Nuclearity in RST and signals of coherence relations

Debopam Das
Department of English and American Studies
Humboldt University of Berlin, Germany
dasdebop@hu-berlin.de

Abstract

We investigate the relationship between the notion of nuclearity as proposed in Rhetorical Structure Theory (RST) and the signalling of coherence relations. RST relations are categorized as either mononuclear (comprising a nucleus and a satellite span) or multinuclear (comprising two or more nuclei spans). We examine how mononuclear relations (e.g., Antithesis, Condition) and multinuclear relations (e.g., Contrast, List) are indicated by relational signals, more particularly by discourse markers (e.g., *because, however, if, therefore*). We conduct a corpus study, examining the distribution of either type of relations in the RST Discourse Treebank (Carlson et al., 2002) and the distribution of discourse markers for those relations in the RST Signalling Corpus (Das et al., 2015). Our results show that discourse markers are used more often to signal multinuclear relations than mononuclear relations. The findings also suggest a complex relationship between the relation types and syntactic categories of discourse markers (subordinating and coordinating conjunctions).

1 Introduction

Nuclearity in Rhetorical Structure Theory (RST) is explained in terms of relative importance of text spans (Mann and Thompson, 1988). The span perceived (by the reader) to be more important or central to the writer's purpose is called the nucleus, and the span perceived to be less important or peripheral to the writer's purpose is called the satellite. RST relations having spans with equal and unequal importance are known as multinuclear and mononuclear relations, respectively. Examples of multinuclear relations include Contrast, List or Sequence, and examples of mononuclear relations include Condition, Elaboration, Evidence or Summary. The notion of nuclearity in RST represents a symmetric-asymmetric divide, which also parallels with the distinction between non-hierarchical and hierarchical relations: Multinuclear relations are symmetrical or non-hierarchical relations, and mononuclear relations are asymmetric or hierarchical relations.

Coherence relations, whether multinuclear or mononuclear, are often signalled by discourse markers (henceforth DMs)[1]. For example, a Contrast relation (multinuclear) can be indicated by the DM *but*, and an Evidence relation (mononuclear) can be conveyed through the DM *because*. Research on the signalling phenomenon in discourse has, however, more recently shown that coherence relations can well be indicated by other textual signals such as certain lexical expressions or syntactic features, both in addition to or in the absence of DMs (Das and Taboada, 2018). For instance, in the Penn Discourse Treebank (PDTB 3.0) (Webber et al., 2018, p. 10) the Condition relation between the text segments (within square brackets) in Example 1 is conveyed through the use of auxiliary inversion (underlined)[2]:

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

In this paper, we investigate the relationship between the notion of nuclearity in RST and the signalling of coherence relations. We examine whether nuclearity has a role to play in relation marking, and whether multinuclear and mononu-

[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 connectives' is deemed to be more appropriate, we prefer to use the term discourse markers in the spirit of the RST Signalling Corpus (Das et al., 2015), which we base our analyses on.

[2] In the 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.

Proceedings of Discourse Relation Parsing and Treebanking (DISRPT2019), pages 30–37
Minneapolis, MN, June 6, 2019.

clear relations differ in terms of signalling. However, since a complete analysis of all kinds of relational signals is beyond the scope of the present paper, we constrain our analysis only to DMs, and do not consider any other types of signalling[3]. We address the following research questions:

1. How are mononuclear and multinuclear relations signalled in text?
2. Does one category employ more DMs than the other?
3. What types of DMs (subordinating and coordinating conjunctions) are used to indicate these two categories of relations?

DMs, although primarily representing a functional category, are generally considered to belong to four different syntactic classes: coordinating conjunction (like *and* and *but*, as when they connect two coordinated clauses), subordinating conjunction (like *if* and *since*, as when they connect a subordinate adjunct clause to a main clause), prepositional phrases (like *in addition* and *as a result*, as when they connect two main clauses or sentences), and adverbial phrases (like *however* and *nevertheless*, as used much like the above-mentioned prepositional phrases). The question that we aim to address here is to what degree these canonical signal types correspond to the categories of nuclearity in RST. More simply, we examine, for example, to what extent coordinated conjunctions are used to indicate multinuclear relations, and to what extent subordinated conjunctions are used to signal mononuclear relations. For this purpose, we examine the signalling of mononuclear and multinuclear relations in the RST Signalling Corpus (Das et al., 2015), a corpus annotated for relational signals, which is built upon the RST Discourse Treebank (Carlson et al., 2002), a corpus annotated for coherence relations.

