., 2013; Koehn, 2005] in English, Dutch, German, French, and Spanish), and she identified segment-internal signals⁵, such as complex phrases, lexical items, modal markers, and verbal inflection, in addition to discourse connectives.
- 14 A recent attempt to provide a comprehensive coverage of diverse types of signalling devices is presented by the RST Signalling Corpus (Das & Taboada,

5. Hoek (2018) makes a distinction between discourse connectives and segment-internal signals: discourse connectives are not part of the text segments they connect, while segment-internal signals are integral part of the propositional content of the text segments.

2018a). In the present paper, we also use the corpus to examine multiple signals of coherence relations. We provide an overview of the RST Signalling Corpus in the next subsection.

2.1. RST Signalling Corpus

15 The RST Signalling Corpus (henceforth the RST-SC) is a corpus of signals of coherence relations. The corpus is built upon the RST Discourse Treebank (henceforth the RST-DT) (Carlson et al., 2002), which provides annotation of over 20,000 coherence relations in news articles from *The Wall Street Journal*. The original RST-DT contains annotations for spans (segmentation of the discourse into elementary discourse units) and relations between those spans, labeled for the type of RST relation. The RST-DT is widely used for discourse analysis and discourse parsing studies (Braud et al., 2016; Egg & Redeker, 2010; Feng & Hirst, 2014; Hernault et al., 2010; Sporleder & Lascarides, 2008).

16 In the RST-SC, the existing relations in the RST-DT are further annotated for signalling information. For each relation, the corpus provides information about whether it was signalled or not and, if signalled, by what type and how many signals. An important aspect of the signalling annotation in the RST-SC, as Das and Taboada (2018b) acknowledge, is that the annotated signals are positive signals, that is, indicators that a relation exists. This does not mean that such signals are used exclusively to indicate the relation (as it is evident in the many-to-many correspondences between relations and their signals). It also means that the signals, as textual devices, are not exclusively used to mark a relation; they may well have other purposes in the text (for instance, a pronoun, an instance of the reference signal type, also contributes to cohesion, in addition to signalling a relation). In a sense, this means the signals are compatible with a relation, not necessarily indicators of the relation exclusively.

17 The signals in the RST-SC include DMs, and eight types of other textual signals: *reference*, *lexical*, *semantic*, *morphological*, *syntactic*, *graphical*, *genre* and *numerical* features. These signals are organized hierarchically in a taxonomy of three levels: signal class, signal type, and specific signal. The top level, signal class, has three tags representing three major classes of signals: *single*, *combined*, and *unsure*. For each class, a second level is identified; for example, the class *single* is divided into the aforementioned nine signal types. Finally, the third level in the hierarchy refers to specific signals; for example, *reference* type has four specific signals: *personal*, *demonstrative*, *comparative* and *propositional*. The hierarchical organization of the signalling taxonomy is provided in Figure 1. Note that subcategories in the figure are only illustrative, not exhaustive. For the detailed taxonomy and more information about the definitions of signals, see Das and Taboada (2018a) and the annotation manual of the corpus (Das & Taboada, 2014)⁶.

6. See: http://www.sfu.ca/~mtaboada/docs/publications/RST_Signalling_Corpus_Annotation_Manual.pdf.

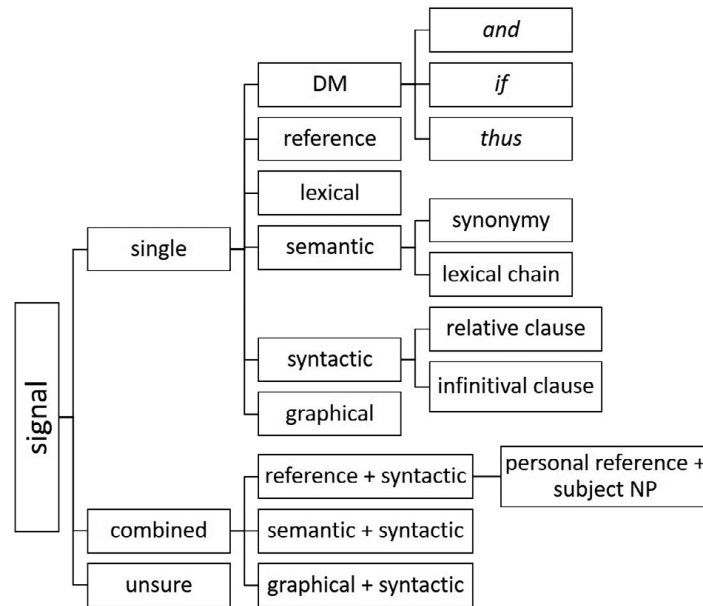


Figure 1 – Hierarchical taxonomy of signals in the RST Signalling Corpus

18 A *single* signal is made of one (and only one) feature used to indicate a particular relation. In Example [9] below, the DM “because”, which is a single signal, is used to signal the Explanation-argumentative relation.

- [9] [The Christmas quarter is important to retailers]N⁷ [because it represents roughly a third of their sales and nearly half of their profits.]S
(file no: wsj-0640)
[Relation: Explanation-argumentative]

19 In Example [10], the Interpretation relation is indicated by a lexical signal, the expression “That means”⁸, a single signalling feature.

- [10] [Production of full-sized vans will be consolidated into a single plant in Flint, Mich.]N [That means two plants – one in Scarborough, Ontario, and the other in Lordstown, Ohio – probably will be shut down after the end of 1991...]S
(file no: wsj-2338)
[Relation: Interpretation]

7. N and S refer to the nucleus and satellite span, respectively, as recognized in RST, to mark the relative importance of text segments in a relation.

8. The expression *that means* would qualify as an AltLex in the PDTB (Prasad et al., 2007), and as a secondary connective according to Rysová and Rysová (2018).

20 A *combined* signal comprises two single signals or features that work in combination with each other to signal a particular relation. In Example [11], two types of single signals, “reference” and “syntactic feature”, operate together to signal the Elaboration-general-specific relation. The reference feature indicates that the word “These” in the satellite span is a demonstrative pronoun because it refers back to the object “\$100 million of insured senior lien bonds”, mentioned in the nucleus span. Syntactically, the demonstrative pronoun, “These”, is also in the subject position of the sentence the satellite span starts with, providing more detail about the object “\$100 million of insured senior lien bonds” in the Elaboration-general-specific relation. Therefore, the combined signal, comprising the reference and the syntactic feature – in the form of a demonstrative reference plus a subject NP – functions here as a signal for the Elaboration-general-specific relation.

- [11] [The issue includes \$100 million of insured senior lien bonds.]N [These consist of current interest bonds due 1990-2002, 2010 and 2015, and capital appreciation bonds due 2003 and 2004...]S
(file no: wsj-1161)
[Relation: Elaboration-general-specific]

21 We would like to point out that both combined and multiple signals involve the use of more than one signal; yet they differ in a significant way, along the dimension of independence of operability. In a combined signal, there are two signals, one of which is the primary (independent) signal and the other signal is dependent on the primary signal. For example, in a combined signal such as “personal reference + subject NP”, the feature *personal reference* is the primary signal because it directly (and independently) refers back to the entity introduced in the first span. In contrast, the feature *subject NP* is the dependent signal because it is used to specify additional attributes of the primary signal. In this particular case, the syntactic role of the personal reference (i.e., a subject NP) in the second span is specified by the use of the second signal *subject NP*. For multiple signals, on the other hand, each signal functions independently and separately from each other, but they all contribute to signalling the relation. For example, in an Elaboration relation with multiple signals, such as a genre feature (e.g., inverted pyramid scheme) and a lexical feature (e.g., indicative word), the signals do not have any connection, but they separately signal the relation⁹.

22 Finally, *unsure* refers to those cases in which no signal was found, as represented in Example [12]. It was often used when the annotators could not clearly establish whether the relation was signalled or not. This may be the case when there are potential signals, but it is not clear whether they contribute to signalling the relation. For instance, in Example [12], there are reference signals (“Mr. Gelbart” - “He”), but it was difficult to establish whether those contributed to signalling the Evidence relation.

9. It is important to note that in the RST-SC no DMs are found to constitute combined signals; but, as this study shows, they frequently associate with other independent signals in the corpus, as part of multiple signals.

- [12] ["Mastergate" is subtitled "a play on words", and Mr. Gelbart plays that game as well as anyone.]N [He describes a Mastergate flunky as one who experienced a "meteoric disappearance" and found himself "handling blanket appeals at the Bureau of Indian Affairs".]S
(file no: wsj-1984)
[Relation: Evidence]

23 The distribution of relations by signals in the RST-SC (Table 1, from Das & Taboada, 2018a) shows that the overwhelming majority of the relations in the RST-DT are signalled, and also that the majority of signalled relations are indicated by other signals rather than DMs.

24 The distribution also shows that 1,616 relations (7.55% of all relations) contain both a DM and other signals as their indicators. In the present paper, we are primarily interested in this population of relations (and also the DMs that are used to indicate them).

