Dialogue and Discourse 4(2) (2013) 249-281 doi: 10.5087/dad.2013.211

Annotation upon Annotation: Adding Signalling Information to a Corpus of Discourse Relations

Maite Taboada MTABOADA@SFU.CA

Debopam Das DDAS@SFU.CA

Department of Linguistics
Simon Fraser University
8888 University Dr.
Burnaby, B.C. V5A 1S6
Canada

Editors: Stefanie Dipper, Heike Zinsmeister, Bonnie Webber

Abstract

We present an annotation effort that involves adding a new layer of annotation to an existing corpus. We are interested in how rhetorical relations are signalled in discourse, and thus begin with a corpus already annotated for rhetorical relations, to which we add signalling information. We show that a very large number of relations carry signals that can help identify them as such. The detailed, extensive analysis of signals in the corpus can aid research in the automatic parsing of discourse relations.
Keywords: Discourse relations, corpus annotation, discourse markers

1 Introduction

One of the most frequent tasks that corpus and computational linguists perform is to re-use and re-annotate existing resources. Although many valuable annotated corpora exist, they often do not contain all the information and detail that is necessary in every research project. Thus, researchers are left with the need to add information to a corpus that has already been annotated in some form or another. Starting from scratch may not be optimal, since one can build on existing annotations, unsatisfactory as they may be. This is particularly the case with higher-level annotations, those "beyond semantics", because they tend to rely on annotations at lower levels of discourse, such as semantic role annotations relying on part of speech tags. In addition, many annotation efforts are conceived as layers of different kinds of information, sometimes added by different annotators (see Stede, 2007; Cunningham et al., 2011 for examples of a general philosophy of layered text annotation).

In this paper, we present an annotation effort that involves adding a new layer of annotation to an existing corpus. We are interested in how rhetorical relations are signalled in discourse, and thus begin with a corpus already annotated for rhetorical relations, to which we add signalling information.

The issue of signalling is central in research on discourse relations. Identification and classification of relations often hinge on pinpointing lexical or other cues that indicate a relation

 Submitted: 04/12; Accepted: 02/13; Published online: 08/13

is present, with some approaches to coherence relations relying exclusively on signals (mostly discourse markers) to classify relations (Sanders et al., 1992; Knott & Sanders, 1998). In more applied areas, signals are used to help identify relations in applications such as discourse parsing and summarization (Marcu, 2000a; Schilder, 2002; Hanneforth et al., 2003; Polanyi et al., 2004; Sporleder & Lascarides, 2005; Baldridge et al., 2007; Afantenos et al., 2010).

More generally, the issue of signalling in discourse relations needs to be examined from a processing point of view. If we assume that coherence relations are cognitive entities, then we need to find how hearers and readers are able to identify them on the basis of linguistic cues. Successful communication must be based on a relatively unambiguous interpretation of relations, for which clear signals are necessary. Most psycholinguistic research on this matter to date has focused on one particular type of signal, the presence of discourse markers.

In order to understand how relations are processed, and in order to extract them automatically, we need to move beyond signalling by discourse markers, as those seem to be present in only a small fraction of the relations found in corpora (Taboada, 2006, 2009). We believe that the first step in this endeavour is to annotate discourse with an open mind to other types of signalling. The only other available resource that contains signalling information, the Penn Discourse Treebank (Prasad et al., 2008), contains mostly discourse markers[1] as signals. Although the annotation is very detailed and useful, it does not include all of the types of signals that we believe are indicative of rhetorical relations.

Thus, in this paper, we begin with a corpus already annotated for coherence relations, to which we are adding information on how the relations are signalled, including a variety of possible signals. We begin the paper by briefly discussing coherence relations and their signalling. Then we propose a classification of signalling devices, which we use to annotate a corpus. We discuss the corpus annotation, issues with reliability, and the particular types of problems that are associated with annotating discourse phenomena. The corpus annotation reveals a broad spectrum of signalling devices. The paper concludes with some lessons learned from the annotation, and the applications that the corpus will have.

2 Coherence relations

There are many theories of discourse, rhetorical, or coherence relations, but we believe they all refer to fundamentally a similar phenomenon: relations among propositions, which are the building blocks of discourse, and help explain coherence. Although we have worked within Rhetorical Structure Theory (Mann & Thompson, 1988), and will use some of its constructs here, the discussion that follows likely applies to any view of coherence relations.

In RST, relations are defined through different fields, the most important of which is the Effect, the intention of the writer (or speaker) in presenting their discourse. Relation inventories are open, and the most common ones include names such as Cause, Concession, Condition, Elaboration, Result or 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 are then built out of basic clausal units that enter into rhetorical relations with each other, in a recursive manner. Mann and Thompson proposed that most texts can be analyzed in

[1] In the Penn Discourse Treebank, relations are also annotated as being signalled by indicative phrases. These relations are known as AltLex (Alternative Lexicalization) relations (Prasad et al., 2010).

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 present an RST analysis from the RST Discourse Treebank (Carlson et al., 2002), the corpus that we have chosen to annotate. In it, we can see the text divided into units, or spans, and how rhetorical relations hold across spans. In this case, all the relations are nucleus-satellite, with relations embedded throughout the example. The analysis itself may be questioned in terms of standard RST practice. For instance, unit 4 should probably not be considered a span, and instead included as a unit with the noun that it modifies (*amount*). We are, however, working with an existing annotation, and will use the relations in the corpus as they are.

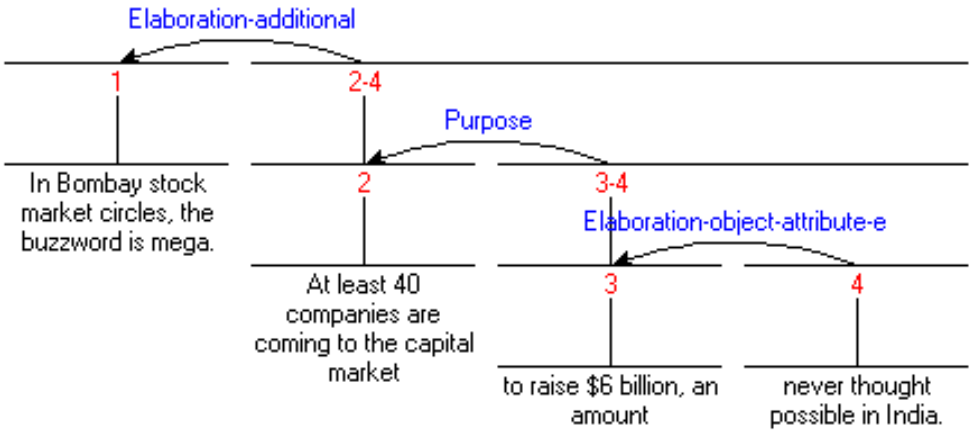

Figure 1. Sample RST analysis from the RST Discourse Treebank

There has been a long and lively debate about how coherence relations, interpreted as rhetorical relations in RST or in other theories (e.g., Polanyi & Scha, 1983; Sanders et al., 1993; Asher & Lascarides, 2003), are recognized and interpreted, that is, their cognitive status: Are relations present in the minds of speakers and hearers[2] or are they analysis constructs? The former postulates that coherence relations are part of the process of constructing a coherent text representation. In Rhetorical Structure Theory (Mann & Thompson, 1988), the relations are postulated as being recognizable to an analyst, and in general to a reader. The process is one of uncovering the author's intention in presenting pieces of text in a particular order and combination. In carrying out an RST analysis of a text, "the analyst effectively provides plausible reasons for why the writer might have included each part of the entire text" (Mann & Thompson, 1988: 246). But further cognitive claims have not been strong within RST.

Support for the cognitive status of coherence relations comes from experimental work on the effect of particular types of relation on text comprehension. Sanders and colleagues have best

[2] We will use speakers/hearers and writers/readers interchangeably. It is arguably the case that most of what can be said about coherence relations applies equally to spoken and written discourse. Indeed, if we postulate psychological validity, both forms of discourse must be accounted for.

articulated this view. In Knott and Sanders (1998), they argue that text processing consists of building a representation of the information contained in the text. Part of the process of building involves integrating individual propositions in the text into a whole. Coherence relations model the ways in which propositions are integrated. The evidence presented comes from two different sources. First of all, studies have shown differences in processing different types of relations, mostly causal versus non-causal (e.g., Trabasso & Sperry, 1985; and references in Knott & Dale, 1994). Secondly, the presence of connectives indicating coherence relations tends to facilitate text comprehension. If coherence relations were not cognitive entities, then there should not be any effect in indicating their presence. The conclusion is, then, that processing coherence relations is part of understanding text. The evidence on the production side is not as abundant, however.

This line of research has explored the identification and classification of coherence relations through discourse markers (or connectives). The problem with such an approach is that it does not address the issue of unsignalled relations. It is clear to most researchers that one can postulate relations (and presumably, readers understand them) even when they are not signalled. If all relations are of the same type, that is, if all relations are cognitive entities, then signalling through discourse markers only facilitates their comprehension. Lack of signalling does not mean that no relation is present.

In the following section we further discuss the signalling problem, and show that signalling has been understudied, focusing mostly on discourse markers.

3 The signalling of discourse relations

In this paper, by *signalling* we mean the cues that indicate that a coherence relation is present, such as the conjunction *because* as a clue that a causal relation is being presented. We use the term *signalling* rather than *marking* because the latter has been associated with discourse markers, one of many possible signalling devices.

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 unsignalled. Whereas it is true that many coherence relations (under whatever definition) are not signalled by a discourse marker, that is, they are *implicit*, it is also often the case that other markers have been understudied (Taboada & Mann, 2006; Taboada, 2009). Our goal in this paper is to push that line of research further. We explore how many, and what types of cues can be found if we study signalling beyond discourse markers. A secondary goal aims at discovering whether unsignalled or implicit relations can be said to exist at all. If we postulate psychological validity for coherence relations, that is, if we assume that coherence relations are present in discourse and that they are recognized by speakers, then there must be signals through which speakers identify relations when parsing discourse.

If, as Spooren (1997) suggests, underspecified or unsignalled relations obey the Cooperative Principle (Grice, 1975) and the Quantity maxim ("say no more than necessary")[3], then unsignalled relations are such because no signal is necessary. Psycholinguistic experiments have shown that certain relations are processed faster when a connective is present. Haberlandt (1982), for instance, found that causal and concessive connectives between two sentences resulted in a faster processing of the second sentence. This was compared to pairs of sentences with no

[3] Spooren actually makes reference to Horn's (1984) take on the Cooperative Principle, which can be summarized as "say no more than necessary".

connective between them. The conclusion was that the lack of connective necessitated inference, which resulted in longer processing times. Sanders et al. (2007) showed that explicitly marked relations led to better performance in text comprehension questions, both in laboratory and realistic situations.

