LACUS Forum **40:3.8-20 (2020)**

SIGNALING SUBJECT MATTER AND PRESENTATIONAL COHERENCE RELATIONS IN DISCOURSE: A CORPUS STUDY

DEBOPAM DAS & MAITE TABOADA
Simon Fraser University

Abstract: In this study we examine how subject matter and presentational coherence relations in Rhetorical Structure Theory (Mann & Thompson 1988) are signaled in written discourse, and whether they differ quantitatively or qualitatively in terms of the signaling devices involved. By signaling we mean textual signals (discourse markers such as *although, because,* and *thus*, and also signals such as *tense, lexical chains,* or *punctuation*) that indicate a relation is present. We hypothesize that, because of their different nature, subject matter and presentational relations may be indicated by different signals, or at different frequencies. We conducted a corpus study examining 40 articles, comprising 1,306 relations, from the RST Discourse Treebank (Carlson, Marcu, & Okurowski 2002). We identified the signals for those relations and added a new layer of annotation to them, to include signaling information. Results from our corpus analysis show that subject matter and presentational relations do not differ quantitatively or qualitatively in terms of signaling. These findings, although negative, point to a number of theoretical possibilities about the validity and accuracy of subject matter and presentational relation classification.

Keywords: Rhetorical Structure Theory, subject matter relations, presentational relations, signals of relations, corpus study, RST Discourse Treebank

Languages: English

THE INTERPRETATION OF A TEXT requires not only an understanding of the meanings of individual discourse components (clauses, sentences, etc.), but also an understanding of how the meanings of those components are connected to each other. Coherence relations (also known as discourse or rhetorical relations) refer to the types of semantic or pragmatic connections that bind one discourse component to another. For example, in the following text,

(1) John could not go to the party. He was busy with his work.

there are two components: (i) *John could not go to the party* and (ii) *He was busy with his work*. These components are connected to each other by a *causal* relation: John's inability to go to the party is caused by the fact that he was doing his work.

1. COHERENCE RELATIONS AND RHETORICAL STRUCTURE THEORY. Coherence relations have been extensively investigated in the framework of Rhetorical Structure Theory or RST (Mann & Thompson 1988). RST is a functional theory of text organization. It describes what parts a text is made of, what kinds of relationships exist between these parts, and how these parts are organized with respect to each other to constitute a coherent piece of discourse. RST originally emerged as computational theory for analyzing text structure, but later the use of RST has been extended to various other applications, including the analysis of spoken discourse. In RST, relations are defined through different fields, the most important of which is the *Effect* (achieved on the text receiver), referring to the intention of the writer (or speaker) in presenting their discourse. Relation inventories are open, and the most common relation taxonomies include names such as *Cause, Concession, Condition, Elaboration, Result* and *Summary*. Relations can be multinuclear, reflecting a paratactic relationship, or nucleus-satellite, a hypotactic type of relation. The names nucleus and satellite refer to the relative importance of each of the relation components.

Texts, according to RST, are built out of clausal units (smallest discourse components) that enter into rhetorical (coherence) relations with each other, in a recursive manner. Mann and Thompson proposed that most texts can be analyzed in their entirety as recursive applications of different types of relations. In effect, this means that an entire text can be analyzed as a tree structure, with clausal units being the branches and relations the nodes.

In **Figure 1**, we provide an RST analysis of a text taken from the RST Discourse Treebank (Carlson, Marcu, & Okurowski 2002).