25 The RST-SC includes 29,297 signal tokens for 21,400 relation instances, with a breakdown into 24,220 (82.7%) single signals, 3,524 (12.0%) combined signals, and 1,553 (5.3%) unsure cases (Das & Taboada, 2018a). The number of signal tokens (29,297) is higher than the relation instances (21,400). This is because relations are often indicated by more than one signal, thereby generating more signal tokens than the number of relations in the corpus (with an average of 1.37 signals for each relation). Table 2 presents the distribution of the (signalled) relations¹⁰ with respect to their (multiple) signals¹¹.

Relation type	Signalling type	#	%
Signalled relations	Relations exclusively signalled by DMs	2,280	10.65
	Relations exclusively signalled by other signals	15,951	74.54
	Relations signalled by both DMs and other signals	1,616	7.55
	TOTAL	19,847	92.74
Unsignalled relations	Relations not signalled by DMs or other signals	1,553	7.26
TOTAL		21,400	100.00

Table 1 – Distribution of signalled and unsignalled relations in the RST Signalling Corpus

10. For distinctions between relations and relation definitions, please consult the original RST paper (Mann & Thompson, 1988) and the RST-DT annotation manual (Carlson & Marcu, 2001).

11. Other views of the data, with different combinations of signals and relations, are available in Das (2014).

Multiple signals	# of signalled relations	% of signalled relations	Common signalled relations	# of common signalled relations	% of common signalled relations
1 signal or more	19,847	92.74	Elaboration-additional	4,043	97.56
			Attribution	3,061	99.71
			Elaboration-object-attribute	2,685	99.52
			List	1,843	94.27
			Circumstance	635	89.44
			Purpose	526	97.95
			Explanation-argumentative	392	64.69
			Antithesis	369	91.79
2 signals or more	7,901	36.92	Elaboration-additional	2,745	66.24
			List	861	44.04
			Elaboration-general-specific	185	39.11
			Contrast	182	41.84
			Circumstance	148	20.85
			Example	142	42.77
			Antithesis	133	33.08
			Elaboration-set-member	108	83.72
3 signals or more	2,614	12.22	Elaboration-additional	1,561	37.67
			Elaboration-general-specific	104	21.99
			Summary	62	74.70
			List	56	2.86
4 signals or more	725	3.39	Elaboration-additional	552	13.32
			Summary	35	42.17
			Elaboration-general-specific	25	5.29

5 signals or more	98	0.46	Elaboration- additional	80	1.93
			Elaboration- general-specific	6	1.27
6 signals	9	0.04	Elaboration- additional	9	0.22

Table 2 – Distribution of relations with (multiple) signals

26 Table 2 shows that a significant proportion (as well as a wide variety) of the signalled relations in the RST-SC contains more than one signal (36.92%). The distribution also shows what proportion of a relation in the RST-SC is conveyed by increasingly growing numbers of signals. For example, out of all List relations (1955 in total)¹² in the corpus, 94.27% of them are indicated by at least one signal, 44.04% of them contain two or more signals, and 2.86% of them are conveyed through three or more signals. Furthermore, Table 2 also shows that certain relations (e.g., Elaboration-general-specific) can employ even higher number of signals (as many as six signals for Elaboration-additional). The distribution clearly indicates that the co-occurrence of multiple signals for coherence relations in text is a quite common and pervasive phenomenon.

3. Research objectives and hypotheses

27 In this paper, we investigate the co-occurrence of multiple signals when they are used to simultaneously indicate the same relation. In particular, we focus on DMs when they are accompanied by other signalling device(s) to mark a relation.

28 As already mentioned earlier, although DMs are the most prototypical signals of coherence relations, they are frequently seen to co-occur with other signalling types. We provide some more examples of this from the RST-SC (Example [13]) and PDTB 3.0 (Example [14]).

[13] [Tele-Communications has a 21.8% stake,]N [while Time Warner has a 17.8% stake.]N

(file no: wsj-1190)

[Relation: Comparison]

[Signal 1: DM (“while”); Signal 2: syntactic (parallel constructions: *X has a Y*)]

[14] [... nor can the defendant be compelled to take the stand as a witness,] [thus forcing him to “take the Fifth”]

(file no: wsj-2377)

[Signal 1: DM (“thus”); Signal 2: AltLex (“forcing”)]

12. Source: Das and Taboada (2018b).

29 The co-occurrence of DMs and other signals for relation marking can arise from (either/both of the) two possible sources: (a) the signals themselves which work together to convey a relation, and (b) the relations themselves which require more than one signal for their complete and successful realization. The former possibility alludes to the relative signalling force (strength) of individual signals, while the latter invokes the semantics of the marked relation in question. We formulate these possibilities in two research questions (and corresponding hypotheses).

1. Do DMs that co-occur with other signals constitute weak signals, so that the relations which already include those DMs also require some sort of extra signalling by other devices? The conjecture seems plausible because some DMs are inherently ambiguous, and they can indicate more than one relation type (e.g., the DM *and* can signal both List and Elaboration relations). Consequently, relations which include these DMs might also need the presence of some other signal(s) to convey their specific meanings and intended effect.
2. Can the use of multiple signals (specifically with DMs) be triggered not by the signal types, but by the relation types themselves? It might be case that some relations types require multiple signals in cases where it is important to distinguish (e.g., between a semantic and pragmatic reading of a relation), and the other signals help express that difference.

30 The rationale for the first hypothesis rests on the fact that DMs are of different nature from a semantic and syntactic point of view. There are many perspectives on the classification of DMs, but most agree that they have different meaning and contribute different meaning to sentences. From a syntactic point of view, DMs can also be categorized according to their grammatical status (conjunction, preposition, adverbial) (Fraser, 1999). From a pragmatic point of view, DMs may relate propositional content of utterances, but they may also contribute to topic management, or to the interactional aspects of conversation (Schiffrin, 1987). It is plausible, then, that different types of DMs will have different co-occurrence patterns with other signals. By “weak signals” here we mean signals that do not have a direct correlation with any specific relation, i.e., that are found in multiple relations and can therefore be interpreted as signalling various relations. They are weak in that they do not point to any one relation in particular.

31 With respect to the second hypothesis, that relations trigger different patterns of signalling, the rationale is similar: we know that there exist different types of relations. In RST, for instance, relations are grouped into two main categories: subject matter and presentational (see Section 5). In the Cognitive approach to Coherence Relations or CCR (Sanders et al., 2018; Sanders et al., 1992), as we will also see in Section 5, parameters of relations include the order of segments, the source of coherence or the polarity of the relation. We explore whether such parameters contribute to differences in signalling of different types of relations.

32 We address these questions by examining the annotation of multiple signals (comprising a DM plus some other signalling feature) in the RST-SC. We should point out that our research addresses cases where multiple signalling is used to signal a single relation, i.e., where the DM and the other signals contribute to enabling the processing of one relation between text spans. In some cases elsewhere, multiple signals are present because multiple simultaneous relations are postulated. Rohde et al. (2018) present an extensive study of such cases. We assume, however, that the annotations in the RST-SC are of one relation at a time (as they have been provided in the corpus), for which multiple signals were found.

4. Relations, discourse markers and other signals in the RST Signalling Corpus

33 In order to examine the source of signalling of coherence relations by multiple signals, a good starting point is to identify the signals that often associate with each other for marking a relation, and also the relations which are realized through the use of more than one signal. For our purpose, we extract from the RST-SC (1) the instances of those DMs that co-occur (and also do not co-occur) with other signal(s), and also (2) the instances of those relations that contain (and also do not contain) multiple signals including DMs. In order to extract those tokens, we use UAM CorpusTool (O'Donnell, 2009), which was also used to annotate the corpus. The tool provides an efficient tag-specific search option for finding required annotated segments, and it also provides various types of statistical analyses of the corpus.

34 As shown in Table 1, we are primarily interested in the 1,616 relations (7.55% of 21,400 relations) which have a DM and some other signals as their indicators. A closer look at the DMs which occur in these relations shows their types, as provided in Table 3.

35 Overall, over 20 DMs (with a minimum of 10 instances) co-occur with other signals while signalling a relation in the RST-SC. The two most frequent DMs co-occurring with other signals are *and* and *but* (with over 300 tokens), while the other moderately frequent DMs of this type include *while*, *however*, *because*, *or* and *in addition*. The table also shows how frequently these DMs co-occur with other signals with respect to their overall occurrence in the corpus. For instance, some DMs almost always co-occur with other signals: *in addition* (96% of its overall occurrence), *moreover* (93.33%), and *so far* (90.91%). On the other hand, some DMs show moderate level of association with other signals, such as *and* (60.50%), *however* (60.87%) and *or* (63.41%), while some DMs, like *as* (16.87%), *when* (13.70%) and *after* (17.82%), appear in conjunction with other signals infrequently with respect to their overall occurrence.

36 Next, we examine what relations these DMs indicate while co-occurring with other signals. The distribution of relations (with a minimum of five instances) indicated by the most frequent DMs in the list (see Table 3) is provided below (in Table 4).