The paper is organized as follows: In Section 2, we provide the distribution of mononuclear and multinuclear relations in the RST Discourse Treebank. Section 3 provides a brief introduction to the RST Signalling Corpus, with a special focus on DMs. In Section 4, we present the results, reporting on the distributions of mononuclear and multinuclear relations with respect to DMs. Section 5 reflects on the implications of the results, and outlines a few potential future developments of this work. Finally, Section 6 summarizes the paper, and provides the conclusion.

[3]We believe that the signalling of relations beyond discourse markers constitutes an important topic, and is worthy of investigation in its own right. We discuss the prospects of conducting a similar analysis of other signals for nuclearity in Section 5.

Project	# rel	# mono	# multi	# both
M&T[4]	23	21	2	0
PCC[5]	31	26	5	0
GUM[6]	20	16	3	1
Span TB[7]	28	22	6	0
DiZer[8]	32	26	6	0
Website[9]	25	21	4	0
RST-DT[10]	78	53	8	17

Table 1: Distribution of mononuclear and multinuclear RST relations in RST-based studies

2 Nuclearity and RST-DT

In RST-based research, just like the way relational inventories differ from studies to studies, so does the number of mononuclear and multinuclear relations within an inventory, as shown in Table 1.

In this study, we examine the relations from the RST Discourse Treebank (henceforth the RST-DT) (Carlson et al., 2002). The corpus, as the distribution (in Table 1) shows, uses a large set of 78 relations, including 53 mononuclear and 8 multinuclear relations. Most importantly, unlike other RST-based projects (in Table 1) that only distinguish between mononuclear and multinuclear relations (exception: the GUM corpus), the RST-DT includes an additional category for relations that can appear as both mononuclear or multinuclear[11]. The taxonomy of the RST-DT relations in terms of nuclearity is provided in Table 2.

The RST-DT contains a total of 20,123 relations, which were expanded to 21,400 relations for the signalling annotation in the RST Signalling Corpus (Das and Taboada, 2017), as a result of complying with a strict binary branching requirement and thus breaking a multinuclear relation

[4]Mann and Thompson (1988)
[5]Potsdam Commentary Corpus (Stede, 2016)
[6]The GUM corpus (Zeldes, 2017)
[7]RST Spanish Treebank (da Cunha et al., 2011)
[8]DIscourse analyZER for Brazilian Portuguese (Maziero et al., 2011)
[9]RST website (`http://www.sfu.ca/rst/`)
[10]RST Discourse Treebank (Carlson et al., 2002)
[11]The assignation of nuclearity status on a particular span can sometimes be a matter of considerable difficulty, and the inclusion of the both mono and multi versions of relations in the RST-DT, as Stede (2008) suggests, provided the RST-DT annotators greater freedom in choosing what spans should be labeled nuclear.

Type	Relation
mono	Antithesis, Attribution, Background, Cause, Circumstance, Comment, Concession, Condition, Contingency, Definition, Elaboration-additional, Elaboration-set-member, Elaboration-part-whole, Elaboration-process-step, Elaboration-object-attribute, Elaboration-general-specific, Enablement, Evidence, Example, Explanation-argumentative, Hypothetical, Manner, Means, Otherwise, Preference, Purpose, Restatement, Result, Rhetorical-question, Summary, Temporal-after, Temporal-before
multi	Contrast, Cause-Result, Comment-topic, Disjunction, Inverted-sequence, List, Otherwise, Proportion, Same-unit, Sequence, Textual-organization, Topic-comment
both	Analogy, Comparison, Conclusion, Consequence, Evaluation, Interpretation, Problem-solution, Question-answer, Reason, Statement-response, Temporal-same-time, Topic-drift, Topic-shift

Table 2: Relation types in RST-DT

having more than two nuclei into more than one (multinuclear) relation. These 21,400 relations are divided into 16,526 mononuclear and 4,874 multinuclear relations.