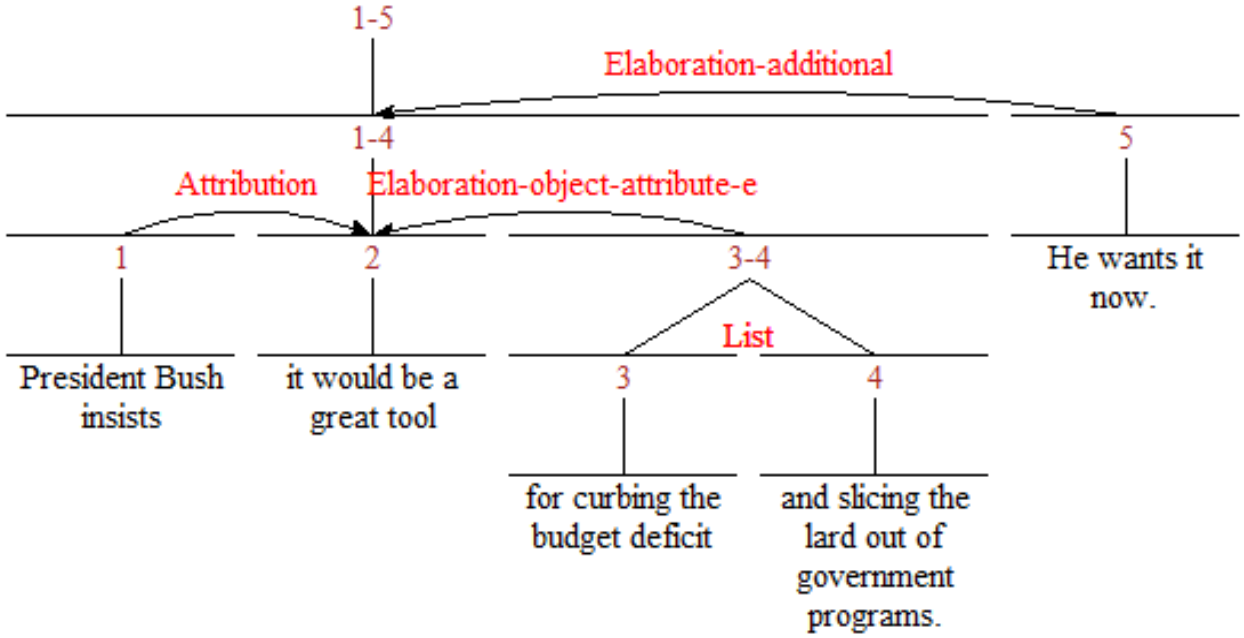

Figure 1. *Graphic representation of an RST analysis*

The RST analysis in **Figure 1** shows that the text comprises five spans which are represented by the cardinal numbers, 1, 2, 3, 4, and 5. In the diagram, the arrowhead points of

spans refer to the nuclei, and the arrow points away from another span that refers to the satellites. Span 1 (satellite) is connected to Span 2 (nucleus) by an *Attribution* relation, and together they make the combined span 1-2. Span 3 (nucleus) and 4 (nucleus) are in a multinuclear *List* relation, and together they make the combined span 3-4. Span 3-4 (satellite) is connected to span 1-2 (nucleus) by an *Elaboration* (more specifically, *Elaboration-object-attribute-e*) relation, and together they make the combined span 1-4. Finally, span 5 (satellite) is connected to span 1-4 (nucleus) by an *Elaboration* (more specifically, *Elaboration-additional*) relation. For more information about the relational definitions, see Carlson & Marcu (2001).

2. SUBJECT MATTER AND PRESENTATIONAL RELATIONS IN RST. Relations in RST, in terms of their intended effects, are divided into two groups: subject matter relations (e.g., *Elaboration, Circumstance, Solutionhood, Cause, Restatement*) and presentational relations (e.g., *Motivation, Background, Justification, Concession*). In subject matter relations, the text producer intends the reader to understand the relation. In presentational relations, on the other hand, the intended effect is to increase some inclination in the reader (positive regard, belief, or acceptance of the nucleus). The distinction is related to the semantic/pragmatic divide proposed by van Dijk to classify discourse connectives according to what type of relation they signal: "[p]ragmatic connectives express relations between speech acts, whereas semantic connectives express relations between denoted facts" (van Dijk 1979: 449). The distinction is also represented by the following labels (corresponding to subject matter and presentational relations, respectively):

- External and Internal (Halliday and Hasan 1976a, Martin 1992).
- Semantic and Pragmatic (Sanders, Spooren, and Noordman 1992a, 1993, Schiffrin 1987, van Dijk 1979, 1977).
- Ideational and Interpersonal. This is a distinction proposed in Systemic Functional Linguistics (Halliday 1985), and applied to coherence relations by others (Maier and Hovy 1993, Redeker 2000). The SFL distinction includes a third category, Textual, which does not fit clearly against the other dual classifications.
- Experiential and Rhetorical (Benwell 1999).
- Causal and Diagnostic, restricted to relations linked by connectives such as *because*, e.g., *The streets are wet because it is raining* versus *It is raining, because the streets are wet* (Traxler et al. 1997).

As it happens with most attempts at classification, the different classifications above represent distinctions that are not fully equivalent. The semantic/pragmatic distinction proposed by van Dijk (1979) separates relations between propositions (semantic) from those between speech acts (pragmatic). This distinction seems to be orthogonal to the subject matter/presentational divide in RST, since both propositions and speech acts can be in the same types of relations. Events denoted by propositions may be related through *Cause* (subject matter type) or *Background* (presentational type) relations; speech acts can be equally related through either type of relation. RST has concentrated on intended effects, more than on how they are achieved, whether it is through presenting relations containing events, facts, propositions or speech acts.