Co-occurrence frequency range	DM	Co-occurrence #	Overall #	% of co-occurrence
Freq. \geq 50	And	631	1,043	60.50
	But	309	615	50.24
	While	73	131	55.73
	However	56	92	60.87
	Because	52	162	32.10
50 > Freq. > 20	As	28	166	16.87
	Or	26	41	63.41
	In addition	24	25	96.00
	When	23	168	13.70
20 \geq Freq. \geq 10	Now	19	35	54.29
	Although	17	62	27.42
	After	18	101	17.82
	Meanwhile	16	31	51.61
	For example	15	32	46.88
	Moreover	14	15	93.33
	Since	14	36	38.89
	Though	14	43	32.56
	Without	14	51	27.45
	Also	13	16	81.25
	In fact	11	14	78.57
	Still	11	23	47.83
	For instance	10	16	62.50
	So far	10	11	90.91

Table 3 – Discourse markers co-occurring with other signals
in the RST Signalling Corpus

And		But		While	
Relations	#	Relations	#	Relations	#
List	443	Contrast	92	Comparison	16
Elaboration-additional	61	Antithesis	74	List	16
Sequence	30	Elaboration-additional	46	Contrast	16
Circumstance	13	Concession	36	Antithesis	10
Contrast	11	List	11	Concession	7
Consequence-s	9	Comparison	10	Circumstance	3
Comparison	8	Circumstance	7	Antithesis-e	2
Consequence-n	7	Interpretation-s	6	Elaboration-additional-e	1
Consequence	6	Explanation-argumentative	5	Concession-e	1
Cause	4	Background	4	Elaboration-additional	1
However		Because		As	
Relations	#	Relations	#	Relations	#
Antithesis	15	Reason	21	Circumstance	15
Contrast	12	Explanation-argumentative	9	Explanation-argumentative	3
Elaboration-additional	11	Consequence-n	7	Result	3
Concession	5	Cause-Result	5	Elaboration-additional	2
Interpretation-s	3	Result	4	Comparison	1
Comparison	2	Cause	2	List	1
List	2	Circumstance	1	Analogy	1
Consequence-s	1	Evidence	1	Interpretation-s	1
Circumstance	1	Result-e	1	Example	1
Problem-solution	1	Reason-e	1	Consequence-n	1
<i>A note on suffixes: -n stands for nucleus, -s stands for satellite, and -e indicates that the relation is embedded. For more information, see the RST-DT annotation manual (Carlson & Marcu, 2001).</i>					

Table 4 – Distribution of relations by discourse markers
(when co-occurring with other signals) in the RST Signalling Corpus

37 The distribution of relations in Table 4 shows that certain DMs are not only used very frequently in conjunction with other signals, but when they are used as such, they signal a wide variety of coherence relations in the corpus. For example, DMs such as *and* and *but*, when used as part of multiple signals, are used to indicate as many as nine relation types (with a minimum of five tokens) in the corpus. The other DMs with relatively lower frequencies, such as *while*, *however* and *because* (excluding the DM *as*), are also seen to signal many different relations when they occur with other signals. In other words, the correlation between the DMs and the relations signalled by them appears to be one-to-many.

38 Sometimes, the multiplicity of relations for a DM, however, seems to arise from the relation taxonomy of the original corpus (RST-DT) itself. This is because the RST-DT taxonomy employs a hierarchical organization of relations, placing similar relation types into broad relation groups¹³. For example, the Contrast group includes three relations: Contrast (multinuclear), Antithesis and Concession (both mononuclear). When it comes to relation marking, a broad contrast in text can often be (prototypically) signalled by a DM such as *while* or *however*. However, since the taxonomy distinguishes finer relation types, the relation marking by these DMs is sometimes distributed among similar yet several relation names (Contrast or Antithesis or Concession).

39 In addition to identifying the most frequently co-occurring DMs, it is also important to find out what types of other signals associate with them. In Table 5, we provide the distribution of those other signal types. An obvious point to observe in Table 5 is that among different types of other signals, the DMs that frequently occur as part of multiple signals all go with the *semantic* feature¹⁴. Some of these DMs, in addition to the semantic feature, are also accompanied by other types. For example, *but* co-occurs with the *lexical* and *reference* features, other than with the *semantic* type.

40 However, unlike the DMs in Table 3, 4 and 5, there are some DMs which either co-occur with other signals very rarely with respect to their overall distribution in the corpus, or do not co-occur with other signals at all (Table 6). We also provide the distribution of the relations which are indicated by these DMs, as presented in Table 7.

13. The RST-DT includes 16 broad relation groups, which are further divided into 78 specific relation types.

14. Distribution-wise, the semantic feature constitutes the second most frequent signal type (after the *syntactic* feature) in the RST-SC, accounting for 24.80% of all signal tokens (7,265 instances out of a total of 29,297 signal tokens) (Das & Taboada, 2018a).

Co-occurring DM	Total co-occurrence #	Co-occurring single signal			Co-occurring combined signal		
		Type	#	%	Type	#	%
And	631	Semantic	467	74.00	syntactic + semantic	88	13.95
		Lexical	43	6.81	reference + syntactic	22	3.49
		Syntactic	31	4.91	semantic + syntactic	18	2.85
		Morphological	26	4.12			
But	309	Semantic	277	89.64	reference + syntactic	17	5.50
		Lexical	30	9.71	semantic + syntactic	17	5.50
		Reference	15	4.85	syntactic + semantic	10	3.24
While	73	Semantic	46	63.01	syntactic + semantic	21	28.77
However	56	Semantic	39	69.64	semantic + syntactic	7	12.50
Because	52	Semantic	50	96.15	—	—	—
As	28	Lexical	9	32.14	—	—	—
		Semantic	9	32.14			

Table 5 – Distribution of major types of other signals that co-occur with discourse markers

DM	Co-occurrence #	Overall #	% of co-occurrence
Despite	0	30	0.00
Even if	0	14	0.00
Unless	0	14	0.00
If	4	180	2.22
Once	1	21	4.76
For	1	18	5.56
Until	1	18	5.56
Before	5	60	8.33

Table 6 – Discourse markers that do not or very rarely co-occur with other signals in the RST Signalling Corpus

Despite		Even if		Unless		If	
Relation	#	Relation	#	Relation	#	Relation	#
Concession	24	Concession	9	Condition	12	Condition	162
Antithesis	6	Condition	2	Otherwise	2	Circumstance	7
		Antithesis	1			Contingency	4
		Hypothetical	1			Hypothetical	2
		Circumstance	1			Antithesis	2
						Elaboration-general-specific	1
						Contrast	1
						Disjunction	1
Total	30	Total	14	Total	14	Total	180
Once		For		Until		Before	
Relation	#	Relation	#	Relation	#	Relation	#
Circumstance	15	Reason	8	Circumstance	9	Temporal-before	31
Condition	3	Consequence	3	Condition	7	Circumstance	14
Temporal-after	1	Elaboration-object-attribute	2	Temporal-before	2	Sequence	6
Concession	1	Purpose	2			Temporal-after	3
Contingency	1	Circumstance	1			Condition	3
		Cause-Result	1			Comparison	1
		Condition	1			Elaboration-object-attribute	1
						Contingency	1
Total	21	Total	18	Total	18	Total	60

Table 7 – Distribution of relations by discourse markers that do not or rarely co-occur with other signals

41

There are a few important aspects to notice about the DMs in Table 6 and the relations in Table 7. The DMs such as *despite*, *unless*, *if* and *before* (in Table 6) are not as profusely distributed in the RST-SC as the DMs in Table 3 (*and*, *but*, *however*, etc.). However, whenever these DMs are used, they typically appear alone, and not with other signals. Furthermore, the range of relations signalled by these DMs appears to be quite restricted (compared to that for the DMs in Table 4). For example, the DM *if* is primarily used (alone) to convey a Condition relation (disregarding the finer distinctions among the individual types, Condition, Contingency, Hypothetical and Otherwise, which together constitute the Condition group in the RST-DT). In other words, the correspondence between these DMs and the relations indicated by them constitutes more of a one-to-one mapping.

Co-occurrence frequency range	Relation	Co-occurrence #	Overall by DM #	% of co-occurrence	Overall # in RST-SC	% of co-occurrence
Freq. ≥ 50	List	533	818	65.16	1,955	27.26
	Elaboration-additional	221	269	82.16	4,144	5.33
	Contrast	167	305	54.75	435	38.39
	Antithesis	128	327	39.14	402	31.84
	Circumstance	90	382	23.56	710	12.68
	Concession	73	262	27.86	293	24.91
50 > Freq. > 20	Comparison	42	72	58.33	265	15.85
	Sequence	41	119	34.45	218	18.81
	Reason	30	114	26.32	206	14.56
	Example	27	52	51.92	332	8.13
	Explanation-argumentative	23	83	27.72	606	3.80
	Result	21	86	24.42	159	13.21
	Consequence-s	21	60	35.00	305	6.89
20 \geq Freq. ≥ 10	Consequence-n	20	82	24.39	165	12.12
	Interpretation-s	17	24	70.83	227	0.75
	Evaluation-s	16	20	80.00	198	8.08
	Disjunction	15	26	57.69	27	55.56
	Cause-result	15	42	35.71	65	23.08
	Background	11	22	50.00	227	4.85

Table 8 – Distribution of relations indicated by a discourse marker and other signals

Switching our attention from signals to relations, we next identify what relations frequently employ multiple signals, including a DM.