3 RST Signalling Corpus

The RST Signalling Corpus (henceforth the RST-SC) (Das et al., 2015) provides signalling annotation for the coherence relations that are present in the RST-DT. The RST-SC implements a wide perspective of signalling, and provides annotation for a large variety of textual signals, such as reference, lexical, semantic, syntactic, graphical and genre-related features, in addition to DMs. 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 nine signal types (e.g., *reference*, *syntactic*, *graphical*). Finally, the third level in the hierarchy refers to specific signals; for example, *reference* type has four specific signals: *personal, demonstrative, comparative,* and *propositional reference*. The hierarchical organization of the taxonomy is provided in Figure 1[12].

The distribution of relations by signals in the RST-SC (in Table 3, from Das and Taboada (2018)) shows that an 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. Only 3,896 (2,280 + 1,616) relations out of 21,400 relations (18.21% of all relations) are indicated by DMs[13]. These DMs are distributed across 201 different types, which can be further divided into coordinating conjunctions, subordinating conjunctions, prepositional and adverbial phrases.

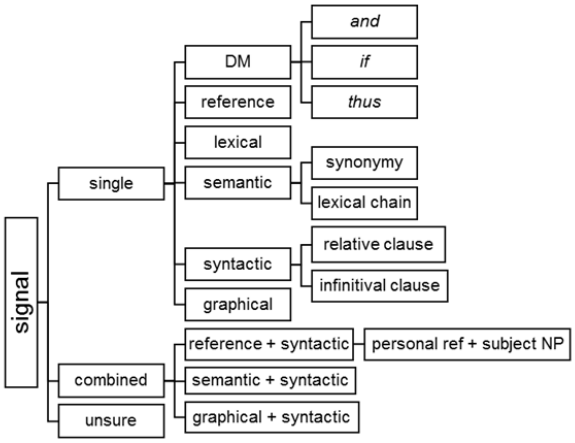

Figure 1: Hierarchical taxonomy of signals in RST-SC

In order to examine the relationship between nuclearity and DMs, we extract from the RST-SC (i) instances of all DMs, and (2) instances of different relations that are indicated by those DMs. A complete analysis of the relationship between the relations (78 types) and DMs (201 types) in the RST-SC, however, could not be covered in the present paper. That is why, we focus only on the most frequently occurring DMs and most frequently occurring relations in the corpus. In order to extract those tokens, we use UAM CorpusTool (O'Donnell, 2008), which was also used to annotate the RST-SC. 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.

[12]Note that subcategories in the figure are only illustrative, not exhaustive. For the detailed taxonomy and definitions of signals, see Das (2014).

[13]One possible reason for the RST-SC for having a lower proportion of DMs than other comparable corpora is that the RST-SC employs a much stricter definition for DMs. For example, the PDTB 3.0 corpus (Webber et al., 2018), containing 45.93% relations with explicit connectives, uses more flexible parameters in the connective definition, and includes very frequently occurring words such as *also, by, from, in* or *like* as connectives. In contrast, the RST-SC considers these expressions not as DMs, but as lexical signals (more specifically, indicative words).

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 3: Distribution of signalled and unsignalled relations in RST-SC

4 Results

We first examine how mononuclear and multinuclear relations are distributed in the RST-SC with respect to signalling. The distribution in Table 4 shows that both mononuclear or multinuclear relations most often contain signals (over 90% of the relations). This also shows, however, that when it comes to the signalling by DMs, multinuclear relations are more often indicated by DMs than mononuclear relations: About 30% of the multinuclear relations contain a DM while only about 15% of the mononuclear relations occur with them.