3. RESEARCH GOALS. In this study, we examine how subject matter and presentational relations are signaled, and whether they differ quantitatively and qualitatively in terms of the types of signaling involved. By signaling we mean textual signals that indicate a relation is present. These include

discourse markers or DMs (e.g., *although, because, since, thus*), but also signals such as *tense, lexical chains*, or *punctuation*. We hypothesize that, because of their different nature, subject matter and presentational relations may be indicated by different types of signals or at different frequencies. Because subject matter relations are meant to be recognized as such, they may have explicit signals, whereas in presentational relations the connection is more abstract, and the effect probably stems more out of the content than the form.

Research on coherence relations has often focused on cues that indicate the presence of a relation, or the lack of such cues, as many relations seem to be unsignaled or implicit. Whereas it is true that many coherence relations are not signaled by a DM, it is also often the case that other signals have been understudied (Taboada and Mann 2006, Taboada 2009). We explore how many and what types of cues can be found if we study signaling beyond DMs, and how they are used to signal subject matter and presentational relations. For this purpose, we undertake a large-scale signaling annotation project, in which we first select a corpus already annotated for coherence relations, then examine the relations in the corpus, and finally add information on how those relations are signaled, including a variety of possible signals.

4. SIGNALS FOR RELIABLE ANNOTATION. The most important aspect of the annotation was to select and classify the types of signals to annotate. We built our taxonomy of signals based on the different classes of relational markers that we identified in our preliminary corpus work (Das and Taboada 2013, Taboada and Das 2013), or that have been mentioned in previous studies on the signaling in discourse (Bateman et al. 2001, Blakemore 1987, 1992, 2002, Corston-Oliver 1998, Dale 1991b, a, Fraser 1990, 1999, 2006, 2009, Halliday and Hasan 1976b, Knott 1996, Lapata and Lascarides 2004, Le Thanh 2007, Lin, Kan, and Ng 2009, Louis et al. 2010, Marcu 1999, 2000, Maziero et al. 2011, Pitler, Louis, and Nenkova 2009, Polanyi et al. 2004, Prasad, Joshi, and Webber 2010, Sanders, Spooren, and Noordman 1992b, 1993, Schiffrin 1987, 2001, Scott and de Souza 1990, Sporleder and Lascarides 2005, Theijssen 2007).

The taxonomy of signals is organized hierarchically in three levels: *signal class*, *signal type*, and *specific signal*. The top level, *signal class*, has tags representing three major classes of signals: *single*, *combined*, and *unsure*. For each class, a second level of types is defined; for example, the class *single* is divided into nine types (*DM, reference, lexical, semantic, morphological, syntactic, graphical, genre*, and *numerical* features). Finally, the third level in the hierarchy refers to the specific signals; for example, *reference type* has four specific signals: *personal, demonstrative, comparative* and *propositional reference*. The taxonomy of signals is illustrated in **Figure 2**. Note that subcategories are only illustrative, not exhaustive. More detail on the taxonomy can be found in Taboada and Das (2013)[1].

In addition, we find that many relations are indicated by combined signals. Combined signals are made of two or more single signals which work together to indicate a particular relation. We have identified 10 broad types of combined signals[2]: (i) *entity* + *positional*, (ii) *entity* + *syntactic* + *lexical*, (iii) *entity* + *syntactic*, (iv) *graphical* + *syntactic*, (v) *lexical* + *positional*, (vi) *lexical*+ *syntactic* + *positional*, (vii) *lexical* + *syntactic*, (viii) *syntactic* + *lexical*, (ix) *syntactic* + *positional*, and (x) *semantic* + *syntactic*.

[1] http://www.sfu.ca/~mtaboada/docs/Taboada_Das_Dialogue_and_Discourse_2013.pdf#!
[2] For more detail on combined signals, see Taboada and Das (2013).