The distribution of relations in Table 8 shows what relations employ other signals in addition to a DM, and how frequently they do so with respect to their overall populations (both DM-specific and general) in the corpus. For example, List is a relation which is frequently indicated by a DM and other signals simultaneously. In the RST-SC, List appears with a DM 818 times, out of which 533 instances of List (65.16%) include an additional other signal. Furthermore, this amounts to 27.26% of all the List relations in the corpus (533 out of 1,955 instances).

We also draw some interesting observations from Table 8. First, there are some relations which are often signalled by a DM, and on many occasions, they also include other signals in addition to the DM. For example, List and Contrast are frequently indicated by a DM, and also quite often by a “DM + other signal” combination. Second, there are some relations which do not frequently include a DM, but when they do so, they also often incorporate other signals. Examples of such relations include Elaboration-additional, Comparison and Interpretation-s. Finally, there are also some relations which are frequently indicated by a DM, but these relations do not often include a “DM + other signal” combination. Concession and Reason are examples of this type.

In addition to identifying the relation types that include multiple signals, we also identify the major types of other signals that co-occur with DMs while signalling such relations. We provide the distribution of those other signals types in Table 9.

Relation	Total co-occurrence #	Co-occurring single signal			Co-occurring combined signal		
		Type	#	%	Type	#	%
List	533	Semantic	390	73.17	syntactic + semantic	100	18.76
		Syntactic	31	5.81			
Elaboration-additional	221	Semantic	68	30.77	semantic + syntactic	87	39.37
		Reference	46	20.81			
		Genre	31	14.03			
Contrast	167	Semantic	151	90.42	syntactic + semantic	12	7.19
Antithesis	128	Semantic	124	96.88	—	—	—
Circumstance	90	Lexical	37	41.11	—	—	—
		Morphological	33	36.67			
		Syntactic	21	23.33			
Concession	73	Semantic	70	95.89	—	—	—

Table 9 – Distribution of major other signals co-occurring with discourse markers for indicating a relation

Relation	Co-occurrence #	Overall by DM #	% of co-occurrence	Overall #	% of co-occurrence
Means	2	5	40.00	130	1.54
Manner	7	26	26.92	96	7.29
Contingency	1	20	5.00	27	3.70
Condition	1	221	0.05	239	0.42
Hypothetical	8	8	100.00	46	17.39
Purpose	1	12	8.33	537	0.02
Enablement	7	7	100.00	31	22.58
Evidence	7	16	43.75	174	4.02
Elaboration-general-specific	6	12	50.00	473	1.27
Elaboration-part-whole	1	2	50.00	44	2.27
Temporal-before	2	38	5.26	44	4.55
Temporal-after	3	69	4.35	93	3.23
Temporal-same-time	1	115	0.87	160	0.63
Summary	0	3	0.00	83	0.00
Restatement	0	4	0.00	140	0.00

Table 10 – Distribution of relations that rarely (or do not) include a discourse marker and other signal combination

46 As we have seen earlier in Table 5, we also notice in Table 9 that out of different types of other signals, the *semantic* feature is common for almost all major relation types that contain multiple signals. The only exception here is Circumstance relations, which are commonly conveyed by lexical, morphological or syntactic features, in addition to a DM.

47 Finally, we provide the distribution of relations which either employ the DM and other signal combinations very rarely or do not include them at all. As Table 10 shows, the relations can be of two types. On the one hand, there are relations which are neither signalled by a DM nor signalled by a “DM + other signal” combination. Examples of this type include Means, Purpose, Elaboration-part-whole and Summary. On the other hand, there are relations which are frequently signalled by a DM, but not by other signals when a DM is already present. Contingency, Condition and Temporal (-before, -after, -same-time) belong to this second type of relations.

5. Discussion

48 As we hypothesize that the possible source of the signalling phenomenon by multiple devices could be the signals or relations themselves, we first identify through our corpus analysis the DMs that are part of multiple signals and also the relations which employ multiple signals, including a DM. We closely examine those DMs and relations in this section.

49 Two DMs that most frequently co-occur with other signals in the RST-SC are *and* and *but*. They are also by far the most frequently occurring DMs in general (whether being used alone or in association with other signals for a relation). It is also seen that *and* and *but* co-occur with other signals for more than half of their overall distribution in the corpus (Table 3). The other DMs of this type with high frequencies include *while*, *however*, *because*, *as* and *or*.

50 The first important thing to notice, other than the fact that DMs often co-occur with other signals, is that DMs with other signals are used to indicate a wide variety of relations in the corpus (Table 4). For example, *and* as part of multiple signals is used to signal as many as nine different relations: List, Elaboration-additional, Sequence, Circumstance, Contrast, Consequence-s, Comparison, Consequence-n, Consequence and Cause. The same is true for *but*, and the relations indicated by *but* (in conjunction with other signals) include Contrast, Antithesis, Elaboration-additional, Concession, List, Comparison, Circumstance, Interpretation-s, Explanation-argumentative and Background. Wider distribution of relations (although not as wide as for *and* and *but*) is also found for other DMs belonging to this group, such as *while*, *however*, *because*, or *as*. The relation sets become even wider if we add relations when these DMs are used alone, and not as part of multiple signals (cf. Das, 2014).

51 A closer look at the relation sets for these DMs shows that relations are both mononuclear and multinuclear. For DMs such as *and* or *but*, this finding is particularly important since *and* and *but* are primarily considered to be conjunctive DMs; yet they are used to connect mononuclear (hierarchical) relations (e.g., Elaboration-additional, Circumstance) in addition to the prototypical multinuclear (non-hierarchical) relations (e.g., List, Contrast or Comparison). This could be a reason why other signals are deployed. Since markers such as *and* do not convey enough information about the relative hierarchy of the spans, other signals are used to help with that. In other words, the additional signals are there not only to help identify the specific relations (Elaboration or List), but the more general type of relation (mononuclear or multinuclear).

52 The other important point is that the relations indicated by these DMs are of different types. For example, the DM *and* is used to convey broad relation types (merging similar fine-grained relation labels) such as List, Elaboration, Sequence, Circumstance and Contrast, while the DM *but* is used to signal relation types such as Contrast, Elaboration, List and Comparison. Among the other less frequent DMs, *while* is used to indicate Comparison, List and Contrast types while *however* is used

for Contrast and Elaboration types. The variety is, however, gradually reduced for the remaining DMs with the lower frequencies. For example, the DM *because* is used to indicate primarily Causal relations (although of different subtypes), and *as* is mainly used to signal Circumstance relations. This tendency seems to correspond with the distribution of those DMs which do not usually co-occur with other signals. We have shown previously (in Tables 6 and 7) that DMs, such as *despite*, *even if*, *unless* and *before*, in general occur alone, and they prototypically signal a much more restricted set of relations (if not always one relation type). For instance, the DM *despite* only signals the Contrast type (specific subtypes: Concession and Antithesis), or the DM *unless* exclusively indicates the Condition type (specific subtypes: Condition and Otherwise). Based on these observations, it seems that the frequency of multiple signals is proportional to the number of the relation types they indicate. In other words, the more frequent a DM is in co-occurring with other signals, the greater number of relations it is used to signal.

53 The versatility of DMs at indicating different relation types calls for them to be categorized as ambiguous signals of coherence relations¹⁵. The ambiguous nature of these DMs makes them generic markers of coherence relations. This seems to parallel with Knott's (1996) notion of superordinate signals, which have a broader semantics and which constitute a hyponymy relationship with specific signals within a hierarchical taxonomy of relational signals (Knott calls them "cue phrases"). Knott proposes to classify cue phrases based on a substitutability test, which essentially relies on retaining or changing a given relation by altering possible variants of a cue phrase. He observes that more generic markers are typically substitutable for less generic ones (but not vice versa). For example, *and*, on most occasions, can be substituted for markers such as *furthermore*, *moreover* and *also*; *but*, in similar ways, is substituted for markers such as *despite this* or *whereas*. This, in turn, implies that superordinate markers are generic simply because they signal a larger class of relations, and specific signals are specific simply by virtue of being the signals of a more restricted class of relations.

54 In sum, ambiguous (and thus generic) DMs generally co-occur with other signals, and when these DMs occur as such, they signal a wide range of relations. This seems to indicate that ambiguity might have a direct bearing with co-occurrence of DMs and other signals. If a DM is ambiguous and can represent a set of many different relations, it potentially makes a particular inter-segmental link ambiguous as well, and thus allows the reader or hearer to interpret the link in terms of more than one relation. It seems plausible then that for such cases the writer or speaker chooses

15. We would like to clarify that the ambiguity in question does not necessarily point to the relations, but it may also arise from the signals themselves. The former view refers to a situation in which two text segments can be linked by more than one relation at the same time: e.g., "[George Bush supports big business.] [He's sure to veto House Bill 1711.]" – Relation 1: Evidence; Relation 2: Cause (from Moore & Pollack, 1992: 540). The latter view considers whether the same signal is capable of signalling one or more relations. This perspective is recently used in compiling discourse connective lexicons, in order to specify the range of relations a connective can point to (Das et al., 2018; Scheffler & Stede, 2016).

to add some extra signals, in order to compensate for the ambiguity of the DMs. This way one may increase the overall signalling force, and convey a more specific relation, which is identical (or closer to) the intended interpretation.