The effects of signalling on recall and some aspects of comprehension have been more mixed. Meyer et al. (1980) found no positive effect on recalling content[4]. They did, however, find that subjects recalled the structure of the original text more faithfully when it was signalled. Millis and Just (1994) saw an increase in processing time but more accurate answers to comprehension questions when a connective was present. Degand and Sanders (2002) report better answers on comprehension questions if the texts include a relational marker. Sanders and Noordman (2000) found that connectives had a positive effect on processing, but no noticeable effect on recall. Sanders and Noordman's conclusion about the recall effect is that the effect of the marker decreases over time, just as the surface representation of the text is lost, but the semantic content is preserved longer. Degand and Sanders (2002) also caution that the mixed results may reflect mixed methodology, where there was no control for different types of connectives, coherence of the texts, evaluation methodology (free recall versus comprehension questions), or reader background.

Other studies have shown that the effect of signalling is different for different types of readers. Meyer et al. (1980) discovered that explicit connectives helped only underachieving students, those readers that need signalling to identify the top-level structure of a text. Britton et al. (1982) also found faster reaction times in a secondary task, but no effect on recall due to signalling, in two types of subjects, with average or low verbal ability (measured in terms of the Scholastic Aptitude Test).

Although it is not the focus of this paper, it is also worth mentioning that some of the experimental work has studied the role of different types of relations. It has consistently been shown that causal relations are processed faster and often lead to better recall than other types of relations (e.g., Keenan et al., 1984; Trabasso & Sperry, 1985; Myers et al., 1987). Other research has shown differences among different relations, such as problem-solution and list (Sanders & Noordman, 2000). It seems clear that coherence relations are different in nature among them. This probably means that their signalling will also be different, not only in terms of whether signalling is present or not, but in terms of which types of signal produce which comprehension effects.

The task of a writer or speaker, then, is one of determining how much signalling is enough. A writer may decide that no connective is necessary because other cues that suffice to identify the relation are present, thus obeying the Quantity maxim or, according to Spooren (1997), the R-principle ("say no more than necessary"). In a study of young (6-7 year old) and older (11-12 year old) children, Spooren found that a number of relations were unsignalled (close to 20%) and, more importantly, that a very large number (between 65 and 75%) were underspecified, that is, they were signalled by general connectives, such as *and*. There was a significant difference between the age groups, with younger children leaving fewer relations implicit, but using more underspecified relations.

[4] Meyer et al.'s (1980) signalling included explicit statements of the structure of the text and connectives. As noted later on in this section, the results were different for different types of students (poor vs. good readers).

The fact that some studies have found no significant effects of signalling on recall may indicate that readers (and hearers) are able to process text and assign relations successfully, even if the effort requires more time with unsignalled relations, or relations that are more weakly signalled (for instance, signalled by an open-class lexical item instead of a connective).

Most of the work reviewed thus far dealt with connectives/discourse markers. The problem is that there are many other types of signals that may facilitate the comprehension process, and those have clearly been understudied.

In previous work (Taboada, 2004, 2006, 2009) we have reported on different types of signals that can be used to identify a relation. Here we summarize that work, and in the next section we provide a more detailed list of the signals used in this study.

Discourse markers are, of course, the most studied signals. In some cases, the taxonomy of discourse markers has been reduced to single-word conjunctions. We have found many multi-word expressions that function as discourse markers, even though some of them may not be conjunctions from a syntactic point of view, such as *in the event that* in the following example, from the RST web site (Mann & Taboada, 2010), which signals a condition relation between spans (1b) and (1c). This is a prepositional phrase that takes a clausal complement.

(1) [Copyright notice]

a. This notice must not be removed from the software,

b. and in the event that the software is divided,

c. it should be attached to every part. [RST Web Site]

One aspect that we have discussed elsewhere is the use of mood and modality to signal relations. For example, a question (as expressed by an interrogative mood) is a potential signal for a Solutionhood relation. Verb finiteness is sometimes the only indicator of a relation, as shown in Example (2), from the RST Discourse Treebank (Carlson et al., 2002). The Circumstance relationship between spans 1 and 2-5 is signalled by the non-finite form of the verb *insisting*.

(2) [1] Insisting that they are protected by the Voting Rights Act, [2] a group of whites brought a federal suit in 1987 [3] to demand that the city abandon at-large voting for the nine-member City Council [4] and create nine electoral districts, [5] including four safe white districts. [RST Discourse Treebank]

Lexical items may also be used to indicate a relation, such as the verb *cause* in a causal relation, or *concede*, as in Example (3), which in this case marks a Concession relation.

(3) [S] Some entrepreneurs say the red tape they most love to hate is red tape they would also hate to lose. [N] They concede that much of the government meddling that torments them is essential to the public good, and even to their own businesses. [RST Discourse Treebank]

In Example (4) there is an Evaluation relation between segments 1 and 2. The author characterizes the narrator of the novel "The Wedding" as a character removed from the main protagonist, Noah, and therefore making the connection between narrator and protagonist quite indirect. The main indicator of this Evaluation relation is the semantic content of the word *indirect*, an adjective conveying subjective content.

(4) [1] The first-person narrator of "The Wedding" is the son-in-law (Wilson) of Noah's daughter Jane. [2] ?????? Talk about indirect. [SFU Review Corpus]

We embrace a view of coherence in discourse whereby coherence relations (also known as relational coherence) and reference and lexical relations (also known as cohesion, or entity-based coherence) are part of what renders a text coherent. This is the view in Poesio et al. (2004), and the principle behind Veins Theory (Cristea et al., 1998). In general, coherence established by lexical means, as part of a more general entity coherence, or cohesion, is a very important aspect of signalling. Karamanis (2007), for instance, assumes that, in the absence of a marker, entity coherence (links among entities in the discourse) signals the relation. As we will see in later sections, cohesion of all types (reference, lexical, etc.) seems to be a strong indicator of coherence. In fact, in the Halliday and Hasan (1976) view of cohesion, cohesion and coherence relations are part of the same system, with coherence relations represented by conjunctive links. Thus, it is not surprising that we see signalling by lexical and other cohesive devices as an extension of signalling by conjunctions and discourse markers.

Other cases are more difficult and subjective to interpret. Example (5) contains two Elaborations embedded within each other. In the first relation, the satellite starts with "Recently, the boards…" and continues to the end of the paragraph, which is longer than displayed in the example here. The only possible signal that an Elaboration relation is present is the adverb *also* before the main verb *voted* in this satellite. The second Elaboration relation has that "Recently, the boards…" sentence plus the next sentence as nucleus. The satellite starts with "The transaction…" and continues for a while. This second satellite has no adverb, punctuation mark, or any other device that indicates an elaboration on what has gone before. Knowledge of the newspaper genre leads us to think that an article, unless other cues are present, proceeds in a series of elaborations.

(5) [N1] American Pioneer Inc. said it agreed in principle to sell its American Pioneer Life Insurance Co. Subsidiary to Harcourt Brace Jovanovich Inc.'s HBJ Insurance Cos. for $27 million. American Pioneer, parent of American Pioneer Savings Bank, said the sale will add capital and reduce the level of investments in subsidiaries for the thrift holding company. [S1] [N2] Recently, the boards of both the parent company and the thrift also voted to suspend dividends on preferred shares of both companies and convert all preferred into common shares. The company said the move was necessary to meet capital requirements. [S2] The transaction is subject to execution of a definitive purchase agreement and approval by various regulatory agencies, including the insurance departments of the states of Florida and Indiana, the company said. […] [RST Discourse Treebank]

Finally, there is the question of punctuation and layout in written texts, including the problem of how these devices correlate with rhetorical relations. There is some work in this area, going back to Hovy and Arens (1991) and Dale (1991), and including research by Bateman (Bateman et al., 2001), which in general shows a good correlation between some forms of layout and rhetorical relations.

It should be fairly clear by now that multiple signals for relations are possible, and that some of them are straightforward to annotate, such as discourse markers, especially conjunctions, whereas some other signals require long-distance dependencies and involve a certain amount of

subjectivity. In our work, we have strived to compile a list of signals that we felt we could annotate reliably. The next section discusses these.

4 Signals for reliable annotation

The most important aspect of the annotation was to select and classify the types of cues to annotate. Discourse markers have been extensively studied, and are relatively easy to identify. Beyond discourse markers, we found other classes of cues that have been mentioned in previous studies, or that we identified in our preliminary corpus work. The classification has a top-level breakdown into discourse markers, morphological, syntactic, semantic, lexical, genre and graphical features, plus heuristics specific to each relation. We started our annotation, as we explain in Section 5, by consulting previous studies for indication of what signalling devices have been found in corpora (Halliday & Hasan, 1976; Blakemore, 1987; Schiffrin, 1987; Fraser, 1990; Scott & de Souza, 1990; Dale, 1991; Blakemore, 1992; Sanders et al., 1992, 1993; Knott & Dale, 1994; Knott, 1996; Corston-Oliver, 1998a; Fraser, 1999; Marcu, 1999, 2000b; Bateman et al., 2001; Schiffrin, 2001; Blakemore, 2002; Lapata & Lascarides, 2004; Polanyi et al., 2004; Sporleder & Lascarides, 2005; Fraser, 2006; Huong, 2007; Prasad et al., 2007; Pardo & Nunes, 2008; Sporleder & Lascarides, 2008; Teijssen et al., 2008; Fraser, 2009; Lin et al., 2009; Pitler et al., 2009; Louis et al., 2010; Prasad et al., 2010). Then, as we annotated more and more relations, we added to our classification. The top-level classification of signals is provided in Figure 2. We briefly discuss this classification below. A full account, with examples of each type, is available as supplementary material to this article[5]. Please note that the subcategories in Figure 2 are illustrative, not exhaustive.

Discourse markers are by far the most studied type of signalling (see references in Taboada and Mann, 2006a, 2006b). Markers are specific to each relation, such as *if* for Condition or *although* for Concession. There is, however, no one-to-one correspondence between markers and relations, and many markers are ambiguous (*and* can indicate a number of discourse relations, in addition to its function as linking device within clauses and phrases).

In our annotation, we mainly followed Fraser's (1999, 2006, 2009) definition of discourse markers, that discourse markers constitute a functional class of linguistic elements drawn from different syntactic classes, such as conjunctions, adverbs and prepositional phrases. They connect discourse segments, and signal a relation between them. In addition, we also followed a number of conditions for considering an expression to be a discourse marker. The conditions are enumerated below.

1. The scope of the function of a discourse marker is a single discourse sequence comprising adjacent text spans in a relation.
2. Discourse markers can be present at the beginning or end of the sentence (or segment), or within the sentence (or segment).
3. Discourse markers signal relations that hold between two adjacent text segments.
4. A discourse marker does not create the relation between text segments. It only guides the interpretation of the relation.