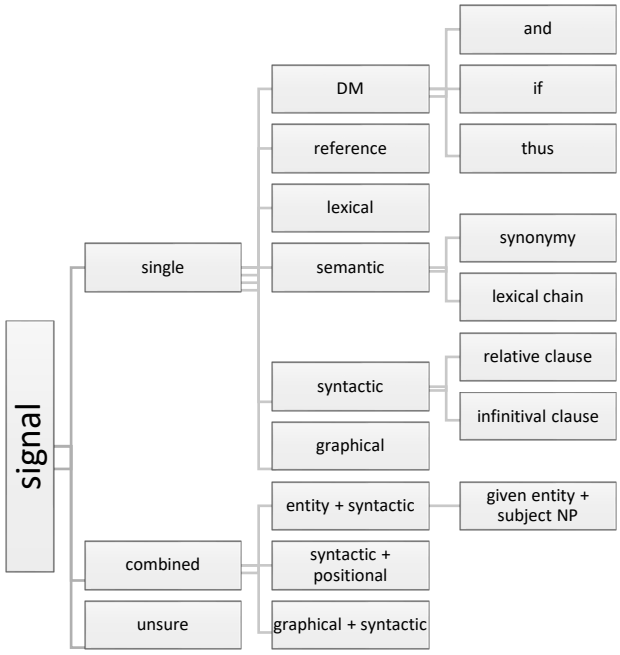

Figure 2. *Hierarchical taxonomy of signals*

5. ANNOTATION PROCESS. For our corpus, we have selected the RST Discourse Treebank or RST-DT (Carlson, Marcu, and Okurowski 2002) which comprises 385 Wall Street Journal articles (financial reports, general interest stories, editorials, etc.) annotated for coherence relations. We chose to use the RST-DT because it is the only available large-scale corpus of RST relations for written texts. The annotated texts in the RST-DT are stored as LISP files which can be opened with RSTTool (O'Donnell 1997) for visual representation.

In our preliminary corpus study, we annotated 1,306 relations in 40 articles which constitute approximately ten percent of the 385 articles in the RST-DT. The annotation process involves examining each relation in the corpus and, assuming the existing relation annotation is correct, searching for cues that indicate that such a relation is present. In some cases, more than one cue may be present. When confronted with a new instance of a particular type of relation, we consult our taxonomy of signals, and find appropriate signal(s) that could best function as the indicator(s) for that relation instance. If our search led us to assigning an appropriate signal (or more than one appropriate signal) to that relation, we declared success in identifying the signal(s) for that relation. If our search does not match any of the signals in the taxonomy, then we examine the context (comprising the discourse components) to discover any potential new signals. If a new signal is identified, we include it in the appropriate category in our existing taxonomy. In this way, we proceed through identifying the signals of the relations in the corpus, and, at the same time, keep on updating our taxonomy with new signaling information, if necessary. We found that after approximately 20 files, or 650 relations, we added very few new signals to the taxonomy.

In the coding task, we provided annotations for signals of coherence relations, or in other words, we added signaling information to the existing relations from the RST-DT. For this purpose, we extracted the signals identified, and documented them along with the relevant information about the relation in question, the document number (the file to which the relation belongs), the status of the spans (nucleus or satellite), and the span numbers (the location of the

spans in the text). We annotated the signaling information in a separate Excel file, since RSTTool does not allow for multiple levels of annotation. For more information about the annotation process, see Das and Taboada (2013)[3] and Taboada and Das (2013).

6. RESULTS. Among the 1,306 relations examined, the distribution of signaled relations (indicated either by DMs or by some other signals) and unsignaled relations (not indicated by any signal) is provided in **Table 1**.

Relation	Tokens	Percentage
Signaled relation	1,129	86.45%
Unsignaled relation	177	13.55%
Total	1,306	
Relations indicated by a DM	251	22.23%
Relations indicated by other signals	878	77.77%
Total	1,129	

Table 1. *Distribution of signaled and unsignaled relations*

The results show that 1,129 relations (86.45%) out of all the 1,306 relations examined are signaled, either by a DM or with the help of some other signaling device. On the other hand, no significant signals are found for the remaining 177 relations (13.55%).

Among the 1,129 signaled relations, we found that DMs are used to signal 251 relations (22.23% of the signaled relations), while 878 relations (77.77% of the signaled relations) are indicated with the help of some other signals.

For the 251 instances of relations signaled by a DM, we have found 58 different DMs. Examples of some of these DMs include *after, although, and, as, as a result, because, before, despite, for example, however, if, in addition, moreover, or, since, so, thus, unless, when* and *yet*. For a full list of these extracted markers, see Taboada and Das (2013).

For the 878 signaled relations without DMs, we have found that wide varieties of signals are used to indicate them. These signals include all the eight types of signals other than DMs mentioned in Section 3, along with numerous specific signals belonging to each type. In our corpus analysis, 81.67% of the signaled relations (922 out of 1,129 signaled relations) are exclusively indicated by a single signal (including DMs), whereas 5.67% of the signaled relations (64 out of 1,129 signaled relations) are indicated by a combined signal. In addition, the distribution also shows that 12.49% of the signaled relations (141 out of 1,129 signaled relations) contain multiple signals[4].

In terms of relation types, we divided the relations in the RST-DT into three groups: subject matter, presentational, and undetermined relations, as shown in **Table 2**.

[3] http://homes.chass.utoronto.ca/~cla-acl/actes2013/Das_and_Taboada-2013.pdf

[4] Multiple signals refer to two or more types of signals (single or combined) which are separately used to indicate a particular relation instance. For more detail, see Taboada and Das (2013).