55 Now, we move our attention from signals to relations. Our second hypothesis conjectured that the relation marking by multiple signals is generated by the relations themselves, rather than the signals involved. Our corpus analysis (Table 8) shows that relations that frequently occur with a DM and other signals are of two types: (1) relations that are in general signalled by DMs (regardless of whether any other signal is present or not), such as List and Contrast, and (2) relations that generally do not contain a DM, but when they do have a DM, they also include some other signals, such as Elaboration-additional, Comparison and Interpretation-s. We examine the semantics of these relations with respect to two parameters: (i) intended effect, as postulated in RST (Mann & Thompson, 1988), and (ii) basic operation, as postulated in the CCR (Sanders et al., 2018; Sanders et al., 1992).

56 RST distinguishes two types of relations, subject matter and presentational, based on the intended effect of the writer. The intended effect for a subject matter relation is that the reader recognizes the relation, which essentially represents a relationship between facts. On the other hand, the intended effect for a presentational relation is to influence the reader (e.g., to increase positive regard, belief, or to accept a claim made in the nucleus), and this relation type encodes a relationship where one part contributes to that influence by presenting additional facts, explanations or an unrealized situation. The original RST relation taxonomy (Mann & Thompson, 1988) provides a list of 23 relations, which are divided into 16 subject matter relations and 7 presentational relations, as shown in Table 11.

Subject matter	Presentational
Elaboration	Motivation (increases desire)
Circumstance	Antithesis (increases positive regard)
Solutionhood	Background (increases ability)
Volitional Cause	Enablement (increases ability)
Volitional Result	Evidence (increases belief)
Non-Volitional Cause	Justify (increases acceptance)
Non-Volitional Result	Concession (increases positive regard)
Purpose	
Condition	
Otherwise	
Interpretation	
Evaluation	
Restatement	
Summary	
Sequence	
Contrast	

Table 11 – Classification of subject matter and presentational relations in Rhetorical Structure Theory

57 We observe that relations in type (1) and (2), as mentioned above, include Contrast, Elaboration(-additional), Interpretation(-s) and Comparison¹⁶, all of which are subject matter relations. However, in addition to these relations, when we also consider those relations which do not usually employ multiple signals (Table 10), such as Means, Purpose, Elaboration-part-whole, Summary, Condition, Contingency and Temporal relations, we find that most of them belong to the subject matter group as well. To summarize the findings, we find that relations that either employ or do not employ a DM and other signals as their indicators are often subject matter relations¹⁷. This may imply that the signalling of relations by multiple markers is an independent phenomenon, and it does not correlate with the intended effects borne out by the relations in any way¹⁸.

58 The CCR (Sanders et al., 2018; Sanders et al., 1992) provides a psychological account of coherence relations. It characterizes relations in terms of four cognitively basic features or primitives: (1) basic operation, (2) source of coherence¹⁹, (3) order of segments and (4) polarity. We examine the issue of multiple signals with respect to the first primitive, i.e., basic operation. According to the basic operation, relations are distinguished as either *causal* or *additive*, based on whether the text segments are strongly or weakly connected. Causal relations are characterized by stronger links, and they encode an implicational relationship between two text segments ($P \rightarrow Q$). Furthermore, the causal category encapsulates conditional relations since they also share the implicational nature. Additive relations, on the other hand, are characterized by weaker links, and they represent a non-implicational relationship between text segments, which are connected through logical conjunction ($P \& Q$).

59 In our study, relations that are frequently signalled by a DM and other signals, such as List, Contrast, Elaboration-additional or Comparison, represent additive relations. In contrast, relations that are not commonly signalled by multiple signals,

16. List and Comparison do not appear in the original RST taxonomy. In order to classify these relations, we draw on our previous work (Das & Taboada, 2013), which investigated the signalling for subject matter and presentational relations in the RST-DT. In that study, we followed the standard classification for 23 RST relations that appear in the corpus. For the new relations (not present in the original RST taxonomy), we categorized them as either subject matter or presentational relations (or even as a third category called *undetermined*), based on their definitions. Following this modified taxonomy, in the present study we consider Comparison as a subject matter relation, and List under the undetermined category.

17. We, however, could not make a similar conclusion about List, which we consider to fall under the *undetermined* category following Das and Taboada (2013), even though it is the most frequently occurring relation with a “DM + other signal” combination.

18. In Das and Taboada (2013), we also did not find any significant quantitative or qualitative difference between the signalling of subject matter and presentational relations. In that study, we, however, constrained our analysis mostly to the distribution of signals when they were used alone. Furthermore, we also examined the signalling phenomenon in a subset of the corpus (only one tenth of the RST-DT).

19. The *source of coherence* primitive classifies coherence relations into two types: semantic and pragmatic relations. The divide is roughly parallel to the distinction of subject matter and presentational relations in RST.

such as Condition and Contingency relations, represent causal relations²⁰. In other words, weaker (additive) relations more often contain other signals in addition to DMs, while stronger (causal) relations are not generally conveyed through such a conjunction of signals. This seems to imply that relations that are additive and weak in nature probably need extra signalling in addition to DMs in order to become adequately realized.

6. Conclusion

60 In trying to identify the source of the usage of multiple signals (especially comprising a DM and other textual signals), we observe that this signalling phenomenon can potentially result from either (or both) signals or relations in question. From the point of view of signals, we found that DMs that are generic and ambiguous in nature often co-occur with other signals to indicate a relation. In contrast, DMs that are relatively unambiguous do not usually associate with other signals. This seems to imply that the inherent ambiguity of DMs calls for extra signalling by other devices, in order to balance out the vagueness or imprecision that might be generated if the DM was used alone to indicate a relation. On the other hand, we have noticed a mixed effect of relations on the usage of multiple signals. In this study, we examine two parameters, by which relations are generally believed to differ: intentionality (of the writer) and causal-additive nature of relations. For the former, we found that subject matter relations in RST may or may not include multiple signals. This seems to imply that intentionality and the signalling phenomenon (with multiple markers) do not have any significant correlation, and that they could be two entirely separate mechanisms. For the latter relational parameter, we observed that it is the additive relations, which are believed to constitute weaker relations, that include more multiple signals than causal relations, which create a stronger link between text segments.

61 In sum, the signalling of coherence relations appears to be a complex phenomenon altogether. The complexity of the matter rises even higher if we add the dimension of plurality into it (as in the case of multiple signals). In our study, we identify multiple plausible sources of relation marking, each of which, either individually or collectively, can be responsible for the usage of multiple signals. Our findings strongly suggest that the signalling of coherence relations by multiple signals is in no way a trivial thing, but it is worthy of investigation in its own right. In the remainder of the conclusion section, we list out a few potential research ideas which we consider would be interesting to examine further.

20. We would like to note that Temporal (-before, -after, -same-time) relations are not usually signalled by a "DM + other signals" combination (Table 10), but they correspond to the additive type. We consider this as a major exception to the general tendency that we observe for the other relations. We leave this issue for future research.

62 An important topic to investigate is the possible combination types of DMs and other signals, and the roles individual signals play in such a combination. In our study, we observed that, in a set of multiple signals, DMs most frequently team up with *semantic* signals, but they also co-occur with *syntactic*, *lexical* or *reference* features (Tables 5 and 9). In a recent study, Hoek (2018) suggests a three-way classification of the combinations of DMs and segment-internal elements²¹:

1. division of labor: a situation in which the segment-internal feature and the connective (or part of the connective) can make each other redundant;
2. agreement: a situation in which the segment-internal feature is independently used to mark a relation, regardless of whether or not a connective is used. In such cases, the relation marking seems to result from an agreement between the feature and connective;
3. general collocation: a situation in which the meanings signalled by the connective and the other signal do not overlap.

63 It would be interesting to see if the classification proposed by Hoek (or a modified version of it) also holds for the DM plus other signal combinations in other corpora such as the RST-SC.

64 In the RST-SC, there are 1,553 relations (7.26% of all 21,400 relations) without any identifiable signals. Another challenging area for future research is the investigation of multiple signals vs. no signals for the otherwise same relation types. In the RST-SC, for example, some instances of Circumstance relation include multiple signals; however, a few of the Circumstance relations do not contain any signal at all. It would be worthwhile to study what might necessitate the same relation types to employ signalling by multiple signalling devices as opposed to no signals at all.

65 In this paper, we focus on the co-occurrence of DMs and other signals. In our corpus analysis, however, we also find instances of relations (although a very few in number) which are signalled by two DMs at the same time, such as the following:

[15] [Although Axa has been rebuffed by Farmers and hasn't had any meetings with management,]S [Mr. Bebear nonetheless appears to be trying to woo the company's executives with promises of autonomy and new-found authority under Axa.]N
(file no: wsj-1178)
[Relation: Antithesis]

66 Very recently, Cuenca and Crible (2019) present a corpus analysis of (sequentially) co-occurring DMs in English conversations. The authors adopt a three-way classification of DMs co-occurrence, originally proposed in Cuenca and Marín (2009):

21. Segment-internal elements include signals such as complex phrases, lexical items, modal markers, and verbal inflection, and they roughly constitute a subset of other signals in the RST-SC.