[5] http://www.sfu.ca/~mtaboada/docs/Taboada_Das_Dialogue_and_Discourse_2013_supplementary_material.pdf

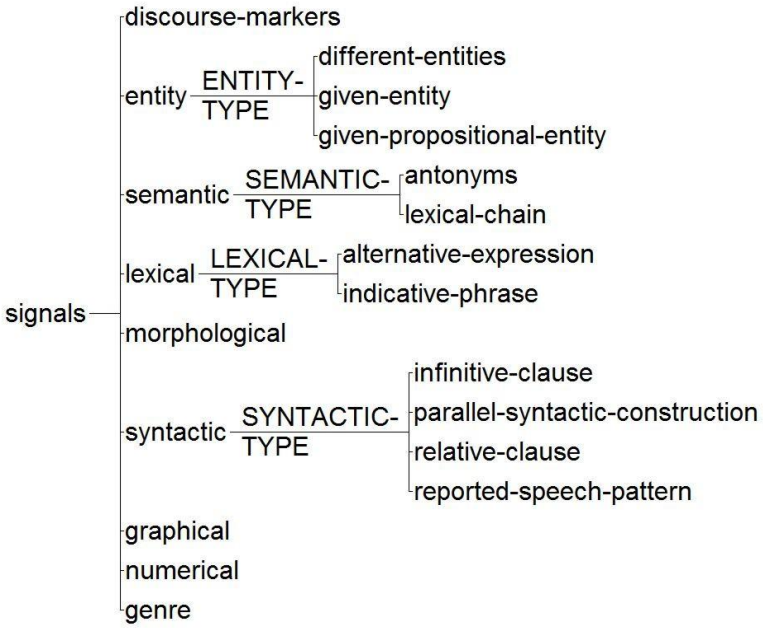

Figure 2. Top-level classification of signals

Entity features include links where entities, similar or dissimilar, help interpret the relation. For example, in (6), which contains a multinuclear List relation with three nuclei, the three distinct entities indicate that a roster of companies is being listed[6].

(6) [Earlier this year, **Tata Iron & Steel Co.**'s offer of $355 million of convertible debentures was oversubscribed.]N

[Essar Gujarat Ltd., a marine construction company, had similar success with a slightly smaller issue.]N

[Larsen & Toubro started accepting applications for its giant issue earlier this month;]N

Many of the **semantic relations** in Halliday and Hasan (1976) can be used to identify relations, such as antonyms as signals of Contrast, or hypernyms as indicators of the satellite(s) in an Elaboration relation. We define these as semantic relations because a semantic link between two or more entities is established, as opposed to the lexical features mentioned below, where a single word or phrase is used, with no connection to other words in the text. The category that includes entities is related to this one. Under semantic relations, however, we include relations that are easily labelled in terms of synonym, antonym, hyponym, etc.

[6] There are many other signals in this example, among them the word *similar* in the second sentence, and the temporal descriptions (*earlier this year; earlier this month*). The example is being used here to illustrate entity features. This will be the case with other examples used to illustrate signals: One particular signal will be highlighted, but other signals may be present in the example.

Lexical features include the use of indicative words and phrases, such as individual words that indicate a relation, for example, the verbs *concede* and *cause* for Concession and Cause respectively. Indicative phrases are some of the more difficult signals to define a priori, but, when they appear, they are unequivocal in their nature as signals. Examples from the current round of analysis include *last year* as an indication of Background, and *at the same time* for Temporal-same-time.

Among **morphological features**, tense is the most prominent one, helping indicate temporal relations (Circumstance in RST terms), or more general Circumstances, as is the case with some instances of non-finite verbs (Taboada, 2006).

At the **syntactic level** there are a host of constructions that help identify a relation. From word order, such as subject-verb inversion for Condition (*Had he known…*) to sentence mood, such as the use of interrogatives to signal Solutionhood.

Graphical and other punctuation features, such as lists and headings, and other forms of layout are sometimes indicators of a relation.

Numerical elements are present in List relations, but also in more subtle ways, when an Elaboration consists of providing a general word (in this case, a number) and then listing the contents of that word. An example is (7), where the nucleus contains the numeral *five*, and the satellite a listing of five names.

(7) [This maker of electronic devices said it replaced all **five** incumbent directors at a special meeting …]N

[Elected as directors were **Mr. Hollander, Frederick Ezekiel, Frederick Ross, Arthur B. Crozier and Rose Pothier**.]S

Genre helps guide the interpretation of relations when the style of the genre is well known to the reader. In the newspaper genre that all the texts in the corpus belong to, it is common to start the text with general information, and to continue with further details. This results in Elaboration relations, with the nucleus being the first sentence or paragraph, and the rest of the article acting as a satellite that expands on the beginning of the text. Other aspects that are specific to newspaper writing are the ways in which the Attribution relation is signalled. These are classified under graphical (quotes and dashes) or syntactic features (verbs of diction such as *say* or *claim*), but it is the fact that the genre is journalistic discourse that provides the interpretation for those signals.

Two other types of signals are not included in our classification above, because they are either too general or too specific. The general class of **discourse features** is less loosely defined, and can include position in the text (Summary tends to appear at the end), Given-New status (in Elaboration and Contrast relations), or genre characteristics (Evaluation relations more common in opinion texts). In some instances, this class overlaps with genre.

Our last category includes **heuristics**, that is, features that are specific to relations. One example is the use of evaluative words (*satisfactory, adequate, success*) in a satellite, which indicate that it is modifying a nucleus in an Evaluation relation.

These broad categories describe single signals, that is, one specific item that indicates the relation. The types of signals described above contain many specific signals in themselves. For

example, the syntactic type includes specific signals such as infinitival clause, participial clause, parallel syntactic construction and reported speech pattern.

In addition, we find that many relations are indicated by combined signals. Combined signals are made of two or more single signals which work in combination with each other to indicate a particular relation. For instance, the List relation between span 2-4 and span 5 in Example (8) in the next section (see Table 1) is indicated by the combined signal Entity + syntactic (more specifically, Given entity + subject NP), along with the single signal Lexical chain (of Semantic Type). We have identified 10 broad types of combined signals: (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. Lists of combined signals can be found in the supplementary material available online (see Footnote 5).

Some relations are also indicated by multiple signals. The difference between combined signals and multiple signals is one of independence of operability. In a combined signal, there are usually two signals, one of which is an independent signal, while the other one is dependent on the first signal. For example, in *given entity + subject NP*, which is a combined signal, *given entity* is the independent signal because it directly (and independently) refers back to the entity introduced in the first span. In contrast, *subject NP* is the dependent signal because it is used to specify additional attributes of the first signal. In this particular case, the syntactic role of the given entity (i.e., a subject NP) in the second span is specified by the use of the second signal *subject NP*. Multiple signals, on the other hand, function independently and separately of each other, but they all contribute to signaling the relation. For example, in an elaboration relation with multiple signals, involving a genre feature (e.g., textual organization) and a lexical feature (e.g., indicative word), the signals do not have any connection, as they refer to two different features which separately signal the relation.

5 Annotation process

For our corpus, we have selected the RST Discourse Treebank (Carlson et al., 2002), a collection of 385 Wall Street Journal articles annotated for rhetorical relations. We elected to use an existing corpus to expedite our research on signalling, even though the corpus may not be ideal. We believe this will be a more and more frequent situation for researchers in discourse, with so many existing annotated corpora available that can be reused and extended. We discuss some of our technical and theoretical difficulties with the layered annotations.

The annotation process involves examining each relation and, assuming the relation annotation is correct, searching for cues that indicate that such relation is present. In some cases, more than one cue may be present. From a theoretical point of view, some of the difficulties that we are encountering are disagreements with the annotations already present in the corpus, from the segmentation (for our approach to segmentation, see Tofiloski et al., 2009) to the application of relation definitions, also including the particular inventory of relations used to annotate the corpus. From a practical point of view, we need to read lengthy texts and examine both parts of a relation, which are sometimes far apart from each other. Our current setup involves opening the files in RSTTool, a graphical interface to annotate RST relations (O'Donnell, 1997), and annotating information about the signalling in a separate Excel file, as RSTTool does not allow for multiple annotations. The annotation, at this point, includes only information about the type and subtype of signal involved, and an indication of what word(s) convey the signalling. It does

not, however, consist of an integrated annotation on the actual RST Discourse Treebank files. A future goal is to find a way to layer our signalling annotation over the existing RST Discourse Treebank, marking both the type of signal and the words we can identify as signals.

In our preliminary corpus study, we annotated 40 articles which constitute approximately ten percent of the 385 articles in the RST Discourse Treebank[7]. The texts in these articles contain 1,304 rhetorical relations. For the annotation of these relations, we performed a sequence of three main tasks: (i) we examined each and every relation in the RST Discourse Treebank, (ii) we identified the signals involved to indicate those relations, and, finally, (iii) we documented information on how the relations are signalled.

We used the list presented in Figure 2 above to identify signals. When confronted with a new instance of a particular type of relation, we consulted our list, and tried to find appropriate signal(s) that could best function as the indicator 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 did not match any of the signals in the list, then we examined the context (comprising the spans) to discover any potential new signals. If a new signal was identified, we included it in the appropriate category in our existing list. In this way, we proceed through identifying the signals of the relations in the corpus, and, at the same time, keep on updating our database with new signalling information, if necessary. We found that after approximately 20 files, or 650 relations, we added very few new signals to the list.

In the coding task, we provided annotations for signals of coherence relations, or in other words, we added signalling information to the existing relations from the RST corpus. For this purpose, we extracted the signals identified, and documented them along with relevant information about the relation in question, the document number (to which the relation belongs), the status of the spans (i.e., nucleus or satellite), and the span numbers (i.e., the location of the spans in the text). We annotated the signalling information in a separate Excel file, since RSTTool, as previously mentioned, does not allow multiple levels of annotation.

5.1 An annotation example

We provide the annotation of a short RST file (file no. 650) with signalling information. The file contains the text in Example (8).

(8) Sun Microsystems Inc., a computer maker, announced the effectiveness of its registration statement for $125 million of 6 3/8% convertible subordinated debentures due Oct. 15, 1999.

The company said the debentures are being issued at an issue price of $849 for each $1,000 principal amount and are convertible at any time prior to maturity at a conversion price of $25 a share.

The debentures are available through Goldman, Sachs & Co.

[7] The 385 articles in the RST discourse Treebank are organized into 385 separate files which are divided into two groups: (i) training documents, comprising 347 files, and (ii) test documents, comprising the remaining 38 files. The 40 articles chosen for annotation are taken from the training set.

The graphical representation of the RST analysis of this text using the RST Tool is provided in Figure 3.