Subject matter	Presentational	Undetermined
Elaboration, Circumstance, Solutionhood, Cause, Result, Purpose, Condition, Otherwise, Interpretation, Evaluation, Restatement, Summary, Sequence, Contrast, Consequence, Comparison, Preference, Analogy, Proportion, Contingency, Hypothetical, Manner, Means, Topic-Comment, Temporal, Topic-Shift, Textual Organization	Motivation, Antithesis, Background, Enablement, Evidence, Justify, Concession, Attribution, Conclusion, Comment, Explanation-argumentative, Reason	Example, Definition, List, Disjunction, Same-Unit

***Table 2**. Relation classification on subject matter/presentational/undetermined basis*

For subject matter and presentational relations, we mainly followed Mann and Thompson's (1988) original classification of RST relations. For the new relations, which were not part of the original RST taxonomy but used exclusively for annotating the RST-DT, we examined the definitions of those relations from the RST-DT annotation manual (Carlson and Marcu 2001), and assigned them to what we believed was the most suitable category. However, there were also some new relations whose definitions did not fit adequately to any of the two major relation types (i.e., subject matter or presentational), and hence, we assigned them to a new category called undetermined relations.

We also found (as shown in **Table 3**) that among the 1,306 relations examined there are 762 subject matter and 358 presentational relations, along with the remaining 186 undetermined relations.

Relation type	Tokens	Percentage
Subject matter	762	58.35%
Presentational	358	27.41%
Undetermined	186	14.24%
Total	1,306	

***Table 3.** Distribution of relation types*

The distribution of relation types with respect to signaling is provided in **Table 4**.

Relation type	Tokens	# relations signaled	# relations unsignaled
Subject matter	762	660 (86.61%)	102 (13.39%)
Presentational	358	300 (83.80%)	58 (16.20%)
Undetermined	186	169 (90.86%)	17 (9.14%)
Total	1,306	1,129 (86.45%)	177 (13.55%)

***Table 4**. Distribution of relation types for signaled and unsignaled relations*

As **Table 4** shows, out of the 762 subject matter relations 660 relations (86.61%) are signaled while the remaining 102 (13.39%) relations are not signaled. For presentational relations, out of 358 relations 300 relations (83.80%) are signaled and 58 (16.20%) relations are unsignaled. Finally, the breakdown for the 186 undetermined relations is into 169 relations (86.45%) signaled and 17 relations (9.14%) unsignaled.

Furthermore, analyzing the distribution of signaled relations of each relation type (as shown in

Table 5), we found that among the 660 signaled subject matter relations 135 relations (20.45%) are indicated by a DM while the remaining 525 relations (79.55%) are indicated by signals other than DMs. For presentational relations, the distribution is between 58 (19.33%) relations with DMs and 242 (80.67%) relations with other signals, out of a total of 300 signaled relations.

Relation type	Signal type	Tokens	Percentage
Subject matter	Indicated by DMs	135	20.45%
	Indicated by other signals	525	79.55%
Total		660	
Presentational	Indicated by DMs	58	19.33%
	Indicated by other signals	242	80.67%
Total		300	

Table 5. *Distribution of signaled relation types indicated by a DM and by other signals*

The results (see **Table 6**) also show that the most frequently used signals for subject matter relations include DMs such as *and, if*, and *but*, and other signals such as *entity, semantic*, and *syntactic* features. For presentational relations, DMs such as *but, because*, and *although*, and other signals such as a *syntactic* feature are the most common signals.

Relation type	Signal type	Most common signal
Subject matter	Indicated by DMs	and, if, but
	Indicated by other signals	entity, semantic, syntactic
Presentational	Indicated by DMs	but, because, although
	Indicated by other signals	syntactic

Table 6*. Most common signals for relation types*

Finally, signaling by the most frequently used individual relations from each relation type is provided in **Table 7**.

Relation type	Relation	DM	Other signals
Subject matter	Circumstance	Yes	syntactic, lexical
	Cause	Yes	semantic
	Condition	Yes	lexical, syntactic
	Elaboration	Yes	entity, semantic, syntactic, genre
	Purpose	-	syntactic
Presentational	Attribution	-	syntactic
	Background	Yes	lexical, semantic, morphological
	Antithesis	Yes	semantic
	Evidence	-	lexical
	Reason	Yes	-

Table 7. *Distribution of most frequently signaled individual relations by type*

We found that almost every subject matter relation (except a few, such as *Purpose*) is signaled

by a DM, while most of the presentational relations are signaled by a DM. In terms of other signals, almost every subject matter relation is signaled by other signals, while the same is true for the presentational type. That is, most individual relations are indicated by other signals, whereas *Attribution* and *Evidence* are exclusively signaled by other signals.