1. juxtaposition: DMs co-occur, but they neither combine with each other syntactically nor semantically (e.g., *and meanwhile*);
2. addition: DMs combine at the local level, but they individually serve distinct functions (e.g., *because, in addition*);
3. composition: the co-occurrence of DMs constitutes a single complex unit, which cumulatively contributes to signalling a discourse function at a global level (e.g., *then OK*).

67 Similarly, the co-occurrence of multiple DMs in written text, in terms of juxtaposition and addition (or similar categories), could also be a very interesting topic to investigate.

References

- AFANTENOS, S., ASHER, N., BENAMARA, F., BRAS, M., FABRE, C., HO-DAC, L.-M., LE DRAOULEC, A., MULLER, P., PÉRY-WOODLEY, M.-P., PRÉVOT, L., REBEYROLLE, J., TANGUY, L., VERGEZ-COURET, M. & VIEU, L. 2012. An Empirical Resource for Discovering Cognitive Principles of Discourse Organization: The ANNODIS Corpus. In N. CALZOLARI, K. CHOUKRI, T. DECLERCK, M. UĞUR DOĞAN, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODIJK & S. PIPERIDIS (eds.), *Proceedings of the 8th International Conference on Language Resources and Evaluation – LREC 2012*. Luxembourg: European Language Resources Association: 2727-2734. Available online: http://www.lrec-conf.org/proceedings/lrec2012/pdf/836_Paper.pdf.
- ALONSO I ALEMANY, L. 2005. *Representing Discourse for Automatic Text Summarization via Shallow NLP Techniques*. PhD dissertation. Universitat de Barcelona.
- AL-SAIF, A. & MARKERT, K. 2010. The Leeds Arabic Discourse Treebank: Annotating Discourse Connectives for Arabic. In N. CALZOLARI, K. CHOUKRI, B. MAEGAARD, J. MARIANI, J. ODIJK, S. PIPERIDIS, M. ROSNER & D. TAPIAS (eds.), *Proceedings of the 7th International Conference on Language Resources and Evaluation – LREC 2010*. Luxembourg: European Language Resources Association: 2046-2053. Available online: http://www.lrec-conf.org/proceedings/lrec2010/pdf/479_Paper.pdf.
- BATEMAN, J., KAMPS, T., KLEINZ, J. & REICHENBERGER, K. 2001. Towards Constructive Text, Diagram, and Layout Generation for Information Presentation. *Computational Linguistics* 27 (3): 409-449.
- BERZLÁNOVICH, I. & REDEKER, G. 2012. Genre-Dependent Interaction of Coherence and Lexical Cohesion in Written Discourse. *Corpus Linguistics and Linguistic Theory* 8 (1): 183-208.
- BRAUD, C., PLANK, B. & SØGAARD, A. 2016. Multi-view and Multi-task Training of RST Discourse Parsers. In Y. MATSUMOTO & R. PRASAD (eds.), *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. Stroudsburg: Association for Computational Linguistics: 1903-1913. Available online: <https://www.aclweb.org/anthology/C16-1179>.

- CARDOSO, P.C.F., MAZIERO, E.G., CASTRO JORGE, M.L.R., SENO, E.M.R., DI FELIPPO, A., RINO, L.H.M., NUNES, M.G.V. & PARDO, T.A.S. 2011. CSTNews – A Discourse-Annotated Corpus for Single and Multi-document Summarization of News Texts in Brazilian Portuguese. In *Anais do III Workshop “A RST e os Estudos do Texto” (Cuiabá, MT, Brasil, Outubro 24-26, 2011)*. Porto Alegre: Sociedade Brasileira de Computação: 88-105. Available online: <http://www.icmc.usp.br/~tasparado/rst2011-cardosoetal1.pdf>.
- CARLSON, L. & MARCU, D. 2001. *Discourse Tagging Reference Manual*. Marina del Rey: University of Southern California, Information Sciences Institute. 1-87. Available online: <https://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf>.
- CARLSON, L., MARCU, D. & OKUROWSKI, M.E. 2002. RST Discourse Treebank, LDC2002T07. Philadelphia: Linguistic Data Consortium. Available online: <https://catalog.ldc.upenn.edu/LDC2002T07>.
- CARTONI, B., ZUFFEREY, S. & MEYER, T. 2013. Using the Europarl Corpus for Cross-linguistic Research. *Belgian Journal of Linguistics* 27 (1): 23-42.
- CORSTON-OLIVER, S.H. 1998. Beyond String Matching and Cue Phrases: Improving Efficiency and Coverage in Discourse Analysis. In *AAAI 1998 Spring Symposium Series, Intelligent Text Summarization*. Palo Alto: Association for the Advancement of Artificial Intelligence (AAAI): 9-15. Available online: <https://pdfs.semanticscholar.org/8a9a/0cecbfdfa287330fcc583d5c7ccd79ca034.pdf>.
- CUENCA, M.J. & CRIBLE, L. 2019. Co-occurrence of Discourse Markers in English: From Juxtaposition to Composition. *Journal of Pragmatics* 140: 171-184.
- CUENCA, M.J. & MARÍN, M.J. 2009. Co-occurrence of Discourse Markers in Catalan and Spanish Oral Narrative. *Journal of Pragmatics* 41 (5): 899-914.
- DALE, R. 1991a. Exploring the Role of Punctuation in the Signalling of Discourse Structure. In S. PREUSS & B. SCHMITZ (eds.), *Proceedings of the Workshop on Text Representation and Domain Modelling: Ideas from Linguistics and AI (Technical University of Berlin, October 9-11, 1991)*. 110-120. Available online: <http://clt.mq.edu.au/~rdale/publications/papers/1991/dscpunct1992.pdf>.
- DALE, R. 1991b. The Role of Punctuation in Discourse Structure. In *Working Notes for the AAAI Fall Symposium on Discourse Structure in Natural Language Understanding and Generation (Asilomar, November 1991)*. 13-14. Available online: <http://clt.mq.edu.au/~rdale/publications/papers/1991/aaai-fall-symposium1991.pdf>.
- DANLOS, L., RYSOVÁ, K., RYSOVÁ, M. & STEDE, M. 2018. Primary and Secondary Discourse Connectives: Definitions and Lexicons. *Dialogue and Discourse* 9 (1): 50-78. Available online: <http://dad.uni-bielefeld.de/index.php/dad/article/download/3734/3617>.
- DAS, D. 2014. *Signalling of Coherence Relations in Discourse*. PhD dissertation. Simon Fraser University, Burnaby, Canada.
- DAS, D., SCHEFFLER, T., BOURGONJE, P. & STEDE, M. 2018. Constructing a Lexicon of English Discourse Connectives. In K. KOMATANI, D. LITMAN, K. YU, A. PAPANGELIS, L. CAVEDON & M. NAKANO (eds.), *Proceedings of the 19th Annual Meeting of the Special Interest Group on Discourse and Dialogue – SIGDIAL 2018 (12-14 July 2018, Melbourne, Australia)*. Stroudsburg: Association for Computational Linguistics: 360-365. Available online: <https://www.aclweb.org/anthology/W18-5042>.

- DAS, D. & TABOADA, M. 2013. Signalling Subject Matter and Presentational Coherence relations in Discourse: A Corpus Study. Paper presented at the *2013 Meeting of the Linguistic Association of Canada and the United States (LACUS)*, Brooklyn College, New York.
- DAS, D. & TABOADA, M. 2014. *RST Signalling Corpus Annotation Manual*. Burnaby: Simon Fraser University, Department of Linguistics. Available online: https://www.sfu.ca/~mtaboada/docs/publications/RST_Signalling_Corpus_Annotation_Manual.pdf.
- DAS, D. & TABOADA, M. 2018a. RST Signalling Corpus: A Corpus of Signals of Coherence Relations. *Language Resources and Evaluation* 52 (1): 149-184.
- DAS, D. & TABOADA, M. 2018b. Signalling of Coherence Relations in Discourse, beyond Discourse Markers. *Discourse Processes* 55 (8): 743-770.
- DAS, D., TABOADA, M. & McFETRIDGE, P. 2015. RST Signalling Corpus, LDC2015T10. Philadelphia: Linguistic Data Consortium. Available online: <https://catalog.ldc.upenn.edu/LDC2015T10>.
- DUQUE, E. 2014. Signaling Causal Coherence Relations. *Discourse Studies* 16 (1): 25-46.
- EGG, M. & REDEKER, G. 2010. How Complex Is Discourse Structure? In N. CALZOLARI, K. CHOUKRI, B. MAEGAARD, J. MARIANI, J. ODIJK, S. PIPERIDIS, M. ROSNER & D. TAPIAS (eds.), *Proceedings of the 7th International Conference on Language Resources and Evaluation – LREC 2010*. Luxembourg: European Language Resources Association: 1619-1623. Available online: http://www.lrec-conf.org/proceedings/lrec2010/pdf/796_Paper.pdf.
- FENG, V.W. & HIRST, G. 2014. A Linear-Time Bottom-Up Discourse Parser with Constraints and Post-editing. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) – ACL-2014*. Stroudsburg: Association for Computational Linguistics: 511-521. Available online: <https://www.aclweb.org/anthology/P14-1048>.
- FRASER, B. 1999. What Are Discourse Markers? *Journal of Pragmatics* 31 (7): 931-952.
- HERNAULT, H., PRENDINGER, H., DUVERLE, D.A. & ISHIZUKA, M. 2010. HILDA: A Discourse Parser Using Support Vector Machine Classification. *Dialogue and Discourse* 1 (3): 1-33. Available online: <https://journals.linguisticsociety.org/elanguage/dad/article/download/591/591-2300-1-PB.pdf>.
- HOEK, J. 2018. *Making Sense of Discourse – On Discourse Segmentation and the Linguistic Marking of Coherence Relations*. PhD dissertation. Utrecht University.
- KAMALSKI, J. 2007. *Coherence Marking, Comprehension and Persuasion: On the Processing and Representations of Discourse*. Utrecht: LOT.
- KIRSCHNER, C., ECKLE-KOHLER, J. & GUREVYCH, I. 2015. Linking the Thoughts: Analysis of Argumentation Structures in Scientific Publications. In C. CARDIE (ed.), *NAACL HLT 2015: Proceedings of the 2nd Workshop on Argumentation Mining*. Stroudsburg: Association for Computational Linguistics: 1-11. Available online: <https://www.aclweb.org/anthology/W15-0501>.
- KNOTT, A. 1996. *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. PhD dissertation. University of Edinburgh.
- KNOTT, A. & DALE, R. 1994. Using Linguistic Phenomena to Motivate a Set of Coherence Relations. *Discourse Processes* 18 (1): 35-62.