The RST analysis shows that the text in Example (8) comprises five spans which are represented in the diagram (in Figure 3) by the numbers, 1, 2, 3, 4, and 5[8], respectively. In the diagram, the arrowhead points from a satellite to a nucleus span. Span 3 (nucleus) and span 4 (nucleus) are in a multinuclear List relation, and together they make the combined span 3-4. Span 2 (satellite) is connected to span 3-4 (nucleus) by an Attribution relation, and together they make the combined span 2-4. A multinuclear List relation holds between spans 2-4 (nucleus) and 5 (nucleus), and together they make the combined span 2-5. Finally, span 2-5 (satellite) is connected to span 1 (nucleus) by an Elaboration (more specifically, Elaboration-addition) relation.

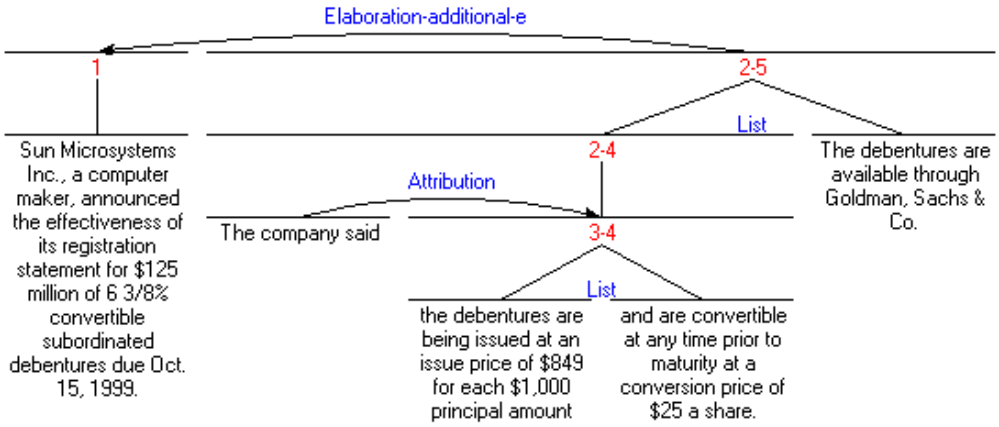

Figure 3. Graphical representation of an RST analysis

We annotated the text in Example (8) with the appropriate signalling information. A detailed description of our annotation for the text is provided in Table 1.

According to our annotation, the Elaboration relation between spans 1 and 2-5 is indicated by three types of signals: (i) Genre; (ii) Entity + syntactic; and (iii) Lexical features. First, the text is part of the newspaper genre (since it is taken from a Wall Street Journal article), and in newspaper texts the content of the first (or the first few) paragraphs is typically elaborated on in the following paragraphs. A reader, being conscious of the fact that he/she is reading a newspaper text, expects the presence of an Elaboration relation between the first paragraph (or the first few paragraphs) and subsequent paragraphs. It is this prior knowledge about the textual organization of the newspaper genre that guides the reader to interpret an Elaboration relation between paragraphs in a news text. In this particular example, the entire first paragraph is the nucleus of the Elaboration relation, with the two following paragraphs being its satellite. Thus, we postulate that the Elaboration relation is conveyed by the genre feature (more specifically by a feature

[8] Spans 2 to 5 do not actually have a label in the corpus. While the labels are inferable, this makes the annotation more complicated with lengthy files.

which we call textual organization). Second, we postulate that a combined signal Entity + syntactic, made of two individual features, is operative in signalling the Elaboration relation (see Section 6 for more information about combined signals). One can notice that the entity *Sun Microsystems Inc.*, mentioned in the nucleus, is elaborated on in the satellite. Syntactically, the entity is also used as the subject NP of the sentence the satellite starts with, representing the topic of the Elaboration relation. Finally, the Elaboration relation is also (perhaps rather loosely) signalled by a lexical feature, lexical overlap. Words such as *debentures* and *convertible* occur in both the nucleus and satellite, indicating the presence of the same topic in both spans, with an elaboration in the second span of some topic introduced in the first span.

The List relation between spans 3 and 4 is conveyed in a straightforward (albeit underspecified) way by the use of the discourse marker *and*.

The Attribution relation between spans 2 and 3-4 is indicated by a syntactic signal, a reported speech pattern in which the reporting clause (span 2) functions as the satellite and the reported clause (span 3-4) functions as the nucleus. The key is the S+V (Subject+Verb) combination with a reported speech verb (*said*).

Source file no.	Nucleus span	Satellite span	Relation name	Marker type identified	Specific marker identified	Explanation – how the relation is signalled
650	1	2-5	Elaboration-additional	Genre	Textual organization	Newspaper: the content of the first paragraph (or the first few paragraphs) is elaborated on in the following paragraphs.
				Entity + syntactic	Given entity + subject NP	*Sun Microsystems Inc.*, mentioned in the nucleus, is the subject of the sentence which the satellite starts with.
				Lexical	Lexical overlap	Words such as *debentures* and *convertible* are in both the spans.
	3/4		List	Discourse marker	*and*	The discourse marker *and* functions as a signal for the List relation.
	3-4	2	Attribution	Syntactic	Reported speech pattern	The reported speech pattern "The company said…" is a signal for the Attribution relation.
	2-4/5		List	Entity + syntactic	Given entity + subject NP	The subject NP of the reported speech in the first span and the subject NP of the sentence in the second span both refer to the same entity: *the debentures*.
				Semantic	Lexical chain	The words *issued* and *available* in the respective spans are semantically related.

Table 1. Annotation of an RST file with relevant signalling information

Finally, the List relation between spans 2-4 and 5 is indicated by two types of signals: (i) Entity + syntactic and (ii) Semantic feature. For the combined feature Entity + syntactic, the specific signal is called Given entity + subject NP, which means that the subject NP of the reported speech within the first span and the subject NP of the sentence in the second span both refer to the same entity (*the debentures*, in this case). For the semantic feature, the specific signal is a lexical chain which means that semantically similar or related words occur in the respective text spans. We notice that words such as *issued* and *available* are semantically related, and they are used in both spans, indicating a List relation holding between them.

After the annotations (with signalling information) are done, we code our annotated data in a separate Excel file. The coded version of our annotation for the text is provided in Table 2.

Source	Nucleus	Satellite	Relation	Marker type	Specific marker
650	1	2-5	Elaboration (-additional)	Genre + (entity + syntactic) + lexical	Textual organization + (given entity + subject NP) + lexical overlap
650	3/4	-	List	Discourse Marker	And
650	3-4	2	Attribution	Syntactic	Reported speech pattern
650	2-4/5	-	List	(Entity + syntactic) + semantic	(Given entity + subject NP) + lexical chain

Table 2. Coding of signalling information for relations in an RST-annotated text

In this way, we completed our annotation task with signalling information for the relations for a total of 40 files in the RST Discourse Treebank.

5.2 Reliability study

As with all annotations, ours carries a certain amount of subjectivity. This is particularly true with discourse annotations and all phenomena beyond semantics, where interpretation of the context and of long-distance features plays a role.

Our list of signals and the annotation procedure were agreed upon after several iterations of the taxonomy and after adding more signals when our initial analysis revealed more than we had originally listed.

To check the validity and reproducibility of our taxonomy, we conducted a reliability study. We selected approximately 10% of the 1,304 relations in the current annotation (see Section 6), coming from two of the texts. One of us had annotated the entire corpus, and the other one annotated those two files, containing 130 relations. We concentrated on whether we agreed on at least one of the signals for the relation. Some relations have multiple signals, and some relations have combined signals. Calculating agreement on those becomes very complex quite quickly, so we stayed with simple signals, and an agreement of at least one signal per relation. Also because of the complexity of the task, we calculated agreement using the top level of the signalling taxonomy, that is, the nine top-level signals from Figure 2. We established whether we agreed on the type of signal, not necessarily on where it was conveyed in the text (e.g., for a lexical chain, we annotated 'semantic', but not what words were involved in the chain).

Agreement was calculated using Cohen's kappa (Siegel & Castellan, 1988), with nominal data, namely, the nine categories in our classification, plus an extra category, "no signal", used to indicate cases where the annotator concluded that there was no identifiable signal. Agreement on

this category is just as important as on the other ones. The kappa value for our study was 0.68, or moderate agreement.

Table 3 presents the disagreements per relation. Of note is the fact that in Elaboration relations, we disagreed in only 17 out of 64 instances (26% of the time), whereas we expected disagreement for that relation to be higher.

In terms of markers, disagreement was higher for genre, where we disagreed in all four cases that it appeared, one annotator identifying genre as a signal, and the other one labelling the instance as 'no signal'. Other markers where disagreement was high were semantic markers (66%, or 20 out of 30 cases) and lexical signals (55%, 5 out of 9 cases). We have, as a consequence, refined our taxonomy of lexical and semantic labels, and believe this will have a positive effect on agreement, to be determined in future agreement studies as we proceed with annotation.

Relation	Agreement	Disagreement
Antithesis	3	-
Attribution	19	1
Background	1	3
Cause-result	-	1
Circumstance	1	1
Condition	2	-
Contrast	3	-
Elaboration	47	17
Example	-	2
Explanation	-	4
Hypothetical	1	-
List	5	-
Manner	-	2
Problem-solution	2	1
Purpose	5	-
Same-unit	6	-
Summary	-	1
Temporal	2	-
Total	97	33

Table 3. Agreement and disagreement per relation

A more general issue as regards reliability studies is whether they are useful at all. In our study, as in most published studies, the level of agreement is considered acceptable, and we do believe that our annotation is reproducible. The larger question is whether providing values for kappa or for similar measures reveals much about the annotation process and its level of difficulty. Reaching such level of agreement after four iterations through the data and after modifying the annotation guidelines is quite different from doing so after a quick explanation of the methodology to a new member of the research group. Spooren and Degand (2010) discuss agreement measures in a similar task, that of coding coherence relations, and conclude that measures beyond kappa are necessary to ensure and measure reliability, such as double coding and discussion of disagreement and agreement cases, and other agreement measures. Those will be part of future reliability tests in our project.

In our case, the reliability study could only be carried out by members of our project, who were familiar with RST, shared similar points of view with regard to what counts as a relation,

and agreed on the list of signals given. Other annotators may disagree with our results, no matter how experienced, or how much time they spend studying our guidelines. We point this out because we feel that too much emphasis is placed on arriving at an acceptable measure of agreement, when an acceptance of the intrinsic difficulty of annotation is what is needed, together with a reasonable explanation of how the annotation was performed.

6 Results

Among the 1,304 relations examined, the distribution of signalled relations (indicated either by discourse markers or by some other signal) and unsignalled relations (not indicated by any signal) is provided in Table 4.