7. DISCUSSION. In an effort to investigate the signaling of coherence relations beyond DMs, we found out that the majority of the relations present in a written discourse are signaled: in our corpus analysis, 1,129 (86.45%) relations out of the 1,306 relations examined are signaled. We also observed that coherence relations can well be indicated by signals other than DMs: in our study, out of the 1,129 signaled relations, 878 (77.77%) relations contain a signal other than a DM. Furthermore, the signals of coherence relations in written discourse are diverse in nature and can be broadly classified into major groups, such as *DM, reference, lexical, semantic, syntactic, graphical* and *genre* features. The individual signal groups also contain different specific signals in themselves. For example, the feature *syntactic* includes specific signals such as *relative clause, participial clause* and *parallel syntactic construction*.

We would like to point out that what we have found are *positive* signals, that is, indicators that a relation exists. This does not mean that such signals are used exclusively to indicate that relation (as we have seen 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. In a sense, this means that the signals are compatible with a relation, not necessarily indicators of that relation exclusively.

Examining the signaling of subject matter and presentational relations, we observed that there is not much statistical difference between the two relation types: around 85% of each type of relation (86.61% for subject matter and 83.80% for presentational relations) present in a corpus are indicated by some linguistic signals. Furthermore, among the signaled relations of each type, around 20% of the relations (20.45% for subject matter and 19.33% for presentational relations) are signaled by a DM while around 80% of relations (79.55% for subject matter and 80.67% for presentational relations) are indicated by signals other than DMs.

Qualitatively, we also found that the signaling of two relation types do not show any noticeable difference. For example, both subject matter and presentational relations are signaled by DMs as well as by other signals. Furthermore, although there are a few differences between the two relation types in using certain signals, the majority of signals for both the subject matter and presentational relations overlap, as both relation types are conveyed through the same textual signals, either a specific DM or certain other signal types.

Finally, for the 177 relations for which we could not identify a signal, there are three different reasons why we believe that is the case. First of all, in some cases we found that there were errors in the existing annotation of relations in the RST-DT, and a relation was postulated, whereas we would not have annotated a relation. In those cases, the lack of signaling is perfectly understandable. Secondly, some of the RST-DT relations are not true RST relations. Relations such as *Comment* or *Topic-shift*, in our opinion, belong in the realm of discourse organization, not together with relations among propositions. Again, finding no signals in those cases is not surprising, as such phenomena are not likely to be indicated by the same type of signal as coherence relations proper. Finally, in many cases, one or both of the annotators had a sense that the relation was clear but could not pinpoint the specific signal used. This is the case with tenuous entity relations, or relations that rely on world knowledge.

In sum, our findings show negative results; i.e., there are no significant quantitative or qualitative differences between subject matter and presentational relations in terms of signaling.

However, these null findings point to a number of theoretical possibilities. First, subject matter and presentational relations can differ in their intended effects, but the difference in their intended effects may not actually lead to different kinds of signaling. In other words, signaling of coherence relations may be independent of their intended effects; otherwise, the difference in effects should have been reflected in the use of signals as well. Second, the dichotomy between subject matter and presentational relations may be theoretically invalid, i.e., relations may not differ in their intended effects. Finally, even if the subject matter vs. presentational classification is correct, or even if relations do differ in their intended effects, the assignment of individual relations to each type may not be absolutely correct, and the classification may require further review.

8. CONCLUSION. The purpose of this paper was to examine whether subject matter and presentational relations in RST differ in terms of signaling. Our results showed that quantitatively there is not a significant difference between the two relation types in terms of how often they are signaled and unsignaled in a text, and also as to the proportion of signals that are discourse markers (DMs) versus other linguistic devices. We also found that the qualitative difference between the two types is also minor in terms of the types of signals (both DMs and other signals) used for them. Thus, we conclude that subject matter and presentational relations do not differ quantitatively or qualitatively in their signaling. However, these findings, although negative, point to a number of theoretical possibilities, suggesting a revision of the theoretical validity of such relation classifications and also a re-examination of the accuracy in assigning individual relations to a particular relation type.

The annotation described in this paper is, however, a preliminary pilot study, comprising only 10% of the total corpus. In future work, we will expand to cover the entire corpus, and examine the difference between the signaling of subject matter and presentational relations more extensively. Parallel to this work, we would also like to examine the signaling of different relation types using other relation classifications, such as the three-way distinction between semantic, pragmatic and expansion relations as proposed by Redeker et al. (2012), and to compare them with the existing classification in RST.

REFERENCES

BATEMAN, JOHN, THOMAS KAMPS, JÖRG KLEINZ, AND KLAUS REICHENBERGER. 2001. "Towards constructive text, diagram, and layout generation for information presentation." *Computational Linguistics* no. 27 (3):409-449.