Next, we find in the RST-SC the relations that are most frequently signalled by DMs (with respect to their overall frequencies in the corpus), and examine what DMs are commonly used to signal those relations. We provide the distribution of DMs for mononuclear relations in Table 5 and the distribution for multinuclear relations in Table 6. The number within parentheses after a relation name in column 1 (labeled Relation, in both tables) refers to the number of instances the relation occurs with a DM in the corpus. The number in column 3 (labeled #, in both tables) refers to the number of instances a DM (in the corresponding row) is used for marking the relation. (Note: CC = coordinating conjunction; SC = subordinating conjunction; PP = preposition (-al phrase); ADV = adverb (-ial) phrase)

Table 5 shows that a mononuclear relation is indicated by different DMs[14], which belong to different syntactic classes (e.g., CC, SC or ADV). For example, Result relations are commonly signalled by the DMs *because* (SC), *and* (CC) and *as a result* (ADV). Similar distribution of DMs for multinuclear relations is shown in Table 6. For example, Temporal-same-time relations[15] are commonly marked by the DMs *while* (SC), *as* (SC) and *and* (CC).

Finally, we extract the most frequently used DMs in the RST-SC, and examine the relations that are signalled by them. In Table 7, we provide the distribution of the common relations for those DMs. The number in column 2 (labeled # DM) refers to the number of instances of a DM in the RST-SC, and the number in column 5 (labeled # Rel) refers to the number of instances for a relation indicated by that DM in the corpus. The distribution shows that the DMs *and* and *but* (both CC) are the two most frequent DMs (with over 600 tokens), followed by other DMs like *as* (SC), *if* (SC), *when* (SC), *because* (SC) and *however* (ADV). As we have seen in Table 5 and 6 that relations are indicated by a wide variety of DMs, Table 7 shows the opposite is also true: Each of the DMs in Table 7 indicates more than one relation in the corpus[16]. Furthermore, the relations indicated by these DMs are distributed for mononuclear and multinuclear categories. For example, the DM *while* is commonly used to indicate, on the one hand, Antithesis and Concession relations, which are mononuclear relations, and on the other hand, Contrast and List relations, which are multinuclear relations.

5 Discussion

As our results show (in Table 4), although over 90% of the RST relations in the RST-SC, regardless of their types (mononuclear or multinuclear), contain some sort of signals, only about

[14]The relations presented in Table 5 (and also Table 6) are indicated by an even wider variety of DMs in the RST-SC (see Das (2014)). The distribution here only records the most frequently used DMs (common DMs) for those relations.

[15]Temporal-same-time relations can be both mononuclear and multinuclear (see Table 2). Table 6 provides the distribution of DMs for Temporal-same-time when it is used as a multinuclear relation.

[16]In Table 7, the DM *if* is shown to indicate only Condition relations. In the RST-SC, however, *if* is also found to signal other relations, such as Circumstance, Contingency or Hypothetical (although with relatively lower frequencies).

Type	Total #	# signalled	% signalled	# with DM	% with DM
mono	16526	15424	93.33	2415	14.61
multi	4874	4423	90.75	1481	30.39

Table 4: Distribution of mononuclear and multinuclear relations by signalling

Relation	DM	#	Type
Concession (264)	but	100	CC
	although	28	SC
	despite	24	PP
	though	24	SC
	while	17	SC
Condition (221)	if	162	SC
	unless	12	SC
Temporal-before (38)	before	31	SC
Antithesis (330)	but	182	CC
	although	28	SC
	however	26	ADV
	though	11	ADV
Temporal-after (69)	after	48	SC
Temporal-same-time (63)	when	29	SC
	as	18	SC
	while	13	SC
Result (87)	because	25	SC
	and	23	CC
	as a result	19	ADV
Reason (112)	because	65	SC

Table 5: Common DMs for mononuclear relations

Relation	DM	#	Type
Disjunction (26)	or	19	CC
Temporal-same-time (52)	while	14	SC
	as	13	SC
	and	9	CC
Contrast (305)	but	186	CC
	however	22	ADV
	while	20	SC
	and	17	CC
Cause-result (42)	and	15	CC
	because	11	SC
Sequence (119)	and	69	CC
	then	20	ADV
List (818)	and	698	CC
	but	19	CC
	while	16	SC

Table 6: Common DMs for multinuclear relations

15% mononuclear and about 30% multinuclear relations contain a DM. This, however, implies that most often both mononuclear and multinuclear relations are conveyed by other textual signals. This is, we believe, an important issue to consider, and we will touch upon this point after we discuss our findings about DMs.