Relation Type	Tokens	Percentage
Signalled relations	1,127	86.43%
Unsignalled relations	177	13.57%
Total	1,304	
Relations indicated by a discourse marker	251	22.27%
Relations indicated by other signals	878	77.91%
Total	1,127	

Table 4. Distribution of signalled and unsignalled relations

The results show that 1,127 relations (86.43%) out of all the 1,304 relations are signalled, either by a discourse marker or with the help of some other signalling device. On the other hand, no significant signals are found for the remaining 177 relations (13.57%).

Among the 1,127 signalled relations, we find that discourse markers are used to signal 251 relations (22.27% of the signalled relations), while 878 relations (77.91% of the signalled relations) are indicated with the help of some other signals.

We need to point out that there are two instances of List relation which are signalled by both a discourse marker and some other signal (which is why the total of 251 plus 878 actually adds up to 1,129). This is because these relations are multinuclear, consisting of three or four nuclei, and we found that while a nucleus is connected to another nucleus by a discourse marker, a third nucleus is related to any of the two former nuclei (in case of a relation with three nuclei), or to a fourth nuclei (in case of a relation with four nuclei) by means of some other signal(s).

For the 251 instances of relations signalled by a discourse marker, we found 58 different discourse markers. Examples of some of these discourse markers 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*. A full list of these extracted markers is available (see Footnote 5).

For the 878 signalled relations without discourse markers, we found that a wide variety of signals are used to indicate them. As mentioned in Section 4, we divide the signals into two broad groups: single and combined signals.

In our corpus analysis, 81.81% of the signalled relations (922 out of 1,127 signalled relations) are exclusively indicated by a single signal (including discourse markers), whereas 5.69% of the signalled relations (64 out of 1,127) are indicated by a combined signal.

We have also noticed that in many cases multiple signals, i.e., two or more types of other signals (single or combined) are separately used to indicate a particular relation instance. For

instance, the Elaboration-additional relation between span 1 and span 2-5 in Example 8 (see Table 1) is indicated by multiple signals: (i) Genre, (ii) Entity + Syntactic, and (iii) Lexical features. The distribution of the signals in our annotation shows that 12.51% of the signalled relations (141 out of 1,127 signalled relations) contain multiple signals. This is an encouraging result for any attempt at automatic identification, as the redundancy in signalling will increase the chances of identification.

The relative distribution of relations with respect to whether they are indicated by a discourse marker, by some other signals, or whether they are unsignalled is provided in Table 5.[9]

No.	Relation group	Relation	# Relations signalled by DMs	# Relations signalled by other markers	# Relations not signalled	Total
1.	Attribution	Attribution	0	228	3	231
		Attribution-negative	0	0	0	0
2.	Background	Background	2	8	6	16
		Circumstance	21	9	9	39
3.	Cause	Cause	2	1	1	4
		Result	3	0	0	3
		Consequence	14	1	12	27
4.	Comparison	Comparison	5	9	4	18
		Preference	0	0	0	0
		Analogy	0	0	0	0
		Proportion	0	0	0	0
5.	Condition	Condition	15	1	1	17
		Hypothetical	1	1	0	2
		Contingency	0	0	0	0
		Otherwise	0	0	0	0
6.	Contrast	Contrast	19	2	2	23
		Concession	13	0	1	14
		Antithesis	25	1	4	30
7.	Elaboration	Elaboration-additional	23	238	41	302
		Elaboration-general-specific	1	16	4	21
		Elaboration-part-whole	0	0	0	0
		Elaboration-process-step	0	0	0	0
		Elaboration-object-attribute	4	179	3	186
		Elaboration-set-member	0	6	1	7
		Example	3	6	8	17

[9] The total number of relations analyzed is actually 1,304. The total in the table shows 1,306, because two relations are counted twice, two instances of List relation indicated by both discourse markers and other signals at the same time.

		Definition	0	2	0	2
8.	Enablement	Purpose	0	39	0	39
		Enablement	0	0	0	0
9.	Evaluation	Evaluation	1	3	1	5
		Interpretation	1	0	9	10
		Conclusion	0	0	0	0
		Comment	0	0	9	9
10.	Explanation	Evidence	0	3	8	11
		Explanation-argumentative	6	1	23	30
		Reason	12	1	4	17
11.	Joint	List	50	27	6	83
		Disjunction	3	0	0	3
12.	Manner-Means	Manner	3	0	0	3
		Means	1	4	0	5
13.	Topic-Comment	Problem-solution	2	2	2	6
		Question-answer	0	0	0	0
		Statement-response	0	2	0	2
		Topic-comment	1	0	0	1
		Comment-topic	0	0	0	0
		Rhetorical-question	0	0	0	0
14.	Summary	Summary	0	0	8	8
		Restatement	0	9	0	9
15.	Temporal	Temporal-before	3	0	0	3
		Temporal-after	7	1	0	8
		Temporal-same-time	3	1	0	4
		Sequence	5	0	0	5
		Inverted-sequence	0	0	0	0
16.	Topic-change	Topic-shift	0	0	4	4
		Topic-drift	0	0	0	0
17.	Same-unit	Same-unit	2	76	3	81
18.	Span	Span	0	0	0	0
19.	Textual Organization	Textual organization	0	1	0	1
	Total		251 (19.25%)	878 (67.33%)	177 (13.57%)	1,306

Table 5. Distribution of relations indicated by a DM (Discourse Marker), of relations indicated by some other signals, and of unsignalled relations

The distribution of relations in Table 5 shows that almost every group of relations is more or less signalled. In particular, we find that relation groups such as Attribution, Elaboration,

Enablement, and Joint are most frequently signalled, either by discourse markers or by some other signal[10]. We also found that that there is only one group of relations, Evaluation, which is rarely indicated by any signal.

Among the signalled relations, discourse markers are most frequently used to signal relations such as Circumstance, Result, Consequence, Condition, Concession, Contrast, Antithesis, Reason and List. In contrast, relations such as Attribution, Background, Comparison, Elaboration-additional, Elaboration-general-specific, Elaboration-object-attribute, Example and Purpose are rarely or never signalled by a discourse marker. Our findings are also parallel to the results presented in our earlier work (Taboada, 2006), where we found that relations such as Concession, Condition and Purpose are most frequently signalled (by a discourse marker), while Background and Summary are rarely signalled (by a discourse marker).

Relations which are mostly indicated by some other signals include Attribution, Elaboration-additional, Elaboration-general-specific, Elaboration-object-attribute, Purpose and Restatement. In contrast, relations which are rarely or never indicated by some other signals include Circumstance, Consequence, Condition, Contrast, Antithesis, Explanation-argumentative and Temporal-after.

Finally, the relations for which no signals (neither a discourse marker nor any other signal) were found include Comment, Summary and Topic-change.

The relation-wise distribution of discourse markers shows that a significant number of relations are frequently signalled by a wide variety of discourse markers. The distribution of the most frequently-occurring discourse markers with respect to the most common relations is provided in Table 6.

Common relation group	Common relation	Most frequently occurring discourse markers
Background (23)	Circumstance (21)	when (5), as (4), with (3),
Cause (19)	Consequence (14)	and (6)
Condition (16)	Condition (15)	if (11), unless (2)
Contrast (57)	Contrast (19)	but (11), however (3)
	Concession (13)	while (3), but (2), though (2)
	Antithesis (25)	but (11), although (3), however (3)
Elaboration (31)	Elaboration-additional (23)	and (8), but (6), as (2), so far (2)
	Example (3)	for example (2)
Explanation (18)	Reason (12)	and (4), because (4), because of (3)
Joint (53)	Disjunction (3)	or (3)
	List (50)	and (44), in addition (2), moreover (2)
Temporal (18)	Sequence (5)	and (4)
	Temporal-after (7)	since (3), after (2)
	Temporal-before (3)	before (3)

Table 6. Distribution of the most frequently occurring DMs with respect to the most common relations signalled by them

The distribution of different discourse markers provided in Table 6 shows what discourse markers are most frequently used to convey a particular relation, and how frequently they are

[10] We exclude Same-unit from this list because Same-unit is not a true coherence relation. In the RST Discourse Treebank it is used to join discontinuous grammatical elements, such as subject NP and VP.

used for signalling that relation[11]. For instance, List relations are most frequently signalled by *and, in addition*, and *moreover*. Out of the 50 instances of List relation, the DMs *and, in addition* and *moreover*, are used 44 (88%), 2 (4%), and 2 (4%) times, respectively. The complete distribution of the discourse markers with respect to the relations is provided online (see Footnote 5).

In an alternate combination, the distribution of the most common relations with respect to the most frequently-occurring discourse markers is provided in Table 7.

Table 7 shows what relations are most frequently signalled by a particular discourse marker, and how frequently they are signalled by that marker. For instance, the discourse marker *but* is most frequently used to signal Contrast and Elaboration relations. Out of the 35 instances of *but*, the relations Contrast and Elaboration are signalled 25 (71.43%) and 6 (17.14%) times, respectively.

Frequently Occurring DM	Common Relation Group	Common Relation
although (5)	Contrast (5)	Antithesis (3)
and (70)	Cause (7)	Consequence (6)
	Elaboration (8)	Elaboration-additional (8)
	Joint (44)	List (44)
	Explanation (4)	Reason (4)
	Temporal (4)	Sequence (4)
as (8)	Background (4)	Circumstance (4)
	Elaboration (2)	Elaboration-additional (2)
because (8)	Cause (2)	Consequence (2)
	Explanation (6)	Explanation-argumentative (2)
		Reason (4)
because of (6)	Explanation (4)	Reason (3)
before (4)	Temporal	Temporal-before (3)
but (35)	Contrast (25)	Antithesis (11)
		Concession (3)
		Contrast (11)
	Elaboration (6)	Elaboration-additional (6)
however (9)	Contrast (6)	Antithesis (3)
		Contrast (3)
if (13)	Condition (11)	Condition (11)
since (5)	Temporal (3)	Temporal-after (3)
when (10)	Background (5)	Circumstance (5)
while (8)	Comparison (3)	Comparison (3)
	Contrast (4)	Concession (3)
with (4)	Background (3)	Circumstance (3)
without (6)	Manner-Means	Manner (3)

Table 7. Distribution of the most common relations with respect to the most frequently occurring discourse markers

The relation-wise distribution of other signals and the other signal-wise distribution of relations show even more diverse relationships between the relations and the other signals. The

[11] The numerical value within parentheses following a relation/relation group refers to the number of instances the relation/relation group is signalled by a DM. The numerical value within parentheses following a DM refers to the number of times it is used to signal the corresponding relation. This applies to Tables 6, 7, 8 and 9.

distribution of the most frequently-used signals with respect to the most common relations is provided in Table 8.