BENWELL, BETHAN. 1999. "THE ORGANISATION OF KNOWLEDGE IN BRITISH UNIVERSITY TUTORIAL DISCOURSE: ISSUES, PEdagogic discourse strategies and disciplinary identity." *Pragmatics* no. 9 (4):535-565.

BLAKEMORE, DIANE. 1987. *Semantic Constraints on Relevance*. Oxford: Blackwell.

BLAKEMORE, DIANE. 1992. *Understanding Utterances: An Introduction to Pragmatics*. Oxford: Blackwel.

BLAKEMORE, DIANE. 2002. *Relevance and Linguistic Meaning : Th e Semantics and Pragmatics of Discourse Markers*. Cambridge: Cambridge University Press.

CARLSON, L., AND D. MARCU. 2001. Discourse Tagging Manual.

CARLSON, LYNN, DANIEL MARCU, AND MARY ELLEN OKUROWSKI. 2002. RST Discourse Treebank, LDC2002T07 [Corpus]. Philadelphia, PA: Linguistic Data Consortium.

CORSTON-OLIVER, SIMON. 1998. Beyond string matching and cue phrases: Improving efficiency and coverage in discourse analysis. Paper read at AAAI 1998 Spring Symposium Series, Intelligent Text Summarization, at Madison, Wisconsin.

DALE, ROBERT. 1991a. Exploring the Role of Punctuation in the Signaling of Discourse Structure. Paper read at Workshop on Text Representation and Domain Modelling: Ideas from Linguistics and AI, at Technical University of Berlin.

DALE, ROBERT. 1991b. The role of punctuation in discourse structure. Paper read at AAAI Fall Symposium on Discourse Structure in Natural Language Understanding and Generation, at Asilomar, CA.

DAS, DEBOPAM, AND MAITE TABOADA. 2013. Explicit and Implicit Coherence Relations: A Corpus Study. Paper read at Canadian Linguistic Association (CLA) Conference, at University of Victoria, Canada.

FRASER, BRUCE. 1990. "An approach to discourse markers." *Journal of Pragmatics* no. 14:383-395.

FRASER, BRUCE. 1999. "What are discourse markers?" *Journal of Pragmatics* no. 31:931 - 953.

FRASER, BRUCE. 2006. "Towards a theory of discourse markers." In *Approaches to Discourse Particles*, edited by K. Fischer, 189 - 204. Elsevier Press.

FRASER, BRUCE. 2009. "An Account of Discourse Markers." *International Review of Pragmatics* no. 1:293–320.

HALLIDAY, MICHAEL A. K. 1985. *An Introduction to Functional Grammar*. London: Arnold.

HALLIDAY, MICHAEL A. K., AND RUQAIYA HASAN. 1976a. *Cohesion in English*. London: Longman.

HALLIDAY, MICHAEL, AND RUQAIYA HASAN. 1976b. *Cohesion in English*. London: Longman.

KNOTT, ALISTAIR. 1996. *A data-driven methodology for motivating a set of coherence relations*. Ph.D. dissertation, University of Edinburgh, Edinburgh, UK.

Lapata, Mirella, and Alex Lascarides. 2004. Inferring sentence-internal temporal relations. Paper read at NAACL-04.

LE THANH, HUONG. 2007. "An approach in automatically generating discourse structure of text." *Journal of Computer Science and Cybernetics* no. 23 (3):212-230.

LIN, ZIHENG, MIN-YEN KAN, AND HWEE TOU NG. 2009. Recognizing implicit discourse relations in the Penn Discourse Treebank. Paper read at 2009 Conference on Empirical Methods in Natural Language Processing, at Singapore.

LOUIS, ANNIE, ARAVIND JOSHI, RASHMI PRASAD, AND ANI NENKOVA. 2010. Using Entity Features to Classify Implicit Discourse Relations. Paper read at SIGDIAL 2010.

MAIER, ELISABETH, AND EDUARD HOVY. 1993. "Organising discourse structure relations using metafunctions." In *New Concepts in Natural Language Generation*, edited by H. Horacek and M. Zock, 69-86. London: Pinter.

MANN, WILLIAM C., AND SANDRA A. THOMPSON. 1988. "Rhetorical Structure Theory: Toward a functional theory of text organization." *Text* no. 8 (3):243-281.

MARCU, DANIEL. 1999. A decision-based approach to rhetorical parsing. Paper read at 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, at College Park, Maryland.

MARCU, DANIEL. 2000. "The rhetorical parsing of unrestricted texts: A surface based approach." *Computational Linguistics* no. 26 (3):395-448.