The crucial difference between mononuclear and multinuclear relations for signalling lies in the proportions of each type of relations containing a DM (15% vs. 30%). We observe that relations that differ according to nuclearity also differ with respect to two additional factors. First, all RST taxonomies (as shown in Table 1) contain significantly higher number of mononuclear relations than multinuclear relations (e.g., the GUM corpus (Zeldes, 2017) has 16 mononuclear (80%), but only 4 multinuclear (20%) relations). Furthermore, with respect to the number of tokens in a corpus, the mononuclear relations also outnumber multinuclear relations. For example, out of 21,400 relations in the RST-DT (Carlson et al., 2002), there are 16,526 mononuclear (77.22%) and only 4,874 multinuclear (22.78%) relations. The relatively lower number of multinuclear relations, both in RST taxonomies and corpora, may imply that mononuclear relations are more basic type of relations than multinuclear relations. If that is borne out, then it might also be case that when relations are multinuclear, they would require more DMs as their signals than mononuclear relations.

The distribution of DMs for mononuclear and multinuclear relations (in Table 5 and 6) shows a complex co-occurrence pattern of nuclearity type and the syntactic membership of DMs. On the one hand, we observe (in Table 5) that mononuclear relations are often conveyed by subordinating conjunctions (SCs). This is evidenced by relations such as Condition, Reason and Temporal-same-time (when used as a mononuclear relation) that (exclusively) employ SCs (among DMs) as their signals. Similarly (in Table 6), a strong association between multinuclear relations and coordinating conjunctions (CCs) is observed for Disjunction which is indicated by the CC *or*.

DM	# DM	Type	Relation	# Rel	Type
and	1043	CC	List	698	multinuclear
			Elaboration-additional	76	mononuclear
			Sequence	66	multinuclear
			Consequence	42	mononuclear
			Circumstance	20	mononuclear
but	615	CC	Contrast	186	multinuclear
			Antithesis	182	mononuclear
			Concession	100	mononuclear
			Elaboration-additional	48	mononuclear
			List	19	multinuclear
if	180	SC	Condition	162	mononuclear
when	168	SC	Circumstance	109	mononuclear
			Temporal-same-time	22	mononuclear
as	166	SC	Circumstance	64	mononuclear
			Temporal-same-time	18	mononuclear
			Comparison	15	mononuclear
because	162	SC	Reason	64	mononuclear
			Explanation-argumentative	35	mononuclear
			Consequence	21	mononuclear
			Result	14	mononuclear
			Cause-result	11	multinuclear
while	131	SC	Antithesis	24	mononuclear
			Contrast	20	multinuclear
			List	16	multinuclear
			Concession	17	mononuclear
			Temporal-same-time	14	multinuclear
after	101	SC	Temporal-after	48	mononuclear
			Circumstance	37	mononuclear
however	92	ADV	Antithesis	26	mononuclear
			Contrast	22	multinuclear
			Elaboration-additional	14	mononuclear
			Concession	11	mononuclear
because of	81	SC	Consequence	21	mononuclear
			Reason	19	mononuclear
			Result	18	mononuclear
although	62	SC	Antithesis	28	mononuclear
			Concession	28	mononuclear
before	60	SC	Temporal-before	31	mononuclear
			Circumstance	14	mononuclear
without	51	PP	Circumstance	21	mononuclear
			Manner	19	mononuclear

Table 7: Common relations for DMs

On the other hand, we observe (in Table 5 and 6) that the opposite pattern also holds, that is, mononuclear relations are often signalled by CCs and multinuclear relations frequently contain SCs. For example, a large proportion of Concession or Antithesis (both mononuclear relations) employ the CC *but* as their signal. Similarly, Temporal-same-time (as a multinuclear relation) are mostly indicated by the SCs *while* and *as*.