Relation group	Relation	Other signal type	Specific signal
Attribution (228)	Attribution (228)	Syntactic (220)	Reported speech pattern (220)
		Genre (4)	Newspaper heuristics (4)
		Lexical (4)	VP cue (4)
Background (17)	Background (8)	Morphological (2)	Change of tense (2)
		Lexical (5)	indicative phrase (5)
	Circumstance (9)	Syntactic + positional (4)	Reduced relative clause + beginning (3)
		Lexical (5)	indicative phrase (5)
Comparison (9)	Comparison (9)	Lexical (8)	indicative phrase (7)
Elaboration (447)	Elaboration-additional (238)	Entity + syntactic (84)	Given entity + subject NP (74), given entity + subject NP (RS) (6)
		Entity (79)	Given entity (77)
		Lexical (8)	Indicative word (4), indicative phrase (4)
		Semantic (133)	Lexical overlap (61), lexical chain (60), phrasal chain (7)
		Syntactic (51)	Relative clause (26), reduced relative clause (10), participial clause (7)
		Genre (38)	Textual organization (32), newspaper heuristics (6)
		Graphical (16)	Parentheses (10), dashes (4)
	Elaboration-object-attribute (179)	Syntactic (167)	Relative clause (85), reduced relative clause (45), infinitival clause (NP) (27)
	Elaboration-general-specific (16)	Entity (5)	Given entity (5)
		Entity + syntactic (5)	Given entity + subject NP (3)
		Graphical (5)	Dash (4)
		Semantic (11)	Lexical chain (5), lexical overlap (5)
Enablement (39)	Purpose (39)	Syntactic (38)	Infinitival clause (37)
Joint (27)	List (27)	Syntactic (14)	Parallel syntactic constructions (9)
		Semantic (7)	Lexical chain (3)
Manner-Means (4)	Means (4)	Lexical + syntactic (4)	Indicative word + participial clause (4)
Summary (9)	Restatement (9)	Graphical (8)	Parentheses (7)

Table 8. Distribution of the most frequently used other signals with respect to the most common relations indicated by them

The relation-wise distribution of different other markers in Table 8 shows what other signals are most frequently used to indicate a particular relation, and how frequently they are used for indicating that relation. For instance, Elaboration-additional relations are most frequently signalled by semantic, syntactic, entity and genre features. More specifically, semantic, entity + syntactic, entity, syntactic and genre features are individually used 133 (55.88%), 84 (35.29%),

79 (33.19%), 51 (21.43%), 38 (15.97%) times, respectively, out of the 238 instances an Elaboration-additional relation is present[12].

In an alternate combination, the distribution of the most common relations with respect to the most frequently-occurring other signals is provided in Table 9.

The distribution of relations with respect to other markers in Table 9 shows what relations are most frequently indicated by a particular other signal, and also how frequently they are indicated by that signal. For instance, the signal *relative clause* is most frequently used to signal Elaboration-object-attribute and Elaboration-additional relations: Out of the 112 instances of *relative clauses*, Elaboration-object-attribute and Elaboration-additional relations are signalled 85 (75.89%) and 26 (23.21%) times, respectively.

Other marker type	Specific other marker	Relation group	Relation
Entity + syntactic (92)	Given entity + subject NP (78)	Elaboration (77)	Elaboration-additional (74), Elaboration-general-specific (3)
	Given entity + subject NP (RS) (7)	Elaboration (7)	Elaboration-additional (7)
Entity (87)	Given entity (84)	Elaboration (83)	Elaboration-additional (77)
Lexical (51)	Indicative phrase (40)	Background (9)	Background (5), Circumstance (4)
		Comparison (7)	Comparison (7)
		Elaboration (11)	Elaboration-additional (4), Elaboration-object-attribute (2), Elaboration-set-member (2), Example (3)
	Indicative word (10)	Elaboration (4)	Elaboration-additional (4)
Semantic (163)	Lexical chain (72)	Elaboration (66)	Elaboration-additional (60), Elaboration-general-specific (5)
	Lexical overlap (67)	Elaboration (66)	Elaboration-additional (61), Elaboration-general-specific (4)
Syntactic (573)	Reported speech pattern (223)	Attribution (220)	Attribution (220)
	Relative clause (112)	Elaboration (112)	Elaboration-object-attribute (85), Elaboration-additional (26)
	Reduced relative clause (55)	Elaboration (55)	Elaboration-object-attribute (45), Elaboration-additional (10)

[12] Note: In signalling relations by discourse markers, a single discourse marker is typically used to signal a particular instance of a relation. However, in signalling relations by signals other than discourse markers, two or more signals are frequently used at the same time to indicate a particular instance of a relation. As a result, the individual distribution score of a particular other signal, unlike that of a discourse marker, is not relative to that of any other signal.

	Infinitival clause (41)	Enablement (37)	Purpose (37)
		Elaboration (3)	Elaboration-additional (3)
	Infinitival clause (NP) (27)	Elaboration (27)	Elaboration-object-attribute (27)
	Participial clause (19)	Elaboration (18)	Elaboration-object-attribute (10), Elaboration-additional (7)
Genre (47)	Textual organization (36)	Elaboration (36)	Elaboration-additional (33)
	Newspaper heuristics (11)	Elaboration (6)	Elaboration-additional (6)
		Attribution (4)	Attribution (4)
Graphical (37)	Parentheses (18)	Elaboration (11)	Elaboration-additional (10)
		Summary (7)	Restatement (7)
	Dashes (12)	Elaboration (11)	Elaboration-additional (4), Elaboration-object-attribute (4)
Lexical + syntactic (18)	PP cue + participial clause (7)	Elaboration (7)	Elaboration-object-attribute (7)
	Indicative word + participial clause (5)	Manner-Means (4)	Means (4)
Syntactic + positional (6)	Reduced relative clause + beginning (3)	Background (3)	Circumstance (3)
	Parallel PP constructions + beginning (2)	Joint (2)	List (2)

Table 9. Distribution of the most common relations with respect to the most frequently-occurring other signals

In the specific case of Elaboration, there are indeed some significant differences in signalling the different types of Elaboration (see Table 8). Among the 447 instances of Elaboration relations, the majority is distributed between Elaboration-additional (238 instances) and Elaboration-object-attribute (179 instances) while the other types of Elaboration have much fewer tokens. Elaboration-additional relations are signalled by a wide variety of signals. The most important types (with higher number of tokens) include (i) entity + syntactic (84), (ii) entity (79), (iii) semantic (133), and (iv) syntactic (51). On the other hand, Elaboration-object-attribute relations are mainly signalled by syntactic features, in particular by features such as relative clause (130) and infinitival clause (27). While we may disagree in principle with the very specific breakdown of Elaboration, in this case it does seem that the annotators of the RST Discourse Treebank were on the right track, distinguishing subtypes that are different in their signalling. It is worth mentioning that Elaboration-object-attribute, a relation that has been questioned as not a true RST relation, but rather a derivative of entity relations (Knott et al., 2001), is actually not signalled through semantic or entity features (which would be the equivalent to the entity or reference relations that Knott et al. postulated). It is most frequently signalled by extensions to the noun that the relation modifies (relative and infinitival clauses).

As for the 177 relation instances for which we could not identify a signal (see Table 4), those include a number of different relation types, but there were three particular relations that were never signaled: Comment, Summary and Topic-shift (21 instances among the three). There are

three different reasons why we believe no signals could be found. First of all, in some cases we found that there were errors in the annotation, and a relation was postulated, whereas we would not have annotated a relation, or we would have proposed a different one. Summary and Elaboration in the RST-Discourse Treebank seem to be used in very similar contexts, so when a Summary was annotated, but we believed the relation was not in fact a summary, it was more difficult to find signals that would identify the relation as Summary. Secondly, some of the RST Discourse Treebank 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. Finding no signals in those cases is not surprising, as such phenomena are not likely to be indicated by the same type of signals 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. What may be happening in those cases is that the relation is being evoked, in the same way frames and constructions may be evoked (Dancygier & Sweetser, 2005). Dancygier and Sweetser propose that, in some constructions, only one aspect of the construction is necessary in order to evoke the entire construction. Such is the case with some instances of sentence juxtaposition, which give rise to a conditional relation reading, as in "Steal a bait car. Go to jail" (the slogan for a car-theft prevention campaign by the Vancouver police). No conditional connective is necessary. The juxtaposition of the two sentences, together with the imperative and a certain amount of world knowledge lead to the conditional interpretation.

We would like to conclude this section by repeating that our results show that relation signalling is much more sophisticated than previously thought, and that a certain level of redundancy is present in many relations. Recent work in the automatic identification of relations has postulated a clear separation between implicit and explicit relations. A series of experiments by Marcu and Echihabi (2002) and Sporleder and Lascarides (2005, 2008) have shown that it is difficult to generalize from "explicit" to "implicit" features, that is, that a classifier built using "explicit" relations does not necessarily identify "implicit" relations correctly (see also the discussion in Stede, 2012). We use quotes around "explicit" and "implicit" because we believe that existing definitions of those terms are too narrow. If by "explicit" we mean relations signalled exclusively by discourse markers, then it may be the case, as Sporleder and Lascarides (2008) conclude, that those two types are different in nature. However, if explicit is extended to include other types of signals, and particularly semantic signals, we believe that the two types may not be that different in nature, and automatic classification may be possible (assuming, of course, complex annotation of the type carried out here, and identification of those semantic relations in unseen data).

7 Discussion: Relation signalling and layered annotations

The first goal is this ongoing annotation effort was to investigate whether signals other than discourse markers exist for coherence relations. In this respect, we can confidently say that this is, indeed, the case: Out of the 1,127 signalled relations, 878 (77.91%) contain a signal other than a discourse marker. Although some of the relations (13.57% of the total 1,304) are not signalled, the overwhelming majority of them are.

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). It also means that the signals,

as linguistic 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 the relation exclusively.

One may argue that the signals that we have identified are quite intricate, and that an automatic system would have a very hard time making use of them. This is especially the case with the semantic and lexical relations, where some of the relations are identified based not only on WordNet-type relations (Fellbaum, 1998), but also on world knowledge. An example from our corpus is an Elaboration relation that relies on the semantic connection between *the Philippine company* in the nucleus and *Luzon Petrochemical Corp.* in the satellite. Identifying that connection may require knowledge about Luzon being a Philippine island, which is beyond the scope of WordNet.

In this paper, we are not, however, directly concerned with the issue of automatic identification. We merely wish to point out that more signals than previously found are present in many of the relations. Automatic identification of relations would require some disambiguation, of the same type that is already necessary for discourse markers, some of which have non-discourse functions (Hirschberg & Litman, 1993).

We will devote the rest of this section to issues having to do with annotating discourse phenomena, and with the difficulties in adding annotations to an existing resource.