MARTIN, JAMES R. 1992. *English Text: System and Structure*. Amsterdam and Philadelphia: John Benjamins.

MAZIERO, ERICK GALANI, THIAGO ALEXANDRE SALGUEIRO PARDO, IRIA DA CUNHA, JUAN-MANUEL TORRES-MORENO, AND ERIC SANJUAN. 2011. DiZer 2.0 – An Adaptable On-line Discourse Parser. Paper read at Anais do III Workshop "A RST e os Estudos do Texto", at Cuiaba, MT, Brasil.

O'DONNELL, MICHAEL. *RSTTool* 1997. Available from http://www.wagsoft.com/RSTTool/.

PITLER, EMILY, ANNIE LOUIS, AND ANI NENKOVA. 2009. Automatic sense prediction for implicit discourse relations in text. Paper read at Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, at Singapore.

POLANYI, LIVIA, CHRIS CULY, MARTIN VAN DEN BERG, GIAN LORENZO THIONE, AND DAVID AHN. 2004. A rule based approach to discourse parsing. Paper read at SigDIAL 2004, at Cambridge, MA.

PRASAD, RASHMI, ARAVIND JOSHI, AND BONNIE WEBBER. 2010. Realization of Discourse Relations by Other Means: Alternative Lexicalizations. Paper read at COLING 2010, at Beijing.

REDEKER, G., I. BERZLÁNOVICH, N. VAN DER VLIET, G. BOUMA, AND M. EGG. 2012. Multi-Layer Discourse Annotation of a Dutch Text Corpus. Paper read at Conference on Language Resources and Annotation (LREC), at Istanbul, Turkey.

REDEKER, GISELA. 2000. "Coherence and Structure in Text and Discourse." In *Abduction, Belief, and Context in Dialogue. Studies in Computational Pragmatics*, edited by W. Black and H. Bunt, 233–63. Amsterdam and Philadelphia: John Benjamins.

SANDERS, TED, WILBERT SPOOREN, AND LEO NOORDMAN. 1992a. "Toward a taxonomy of coherence relations." *Discourse Processes* no. 15 (1):1-35.

SANDERS, TED, Wilbert Spooren, AND LEO NOORDMAN. 1992b. "Toward a taxonomy of coherence relations." *Discourse Processes* no. 15:1–35.

SANDERS, TED, WILBERT SPOOREN, AND LEO NOORDMAN. 1993. "Coherence relations in a cognitive theory of discourse representation." *Cognitive Linguistics* no. 4 (2):93–133.

SCHIFFRIN, DEBORAH. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.

SCHIFFRIN, DEBORAH. 2001. "Discourse markers: language, meaning and context." In *The Handbook of Discourse Analysis*, edited by D. Schiffrin, D. Tannen and H.E. Hamilton, 54–75. Malden, MA: Blackwell.

SCOTT, DONIA, AND CLARISSE SIECKENIUS DE SOUZA. 1990. "Getting the message across in RST-based text generation." In *Current Research in Natural Language Generation*, edited by R. Dale, C. Mellish and M. Zock, 47-73. London: Academic Press.

SPORLEDER, CAROLINE, AND ALEX LASCARIDES. 2005. Exploiting linguistic cues to classify rhetorical relations. Paper read at Recent Advances in Natural Language Processing (RANLP-05).

TABOADA, MAITE. 2009. "Implicit and explicit coherence relations." In *Discourse, of Course*, edited by J. Renkema. Amsterdam: John Benjamins.

TABOADA, MAITE, AND DEBOPAM DAS. 2013. "Annotation upon annotation: Adding signaling information to a corpus of discourse relations." *Dialogue and Discourse* no. 4 (2):249-281.

TABOADA, MAITE, AND WILLIAM C. MANN. 2006. "Rhetorical Structure Theory: Looking Back and Moving Ahead." *Discourse Studies* no. 8 (3):423-459.

THEIJSSEN, DAPHNE. 2007. *Features for automatic discourse analysis of paragraphs*. MA, Radboud University Nijmegen, The Netherlands.

TRAXLER, MATTHEW J., ANTHONY J. SANFORD, JOY P. AKED, AND LINDA M. MOXEY. 1997. "Processing causal and diagnostic statements in discourse." *Journal of Experimental Psychology: Learning, Memory and Cognition* no. 23 (1):88-101.

VAN DIJK, TEUN. 1977. *Text and Context: Explorations in the Semantics and Pragmatics of Discourse*. London: Longman.

VAN DIJK, TEUN. 1979. "Pragmatic connectives." *Journal of Pragmatics* no. 3:447 - 456.

This article was first published at lacus.weebly.com