The complex nature of the co-occurrence of nuclearity types and DM types is further illustrated by Table 7 that presents the distribution of common relations for most frequent DMs in the RST-SC. For example, the SC *while* is used to indicate both mononuclear relations (Antithesis or Concession) and multinuclear relations (Contrast or List).

In sum, DMs are found to signal multinuclear relations more often than mononuclear relations. However, with respect to the DM types, mononuclear and multinuclear relations are indicated by both SCs ad CCs, without having any strong commitment to either type of DMs[17]. The latter finding is in line with Blühdorn (2008), who finds that hierarchy and non-hierarchy at the syntactic level (represented by subordination and coordination, respectively) does not systematically correspond to hierarchical and non-hierarchical coherence relations (in effect, mononuclear and multinuclear relations, respectively) at the discourse level.

Theoretically, the nuclearity status of a span (nucleus or satellite) in a relation is assigned by evaluating it against the other span in terms of to what degree the span is important to the intention of the writer. In practice, however, determining the relative importance of spans may not be a straightforward task. Stede (2008) identifies different factors that influence RST annotators to decide on the nuclearity status of the text segments. These factors include intention of the writer (represented in the nucleus and supported by the satellite), recurrence of an idea across different parts of a text (as a sign of emphasizing importance for a span), digression from the main topic (as a sign of less importance for a span), connectives (in German) and punctuation (e.g., parentheses) that can mark the nucleus-satellite distinction, syntactic structure (main clause vs. subordinate clause), or the RST relation definitions themselves that prescribe the nuclearity status for a span (e.g., reporting clause as the satellite for Attribution relations). If these sources really contribute to identify the nucleus (or distinguish between the nucleus and satellite), an important venture could be to examine whether or how signalling interacts with these factors. It seems that some of the factors are closely associated with the signalling phenomenon. For example, as our results show, certain DMs (or connectives) such as *if* or *although*, which are SCs, are always used to convey mononuclear relations.

The association of potential sources of nuclearity and relation marking can possibly be made more substantial if we adopt a wider perspective of signalling, incorporating other means of signalling beyond DMs. As mentioned in Section 3, the RST-SC exploits many different types of signals, and we argue here that some of these signals may well be correlated with some of the factors affecting nuclearity. We provide a few examples to illustrate this. Syntactic signals such as auxiliary inversion (as shown in Example 1) or certain type of subordinate clauses (e.g., participial or infinitival) may exhibit a strong correlation with the factor *syntactic structure*, as suggested by Stede (2008). Also, parallel syntactic constructions (e.g., *Chris is tall; Pat is short.*) can indicate or predict the presence of a multinuclear relation. Similarly, a graphical signal such as an itemized list (called *items in sequence* in the RST-SC) can be used to signal a multinuclear (List) relation, while the content within parentheses (as also suggested by Stede (2008)) can refer to a satellite span. Furthermore, a reference feature (encoding a co-reference chain) or semantic feature (representing a lexical chain) can indicate the presence of a mononuclear relation (e.g., Elaboration or Restatement). We leave an exploration of the interaction of other relational signals and nuclearity as one of our future endeavors from this study.

Furthermore, a rather specific query about the relationship between nuclearity and signalling relates to the location of the signals, that is, where the signals occur – in nucleus, in satellite, or in both spans. We would like to examine, more particularly, which signal occurs in which span, and how frequently they occur in one (as opposed to in the other) span.

We envisage another related line of develop-

[17] DMs can also belong to two other syntactic classes, PPs or ADVs. However, since we find only a few DMs of these types (four ADVs (*however, though, as a result* and *then*) and two PPs (*despite* and *without*)), we do not include them in the present analysis.

ment concerning what is suggested by Marcu (2000) as the 'strong nuclearity hypothesis'. According to this hypothesis, it is postulated that when a relation holds between two (composite) text spans, it should also hold between the nuclei of those two spans. We would like to examine whether it is possible to motivate the 'strong nuclearity hypothesis' by evidence from the signalling of RST relations. The relevant question to address here would be if we disregard all the satellites in an RST analysis, whether we would still have relevant signals left in the remaining nuclei that can indicate the relations between spans.