One of our main difficulties in annotating discourse phenomena has to do with the more loose definition of what counts as a signal. Although we tried to create a very detailed list of signals, and documented those signals with many examples, it is undeniable that this type of annotation is subjective. Our reliability study shows a decent level of agreement between annotators. As we already discussed in Section 5.2, this is often the case with published studies, and to be expected in a research group where members work closely together and under the same assumptions. The question that we would like to address here is how difficult it is in general to annotate phenomena that are more abstract than, for instance, part of speech tags (which also contain a certain level of abstraction and are by no means straightforward). Our view on this is that phenomena at the discourse level are as easy or as difficult to identify as phenomena at other levels of the language. The main criterion for reliable annotation is a clear set of guidelines and, in particular, a clearly defined taxonomy. We found that distinguishing between signals that belonged in the categories "Entity" and "Semantic" was the basis of many of our disagreements. Initially, we had reserved the category "Entity" for those signals that involved reference to the same referent. The category "Semantic" was reserved for semantic relations that do not necessarily involve same reference, such as synonymy. This distinction works along the lines of Halliday and Hasan's (1976) grammatical versus lexical cohesion, with Entity signals being close to the reference system in Halliday and Hasan's grammatical cohesion. Our Semantic group of signals contains lexical cohesion relations, such as synonyms, antonyms and hypernyms. The problem, however, is that lexical cohesion also includes repetition of the same item which is, strictly speaking, reference to the same referent, and thus Entity in our system. Each one of us had made a different assumption about how to deal with this problem (one including repetition as Entity, the other as Semantic). One of the lessons learned in this process was to stick to the tried and true as much as possible, and rely on existing taxonomies, or else motivate our departure from them.

This lesson leads us to the discussion of the other issue in the annotation of higher-level phenomena. As we have mentioned throughout the paper, we are dealing with an existing corpus, already annotated for discourse relations. We believe that this will be more and more the case,

with so many available resources already annotated for a wide range of phenomena. We found ourselves disagreeing with many of the annotation decisions in the initial corpora, from the number of relations to the definition of what an elementary unit of discourse is. The RST Discourse Treebank uses a very large set of 78 relations, including a high number of subtypes of Elaboration. In practice, this meant that we had to keep all these distinctions in mind as we annotated.

More difficult for our purposes was the fine-grained segmentation. The traditional definition of minimal unit of discourse in RST proposes that clauses should be minimal units, excluding subject and object clauses. In other words, it is mostly adverbial clauses that have a function at the discourse level. Mann and Thompson (1988), in this as in many other aspects, leave the door open for other definitions, if they suit the researcher's purposes. The authors of the RST-Discourse Treebank decided on a segmentation method that classifies all types of clauses as elementary discourse units (EDUs). In particular, noun clauses as objects of verbal processes (*say, tell, claim*) are considered to be units of discourse in the RST Discourse Treebank. Carlson and Marcu (2001) then proposed a new RST relation, Attribution, to connect the reported speech verb and its complement. Similarly, relative clauses and noun clauses that modify nouns (*Alson Lee, who heads the Philippine company…; a contract to build…*) are also elementary discourse units. We found that such level of detail made our annotation quite difficult, in part because we disagree with the notion that noun and relative clauses stand in any kind of discourse relation to the words that they modify.

The clause-internal relations (which specifically represent the relationships between two entities or between an entity and a proposition) mainly include Attribution and Elaboration. These relations are usually signalled by syntactic features. The distribution is provided in Table 10.

Syntactic feature	Relation group	Relation
Infinitival clause (NP) or Noun clause (27)	Elaboration (27)	Elaboration-object-attribute (27)
Participial clause (19)	Elaboration (18)	Elaboration-additional (7), Elaboration-object-attribute (10), Elaboration-general-specific (1)
	Enablement (1)	Purpose (1)
Reduced relative clause (55)	Elaboration (55)	Elaboration-additional (10), Elaboration-object-attribute (45)
Relative clause (112)	Elaboration (112)	Elaboration-additional (26), Elaboration-object-attribute (85), Definition (1)
Reported speech pattern (223)	Attribution (220)	Attribution (220)
	Evaluation (2)	Evaluation (2)
	Statement-response (1)	Statement-response (1)

Table 10. Distribution of clause-internal relations in terms of syntactic features

We found our agreement in annotating these relations to be quite high, as syntactic phenomena tend to be easier to identify. Nonetheless, in most cases we felt that the relation was, in fact, syntactic, rather than a discourse or coherence relation.

As we performed the annotation of signals, we also found ourselves disagreeing with specific aspects of the RST annotation, such as the label for a particular relation or the nucleus-satellite assignation. These types of errors are to be expected in discourse annotation, and we do not take

issue with them, as they are the result of human error and they tend to be localized. One question that arises, however, is whether we should be making corrections in cases of obvious mistakes. Although that would probably make the corpus better, we have decided not to alter it, as it has become a standard in many studies.

In summary, our experience shows that, although layering upon an existing annotation is challenging, the results are certainly worthwhile. We have shown that rhetorical relations have multiple signals associated with them, and we hope to be on our way to determining how those signals can be used to perform automatic identification of relations.

8 Conclusions

We have presented an annotation effort that adds signalling information to an existing corpus of rhetorical relations. The purpose of the study was to determine to what extent rhetorical relations carry signals that may help readers and hearers identify the relation. Research so far has focused mainly on one type of signals, discourse markers, and has thus concluded that the majority of relations are implicit, that is, they contain no overt signal. We have shown that this is not the case and that, although there may still exist some implicit relations, most of the relations in our corpus are explicit, that is, they are signalled, sometimes through multiple signals.

In the process of annotating the corpus, we have discovered and solved a number of issues involving creating accurate and manageable taxonomies of signals, adding information to an existing corpus, and mapping relations and signals to each other.

The annotation described in this paper is a preliminary pilot study, comprising only 10% of the total corpus. In future work, we will expand to cover the entire corpus. The most important qualitative change for the rest of the annotation involves finding a method to layer annotations on top of the existing LISP-style notation for the RST corpus.

The finished corpus has two clear applications. From a psycholinguistic point of view, we hope to be able to use it to determine how hearers and readers use signals to identify relations. Most of the psycholinguistic studies to date have manipulated relations by adding or deleting discourse markers. It would be very useful to extend that work by changing other types of signals, to see what effects that has on comprehension.

The other main application of such an annotated corpus is in discourse parsing. A great deal of recent work (Hernault et al., 2010; Hernault et al., 2011; Mithun & Kosseim, 2011; da Cunha et al., 2012) and also earlier approaches (Corston-Oliver, 1998b; Marcu, 2000a; Schilder, 2002) have used discourse markers as the main signals to automatically parse relations, and almost exclusively at the sentence level. Our extended set of signals, and the fact that they work at all levels of discourse, will probably facilitate this task.

Acknowledgements

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada (Discovery Grant 261104-2008). We would like to thank the audience at the 7th Conference of the Swiss Linguistic Society (Lugano, September 2012), and the reviewers and the editors of the special issue for very useful feedback and comments.

References

Afantenos, Stergos D., Denis, Pascal, Muller, Philippe, & Danlos, Laurence. (2010). Learning recursive segments for discourse parsing. *International Conference on Language Resources and Evaluation (LREC)*. Malta.

Asher, Nicholas, & Lascarides, Alex. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Baldridge, Jason, Asher, Nicholas, & Hunter, Julie. (2007). Annotation for and robust parsing of discourse structure on unrestricted texts. *Zeitschrift für Sprachwissenschaft*(26), 213-239.

Bateman, John, Kamps, Thomas, Kleinz, Jörg, & Reichenberger, Klaus. (2001). Towards constructive text, diagram, and layout generation for information presentation. *Computational Linguistics, 27*(3), 409-449.

Blakemore, Diane. (1987). *Semantic Constraints on Relevance*. Oxford: Blackwell.

Blakemore, Diane. (1992). *Understanding Utterances: An Introduction to Pragmatics*. Oxford: Blackwell.

Blakemore, Diane. (2002). *Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers*. Cambridge: Cambridge University Press.

Britton, Bruce K., Glynn, Shawn M., Meyer, Bonnie J. F., & Penland, M. J. (1982). Effects of text structure on use of cognitive capacity during reading. *Journal of Educational Psychology, 74*(1), 51-61.

Carlson, Lynn, & Marcu, Daniel. (2001). *Discourse Tagging Manual*. Manual. http://www.isi.edu/~marcu/discourse/tagging-ref-manual.pdf.

Carlson, Lynn, Marcu, Daniel, & Okurowski, Mary Ellen. (2002). RST Discourse Treebank, LDC2002T07 [Corpus]. Philadelphia, PA: Linguistic Data Consortium.

Corston-Oliver, Simon. (1998a). Beyond string matching and cue phrases: Improving efficiency and coverage in discourse analysis. *Proceedings of AAAI 1998 Spring Symposium Series, Intelligent Text Summarization* (pp. 9-15). Madison, Wisconsin.

Corston-Oliver, Simon. (1998b). Identifying the linguistic correlates of rhetorical relations. *Proceedings of the Workshop on Discourse Relations and Discourse Markers, COLING/ACL '98* (pp. 8-14).

Cristea, Dan, Ide, Nancy, & Romary, Laurent. (1998). Veins Theory: A model of global discourse cohesion and coherence. *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (ACL-98/COLING-98)* (pp. 281-285). Montréal, Canada.

Cunningham, Hamish, Maynard, Diana, Bontcheva, Kalina, Tablan, Valentin, Aswani, Niraj, Roberts, Ian, . . . Peters, Wim. (2011). *Text Processing with GATE* (6th ed.). Sheffield, UK: Gate.

da Cunha, Iria, San Juan, Eric, Torres-Moreno, Juan Manuel, Cabré, María Teresa, & Sierra, Gerardo. (2012). A symbolic approach for automatic detection of nuclearity and rhetorical relations among intra-sentence discourse segments in Spanish. *Proceedings of CICLing* (pp. 462-474). New Delhi, India.

Dale, Robert. (1991). The role of punctuation in discourse structure. *Proceedings of AAAI Fall Symposium on Discourse Structure in Natural Language Understanding and Generation* (pp. 13-14). Asilomar, CA.

Degand, Liesbeth, & Sanders, Ted. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing, 15*(7-8), 739-758.

Fellbaum, Christiane (Ed.). (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Fraser, Bruce. (1990). An approach to discourse markers. *Journal of Pragmatics, 14*(3), 383-398.

Fraser, Bruce. (1999). What are discourse markers? *Journal of Pragmatics, 31*(7), 931-952.

Fraser, Bruce. (2006). Towards a theory of discourse markers. In K. Fischer (Ed.), *Approaches to Discourse Particles* (pp. 189-204). Amsterdam: Elsevier.