- KNOTT, A. & SANDERS, T.J.M. 1998. The Classification of Coherence Relations and Their Linguistic Markers: An Exploration of Two Languages. *Journal of Pragmatics* 30 (2): 135-175.
- KOEHN, P. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of the 10th Machine Translation Summit (September 12-16, 2005, Phuket, Thailand) – MT Summit X*. 79-86. Available online: <http://www.mt-archive.info/MTS-2005-Koehn.pdf>.
- LAPATA, M. & LASCARIDES, A. 2004. Inferring Sentence-Internal Temporal Relations. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics – HLT-NAACL 2004*. Stroudsburg: Association for Computational Linguistics: 153-160. Available online: <https://www.aclweb.org/anthology/N04-1020>.
- LE THANH, H. 2007. An Approach in Automatically Generating Discourse Structure of Text. *Journal of Computer Science and Cybernetics* 23 (3): 212-230. Available online: <http://vjs.ac.vn/index.php/jcc/article/download/1380/1777>.
- LIN, Z., NG, H.T. & KAN, M.-Y. 2014. A PDTB-Styled End-to-End Discourse Parser. *Natural Language Engineering* 20 (2): 151-184.
- LOUIS, A., JOSHI, A., PRASAD, R. & NENKOVA, A. 2010. Using Entity Features to Classify Implicit Discourse Relations. In R. FERNÁNDEZ, Y. KATAGIRI, K. KOMATANI, O. LEMON & M. NAKANO (eds.), *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue – SIGDIAL 2010 (24-25 September 2010, University of Tokyo, Japan)*. Stroudsburg: Association for Computational Linguistics: 59-62. Available online: <https://www.aclweb.org/anthology/W10-4310>.
- MAK, W.M. & SANDERS, T.J.M. 2013. The Role of Causality in Discourse Processing: Effects of Expectation and Coherence Relations. *Language and Cognitive Processes* 28 (9): 1414-1437.
- MANN, W.C. & THOMPSON, S.A. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8 (3): 243-281.
- MARCU, D. 2000. The Rhetorical Parsing of Unrestricted Texts: A Surface-Based Approach. *Computational Linguistics* 26 (3): 395-448.
- MARCU, D. & ECHIHIABI, A. 2002. An Unsupervised Approach to Recognizing Discourse Relations. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics – ACL-2002*. Stroudsburg: Association for Computational Linguistics: 368-375. Available online: <https://www.aclweb.org/anthology/P02-1047>.
- MARTIN, J.R. 1992. *English Text: System and Structure*. Amsterdam – Philadelphia: J. Benjamins.
- MAZIERO, E.G., PARDO, T.A.S., CUNHA, I. DA, TORRES-MORENO, J.-M. & SANJUAN, E. 2011. DiZer 2.0 – An Adaptable On-line Discourse Parser. In *Anais do III Workshop “A RST e os Estudos do Texto” (Cuiabá, MT, Brasil, Outubro 24-26, 2011)*. Porto Alegre: Sociedade Brasileira de Computação: 1-17. Available online: <http://www.icmc.usp.br/~taspardo/RST2011-MazieroEtAl.pdf>.

- MEYER, T., POPESCU-BELIS, A., ZUFFEREY, S. & CARTONI, B. 2011. Multilingual Annotation and Disambiguation of Discourse Connectives for Machine Translation. In J.Y. CHAI, J.D. MOORE, R.J. PASSONNEAU & D.R. TRAUM (eds.), *Proceedings of the 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue – SIGDIAL 2011 (June 17-18, 2011, Oregon Science and Health University, Portland, Oregon, USA)*. Stroudsburg: Association for Computational Linguistics: 194-203. Available online: <https://www.aclweb.org/anthology/W11-2022>.
- MEYER, T. & WEBBER, B.L. 2013. Implication of Discourse Connectives in (Machine) Translation. In B.L. WEBBER, A. POPESCU-BELIS, K. MARKERT & J. TIEDEMANN (eds.), *Proceedings of the Workshop on Discourse in Machine Translation – DiscoMT (Sofia, Bulgaria, August 9, 2013)*. Stroudsburg: Association for Computational Linguistics: 19-26. Available online: <https://www.aclweb.org/anthology/W13-3303>.
- MOORE, J.D. & POLLACK, M.E. 1992. A Problem for RST: The Need for Multi-level Discourse Analysis. *Computational Linguistics* 18 (4): 537-544. Available online: <https://www.aclweb.org/anthology/J92-4007>.
- MULDER, G. 2008. *Understanding Causal Coherence Relations*. PhD dissertation. Utrecht University.
- MULDER, G. & SANDERS, T.J.M. 2012. Causal Coherence Relations and Levels of Discourse Representation. *Discourse Processes* 49 (6): 501-522.
- O'DONNELL, M. 2009. The UAM CorpusTool: Software for Corpus Annotation and Exploration. In C.M. BRETONES CALLEJAS (ed.), *Applied Linguistics Now: Understanding Language and Mind / La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente*. Almería: Universidad de Almería: 1433-1447. Available online: <https://viancep2012.files.wordpress.com/2012/10/actasalmeria.pdf>.
- OZA, U., PRASAD, R., KOLACHINA, S., SHARMA, D.M. & JOSHI, A. 2009. The Hindi Discourse Relation Bank. In M. STEDE, C.-R. HUANG, N. IDE & A. MEYERS (eds.), *Proceedings of the Third Linguistic Annotation Workshop (LAW III) – ACL-IJCNLP 2009 (Suntec, Singapore, 6-7 August 2009)*. Stroudsburg: Association for Computational Linguistics: 158-161. Available online: <https://www.aclweb.org/anthology/W09-3029>.
- PITLER, E., LOUIS, A. & NENKOVA, A. 2009. Automatic Sense Prediction for Implicit Discourse Relations in Text. In K.-Y. SU, J. SU, J. WIEBE & H. LI (eds.), *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP – ACL-IJCNLP 2009 (Suntec, Singapore, 2-7 August 2009)*. Stroudsburg: Association for Computational Linguistics: 683-691. Available online: <https://www.aclweb.org/anthology/P09-1077>.
- PRASAD, R., DINESH, N., LEE, A., MILTSAKAKI, E., ROBALDO, L., JOSHI, A. & WEBBER, B.L. 2008. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation – LREC 2008*. Luxembourg: European Language Resources Association: 2961-2968. Available online: http://www.lrec-conf.org/proceedings/lrec2008/pdf/754_paper.pdf.
- PRASAD, R., JOSHI, A. & WEBBER, B.L. 2010. Realization of Discourse Relations by Other Means: Alternative Lexicalizations. In C.-R. HUANG & D. JURAFSKY (eds.), *23rd International Conference on Computational Linguistics – COLING 2010 (23-27 August 2010, Beijing, China). Posters Volume*. Beijing: Chinese Information Processing Society of China: 1023-1031. Available online: <https://www.aclweb.org/anthology/C10-2118>.