6 Conclusion

In this paper, we have investigated how the notion of nuclearity correlates with the signalling of coherence relations by discourse markers, which are generally considered to be the most explicit and reliable signals of coherence relations. Based on a corpus analysis of RST relations and relational signals, we have examined how mononuclear and multinuclear relations are signalled by discourse markers. Our results have shown that multinuclear relations are indicated more frequently by discourse markers than mononuclear relations. However, we did not find conclusive evidence as to whether these two relation types are more or less conveyed through coordinating or subordinating conjunctions, the two primary categories of discourse markers. In order to address the complex relationship between nuclearity and signalling more adequately, we have argued for the need to incorporate in the analysis other types of relational signals (such as syntactic, graphical or reference features), which might demonstrate a more substantial correlation between the notion of nuclearity in RST and the signalling of coherence relations.

References

Hardarik Blühdorn. 2008. Subordination and coordination in syntax, semantics, and discourse: Evidence from the study of connectives. In Cathrine Fabricius-Hansen and Wiebke Ramm, editors, *'Subordination' versus 'Coordination' in Sentence and Text: A cross-linguistic perspective*, pages 59–85. John Benjamins, Amsterdam.

Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. 2002. RST Discourse Treebank, LDC2002T07. Philadelphia. Linguistic Data Consortium.

Iria da Cunha, Juan-Manuel Torres-Moreno, and Gerardo Sierra. 2011. On the Development of the RST Spanish Treebank. In *Proceedings of the 5th Linguistic Annotation Workshop*, LAW V '11, pages 1–10, Stroudsburg, PA, USA. Association for Computational Linguistics.

Debopam Das. 2014. *Signalling of Coherence Relations in Discourse*. Phd dissertation, Simon Fraser University, Canada.

Debopam Das and Maite Taboada. 2017. RST Signalling Corpus: A corpus of signals of coherence relations. *Language Resources Evaluation*, 52(1):149–184.

Debopam Das and Maite Taboada. 2018. Signalling of Coherence Relations in Discourse, Beyond Discourse Markers. *Discourse Processes*, 55(8):743–770.

Debopam Das, Maite Taboada, and Paul McFetridge. 2015. RST Signalling Corpus, LDC2015T10. Philadelphia. Linguistic Data Consortium.

William C. Mann and Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Daniel Marcu. 2000. *The Theory and Practice of Discourse Parsing and Summarization*. MIT Press, Cambridge, MA.

Erick Galani Maziero, Thiago Alexandre Salgueiro Pardo, Iria da Cunha, Juan-Manuel Torres-Moreno, and Eric SanJuan. 2011. DiZer 2.0 An Adaptable On-line Discourse Parser. In *Proceedings of the III RST Meeting (8th Brazilian Symposium in Information and Human Language Technology*, pages 1–17, Cuiabá/MT, Brazil.

Michael O'Donnell. 2008. The UAM Corpustool: Software for corpus annotation and exploration. In *Proceedings of the XXVI Congreso de AESLA*, pages 3–5, Almeria, Spain.

Manfred Stede. 2008. RST revisited: Disentangling nuclearity. In Cathrine Fabricius-Hansen and Wiebke Ramm, editors, *'Subordination' versus 'Coordination' in Sentence and Text: A cross-linguistic perspective*, pages 33–58. John Benjamins, Amsterdam.

Manfred Stede. 2016. Rhetorische Struktur. In Manfred Stede, editor, *Handbuch Textannotation: Potsdamer Kommentarkorpus 2.0*. Universitätsverlag, Potsdam.

Bonnie Webber, Rashmi Prasad, Alan Lee, and Aravind Joshi. 2018. The Penn Discourse Treebank 3.0 Annotation Manual. Report, The University of Pennsylvania.

Amir Zeldes. 2017. The GUM corpus: creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51:581–612.