Fraser, Bruce. (2009). An account of discourse markers. *International Review of Pragmatics, 1*, 293-320.

Grice, H. Paul. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech Acts. Syntax and Semantics, Volume 3* (pp. 41-58). New York: Academic Press.

Haberlandt, Karl. (1982). Reader expectations in text comprehension. In J.-F. Le Ny & W. Kintsch (Eds.), *Language and Comprehension* (pp. 239-249). Amsterdam: North-Holland.

Halliday, Michael A. K., & Hasan, Ruqaiya. (1976). *Cohesion in English*. London: Longman.

Hanneforth, Thomas, Heintze, Silvan, & Stede, Manfred. (2003). Rhetorical parsing with underspecification and forests. *Proceedings of HLT-NAACL 2003*. Edmonton, Canada.

Hernault, Hugo, Bollegala, Danushka, & Ishizuka, Mitsuru. (2011). Semi-supervised discourse relation classification with structural learning. *Proceedings of the 12th international conference on Computational linguistics and intelligent text processing (CICLing '11)*. Tokyo, Japan.

Hernault, Hugo, Prendinger, Helmut, duVerle, David A., & Ishizuka, Mitsuru. (2010). HILDA: A discourse parser using Support Vector Machine classification. *Dialogue and Discourse, 1*(3).

Hirschberg, Julia, & Litman, Diane J. (1993). Empirical studies on the disambiguation of cue phrases. *Computational Linguistics, 19*(3), 501-530.

Horn, Laurence. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form and use in context: Linguistic Implications* (pp. 11-42). Washington, DC: Georgetown University Press.

Hovy, Eduard, & Arens, Yigal. (1991). Automatic generation of formatted text. *Proceedings of 9th AAAI Conference*. Anaheim, California.

Huong, Le Thanh. (2007). An approach in automatically generating discourse structure of text. *Journal of Computer Science and Cybernetics, 23*(3), 212-230.

Karamanis, Nikiforos. (2007). Supplementing entity coherence with local rhetorical relations for information ordering. *Journal of Logic, Language and Information, 16*, 445-464.

Keenan, Janice M., Baillet, Susan D., & Brown, Polly. (1984). The effects of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behaviour, 23*(2), 115-126.

Knott, Alistair. (1996). *A Data-Driven Methodology for Motivating a Set of Coherence Relations.* Ph.D. dissertation, University of Edinburgh, Edinburgh, UK. Retrieved from http://citeseer.nj.nec.com/knott96data.html

Knott, Alistair, & Dale, Robert. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes, 18*(1), 35-62.

Knott, Alistair, Oberlander, Jon, O'Donnell, Michael, & Mellish, Chris. (2001). Beyond elaboration: The interaction of relations and focus in coherent text. In T. Sanders, J. Schilperoord & W. Spooren (Eds.), *Text Representation: Linguistic and Psycholinguistic Aspects* (pp. 181-196). Amsterdam and Philadelphia: John Benjamins.

Knott, Alistair, & Sanders, Ted. (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics, 30*, 135-175.

Lapata, Mirella, & Lascarides, Alex. (2004). Inferring sentence-internal temporal relations. *Proceedings of the North American Chapter of the Assocation of Computational Linguistics (NAACL)* (pp. 153-160). Boston, MA.

Lin, Ziheng, Kan, Min-Yen, & Ng, Hwee Tou. (2009). Recognizing implicit discourse relations in the Penn Discourse Treebank. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing* (pp. 343-351). Singapore.

Louis, Annie, Prasad, Rashmi, Joshi, Aravind K, & Nenkova, Ani. (2010). Using entity features to classify implicit discourse relations. *Proceedings of the SIGDIAL conference* (pp. 59-62). Tokyo, Japan.

Mann, William C., & Taboada, Maite. (2010). RST Web Site, from http://www.sfu.ca/rst

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

Marcu, Daniel. (1999). A decision-based approach to rhetorical parsing. *Proceedings of 37th Annual Meeting of the Association for Computational Linguistics (ACL'99)* (pp. 365-372). College Park, Maryland.

Marcu, Daniel. (2000a). The rhetorical parsing of unrestricted texts: A surface based approach. *Computational Linguistics, 26*(3), 395-448.

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

Marcu, Daniel, & Echihabi, Abdessamad. (2002). An unsupervised approach to recognising discourse relations. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)* (pp. 368-375). Philadelphia, Pennsylvania.

Meyer, Bonnie J. F., Brandt, David M., & Bluth, George J. (1980). Use of top-level structure in text: Key for reading comprehension in ninth-grade students. *Reading Research Quarterly, 16*(1), 72-103.

Millis, Keith K., & Just, Marcel A. (1994). The influence of connectives on sentence comprehension. *Journal of Memory and Language, 33*, 128-147.

Mithun, Shamima, & Kosseim, Leila. (2011). Comparing approaches to tag discourse relations. *Proceedings of the 12th international conference on Computational linguistics and intelligent text processing (CICLing '11)* (pp. 328-339). Tokyo, Japan.

Myers, Jerome L., Shinjo, Makiko, & Duffy, Susan A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language, 26*(4), 453-465.

O'Donnell, Michael. (1997). RST-Tool: An RST analysis tool. *Proceedings of the 6th European Workshop on Natural Language Generation*. Duisburg, Germany.

Pardo, Thiago Alexandre Salgueiro, & Nunes, Maria das Graças Volpe. (2008). On the development and evaluation of a Brazilian Portuguese discourse parser. *Journal of Theoretical and Applied Computing, 15*(2), 43-64.

Pitler, Emily, Louis, Annie, & Nenkova, Ani. (2009). Automatic sense prediction for implicit discourse relations in text. *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics* (pp. 683-691). Suntec, Singapore.

Poesio, Massimo, Stevenson, Rosemary, Di Eugenio, Barbara, & Hitzeman, Janet. (2004). Centering: A parametric theory and its instantiations. *Computational Linguistics, 30*(3), 309-363.

Polanyi, Livia, Culy, Christopher, van der Berg, Martin, Thione, Gian Lorenzo, & Ahn, David. (2004). A rule based approach to discourse parsing. *Proceedings of the 5th SIGdial Workshop in Discourse and Dialogue* (pp. 108-117). Cambridge, MA.

Polanyi, Livia, & Scha, R. J. H. (1983). The syntax of discourse. *Text, 3*(3), 261-270.

Prasad, Rashmi, Joshi, Aravind K, & Webber, Bonnie. (2010). Realization of discourse relations by other means: Alternative lexicalizations. *Proceedings of COLING 2010* (pp. 1023-1031). Beijing.

Prasad, Rashmi, Lee, Alan, Dinesh, Nikhil, Miltsakaki, Eleni, Campion, Geraud, Joshi, Aravind K., & Webber, Bonnie. (2008). Penn Discourse Treebank Version 2.0, LDC2008T05 [Corpus]. Philadelphia, PA: Linguistic Data Consortium.

Prasad, Rashmi, Miltsakaki, Eleni, Dinesh, Nikhil, Lee, Alan, Joshi, Aravind K, Robaldo, Livio, & Webber, Bonnie. (2007). The Penn Discourse Treebank 2.0 Annotation Manual. Philadelphia: University of Pennsylvania.

Sanders, Ted, Land, Jentine, & Mulder, Gerben. (2007). Linguistic markers of coherence improve text comprehension in functional contexts. *Information Design Journal, 15*(3), 219-235.

Sanders, Ted, & Noordman, Leo. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes, 29*(1), 37-60.

Sanders, Ted, Spooren, Wilbert, & Noordman, Leo. (1992). Toward a taxonomy of coherence relations. *Discourse Processes, 15*(1), 1-35.

Sanders, Ted, Spooren, Wilbert, & Noordman, Leo. (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics, 4*(2), 93-133.

Schiffrin, Deborah. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.

Schiffrin, Deborah. (2001). Discourse markers: Language, meaning and context. In D. Schiffrin, D. Tannen & H. E. Hamilton (Eds.), *The Handbook of Discourse Analysis* (pp. 54-75). Malden, MA: Blackwell.

Schilder, Frank. (2002). Robust discourse parsing via discourse markers, topicality and position. *Natural Language Engineering, 8*(2/3), 235-255.

Scott, Donia, & de Souza, Clarisse Sieckenius. (1990). Getting the message across in RST-based text generation. In R. Dale, C. Mellish & M. Zock (Eds.), *Current Research in Natural Language Generation* (pp. 47-73). London: Academic Press.

Siegel, Sidney, & Castellan, N.J. (1988). *Nonparametric Statistics for the Behavioral Sciences.* New York: McGraw-Hill.

Spooren, Wilbert. (1997). The processing of underspecified coherence relations. *Discourse Processes, 24*, 149-168.

Spooren, Wilbert, & Degand, Liesbeth. (2010). Coding coherence relations: Reliability and validity. *Corpus Linguistics and Linguistic Theory, 6*(2), 241-266.

Sporleder, Caroline, & Lascarides, Alex. (2005). Exploiting linguistic cues to classify rhetorical relations. *Proceedings of Recent Advances in Natural Language Processing (RANLP-05)*. Borovets, Bulgaria.

Sporleder, Caroline, & Lascarides, Alex. (2008). Using automatically labelled examples to classify rhetorical relations: An assesment. *Natural Language Engineering, 14*(3), 369-416.

Stede, Manfred. (2007). *Korpusgestützte Texanalyse: Grundzüge der Ebenen-orientierten Textlinguistik*. Tübingen: Gunter Narr.

Stede, Manfred. (2012). *Discourse Processing*: Morgan & Claypool.

Taboada, Maite. (2004). *Building Coherence and Cohesion: Task-Oriented Dialogue in English and Spanish*. Amsterdam and Philadelphia: John Benjamins.

Taboada, Maite. (2006). Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics, 38*(4), 567-592.

Taboada, Maite. (2008). SFU Review Corpus [Corpus]. Vancouver: Simon Fraser University, http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html.

Taboada, Maite. (2009). Implicit and explicit coherence relations. In J. Renkema (Ed.), *Discourse, of Course* (pp. 127-140). Amsterdam and Philadelphia: John Benjamins.

Taboada, Maite, & Mann, William C. (2006). Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Studies, 8*(3), 423-459.

Teijssen, Daphne, van Halteren, Hans, Verberne, Suzan, & Boves, Lou. (2008). Features for automatic discourse analysis of paragraphs. *Proceedings of the 18th Meeting of Computational Linguistics in the Netherlands (CLIN 2007)* (pp. 53-68). Leuven, Belgium.

Tofiloski, Milan, Brooke, Julian, & Taboada, Maite. (2009). A syntactic and lexical-based discourse segmenter. *Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics* (pp. 77-80). Singapore.

Trabasso, Tom, & Sperry, Linda L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language, 24*(5), 595-611.