- PRASAD, R., MILTSAKAKI, E., DINESH, N., LEE, A., JOSHI, A., ROBALDO, L. & WEBBER, B.L. 2007. *The Penn Discourse Treebank 2.0 Annotation Manual*. IRCS technical reports series 203. Philadelphia: University of Pennsylvania ScholarlyCommons. Available online: https://repository.upenn.edu/cgi/viewcontent.cgi?article=1203&context=ircs_reports.
- PRASAD, R., WEBBER, B.L. & LEE, A. 2018. Discourse Annotation in the PDTB: The Next Generation. In H. BUNT (ed.), *Proceedings of the 14th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (Santa Fe, New Mexico, USA, August 25, 2018)*. Tilburg: Tilburg Center for Cognition and Communication, Tilburg University: 87-97. Available online: <https://www.aclweb.org/anthology/W18-4710>.
- REDEKER, G., BERZLÁNOVICH, I., VAN DER VLIET, N., BOUMA, G. & EGG, M. 2012. Multi-layer Discourse Annotation of a Dutch Text Corpus. In N. CALZOLARI, K. CHOUKRI, T. DECLERCK, M. UĞUR DOĞAN, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODIJK & S. PIPERIDIS (eds.), *Proceedings of the 8th International Conference on Language Resources and Evaluation – LREC 2012*. Luxembourg: European Language Resources Association: 2820-2825. Available online: http://www.lrec-conf.org/proceedings/lrec2012/pdf/887_Paper.pdf.
- RENKEMA, J. 2004. *Introduction to Discourse Studies*. Amsterdam – Philadelphia: J. Benjamins.
- ROHDE, H., JOHNSON, A., SCHNEIDER, N. & WEBBER, B.L. 2018. Discourse Coherence: Concurrent Explicit and Implicit Relations. In I. GUREVYCH & Y. MIYAO (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) – ACL 2018 (July 15-20, 2018, Melbourne, Australia)*. Stroudsburg: Association for Computational Linguistics: 2257-2267. Available online: <https://www.aclweb.org/anthology/P18-1210>.
- RYSOVÁ, M. & RYSOVÁ, K. 2018. Primary and Secondary Discourse Connectives: Constraints and Preferences. *Journal of Pragmatics* 130: 16-32.
- RYSOVÁ, M., SYNKOVÁ, P., MÍROVSKÝ, J., HAJČOVÁ, E., NEDOLUZHKO, A., OCELÁK, R., PERGLER, J., POLÁKOVÁ, L., SCHELLER, V., ZDEŇKOVÁ, J. & ZIKÁNOVÁ, Š. 2016. Prague Discourse Treebank 2.0. Prague: Charles University, Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics (UFAL). Available online: <http://hdl.handle.net/11234/1-1905>.
- SANDERS, T.J.M., DEMBERG, V., HOEK, J., SCHOLMAN, M.C.J., ASR, F.T., ZUFFEREY, S. & EVERS-VERMEUL, J. 2018. Unifying Dimensions in Coherence Relations: How Various Annotation Frameworks Are Related. *Corpus Linguistics and Linguistic Theory* (Ahead of print).
- SANDERS, T.J.M. & NOORDMAN, L.G.M. 2000. The Role of Coherence Relations and Their Linguistic Markers in Text Processing. *Discourse Processes* 29 (1): 37-60.
- SANDERS, T.J.M. & SPOOREN, W.P.M. 2007. Discourse and Text Structure. In D. GEERAERTS & H. CUYCKENS (eds.), *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press: 916-941.
- SANDERS, T.J.M. & SPOOREN, W.P.M. 2009. The Cognition of Discourse Coherence. In J. RENKEMA (ed.), *Discourse, of Course: An Overview of Research in Discourse Studies*. Amsterdam – Philadelphia: J. Benjamins: 197-212.
- SANDERS, T.J.M., SPOOREN, W.P.M. & NOORDMAN, L.G.M. 1992. Toward a Taxonomy of Coherence Relations. *Discourse Processes* 15 (1): 1-35.

- SANDERS, T.J.M., SPOOREN, W.P.M. & NOORDMAN, L.G.M. 1993. Coherence Relations in a Cognitive Theory of Discourse Representation. *Cognitive Linguistics* 4 (2): 93-133.
- SCHEFFLER, T. & STEDE, M. 2016. Adding Semantic Relations to a Large-Coverage Connective Lexicon of German. In N. CALZOLARI, K. CHOUKRI, T. DECLERCK, S. GOGGI, M. GROBELNIK, B. MAEGAARD, J. MARIANI, H. MAZO, A. MORENO, J. ODIJK & S. PIPERIDIS (eds.), *Proceedings of the 10th International Conference on Language Resources and Evaluation – LREC 2016*. Luxembourg: European Language Resources Association: 1008-1013. Available online: http://www.lrec-conf.org/proceedings/lrec2016/pdf/274_Paper.pdf.
- SCHIFFRIN, D. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.
- SCOTT, D.R. & SOUZA, C.S. DE 1990. Getting the Message across in RST-Based Text Generation. In R. DALE, C.S. MELLISH & M. ZOCK (eds.), *Current Research in Natural Language Generation*. New York: Academic Press: 47-73.
- SPORLEDER, C. & LASCARIDES, A. 2005. Exploiting Linguistic Cues to Classify Rhetorical Relations. In G. ANGELOVA, K. BONTCHEVA, R. MITKOV, N. NICOLOV & N. NIKOLOV (eds.), *Proceedings of Recent Advances in Natural Language Processing – RANLP 2005 (21-23 September 2005, Borovets, Bulgaria)*. Shoumen: INCOMA: 532-539. Available online: <http://lml.bas.bg/ranlp2005/DOCS/RANLP2005.pdf>.
- SPORLEDER, C. & LASCARIDES, A. 2008. Using Automatically Labelled Examples to Classify Rhetorical Relations: An Assessment. *Natural Language Engineering* 14 (3): 369-416.
- TABOADA, M. 2006. Discourse Markers as Signals (or Not) of Rhetorical Relations. *Journal of Pragmatics* 38 (4): 567-592.
- TABOADA, M. 2009. Implicit and Explicit Coherence Relations. In J. RENKEMA (ed.), *Discourse, of Course: An Overview of Research in Discourse Studies*. Amsterdam – Philadelphia: J. Benjamins: 127-140.
- TABOADA, M. & MANN, W.C. 2006. Rhetorical Structure Theory: Looking Back and Moving Ahead. *Discourse Studies* 8 (3): 423-459.
- THEIJSEN, D. 2007. *Features for Automatic Discourse Analysis of Paragraphs*. Master thesis. Nijmegen: Radboud University Nijmegen, Department of Linguistics. Available online: http://daphnetheijssen.ruhosting.nl/aboutme/MA_thesis_DTheijssen.pdf.
- THEIJSEN, D., VAN HALTEREN, H., VERBERNE, S. & BOVES, L. 2008. Features for Automatic Discourse Analysis of Paragraphs. In S. VERBERNE, H. VAN HALTEREN & P.-A. COPPEN (eds.), *Computational Linguistics in the Netherlands 2007: Selected Papers from the 18th CLIN Meeting (CLIN 2007)*. Utrecht: LOT: 53-68. Available online: <https://dspace.library.uu.nl/bitstream/handle/1874/296792/bookpart.pdf?sequence=2&isAllowed=y>.
- VAN DER VLIET, N. & REDEKER, G. 2014. Explicit and Implicit Coherence Relations in Dutch Texts. In H. GRUBER & G. REDEKER (eds.), *The Pragmatics of Discourse Coherence: Theories and Applications*. Amsterdam – Philadelphia: J. Benjamins: 23-52.
- VERSLEY, Y. 2013. Subgraph-Based Classification of Explicit and Implicit Discourse Relations. In A. KOLLER & K. ERK (eds.), *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers*. Stroudsburg: Association for Computational Linguistics: 264-275. Available online: <https://www.aclweb.org/anthology/W13-0123>.

- WEBBER, B.L., PRASAD, R., LEE, A. & JOSHI, A. 2019. *The Penn Discourse Treebank 3.0 Annotation Manual*. Philadelphia: University of Pennsylvania. Available online: <https://catalog.ldc.upenn.edu/docs/LDC2019T05/PDTB3-Annotation-Manual.pdf>.
- ZEYREK, D., DEMİRŞAHİN, I., SEVDİK-ÇALLI, A.B., BALABAN, H.Ö., YALÇINKAYA, I. & TURAN, Ü.D. 2010. The Annotation Scheme of the Turkish Discourse Bank and an Evaluation of Inconsistent Annotation. In N. XUE & M. POESIO (eds.), *Proceedings of the Fourth Linguistic Annotation Workshop (LAW IV) – ACL 2010 (15-16 July 2010, Uppsala, Sweden)*. Stroudsburg: Association for Computational Linguistics: 282-289. Available online: <https://www.aclweb.org/anthology/W10-1844>.
- ZEYREK, D., MENDES, A. & KURFALI, M. 2018. Multilingual Extension of PDTB-Style Annotation: The Case of TED Multilingual Discourse Bank. In N. CALZOLARI, K. CHOUKRI, C. CIERI, T. DECLERCK, K. HASIDA, H. ISAHARA, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODIJK, S. PIPERIDIS, T. TOKUNAGA, S. GOGGI & H. MAZO (eds.), *Proceedings of the 11th International Conference on Language Resources and Evaluation – LREC 2018*. Luxembourg: European Language Resources Association: 1913-1919. Available online: <https://www.aclweb.org/anthology/L18-1301>.