Coherence Errors in Iranian EFL Learners' Writing: A Rhetorical Structure Theory Approach[1]

Alireza Ahmadi*[2]
Salma Parhizgar[3]

Abstract

One of the key elements in the organization of any piece of writing is its coherence. To date, many propositions have been given regarding the definition, analysis, and evaluation of text coherence. In the current study, Mann and Thompson's (1988) Rhetorical Structure Theory (RST) was adopted as the method of text analysis to detect the coherence breaks in writing samples. In order to see what problems Iranian EFL learners have with regard to text coherence, 64 essays in descriptive and argumentative genres written by male students of a language institute in Shiraz were analyzed. The essays were analyzed for discourse errors using RST. The findings indicated that Iranian EFL learners committed eight different types of coherence errors, namely *irrelevant content*, violation *of completedness*, *violation of connectedness*, *incorrect place*, *incorrect relation*, *crossed dependency*, *scattered units*, and *topic*. The reason behind these errors partly came from the learners' tendency to write in an inductive order, and partly from their inability to coherently connect the constituent parts of their texts together. Genre difference was also proved to be significant in the number of coherence relations and in the type and

number of coherence errors. In general, descriptive writing samples were more coherent than argumentative ones.

Introduction

Discourses are more than mere sequences of sentences. Consequently, the meaning of a coherent discourse is more than the sum of the meanings of its diverse parts. But what is it that is more than simple successive utterances? It is coherence which helps us understand and convey substantially more meaning than the meaning of individual sentences alone (e.g., Egg & Redeker, 2006). In fact, one of the key elements in the organization of any piece of writing is its coherence. No single written text could ever be fully comprehensible unless some level of go-togetherness is reflected through all of its constituent parts. In other words, a text is said to be coherent when its sentences flow smoothly from one to another without any gaps.

In comparison to the more familiar language components such as grammar or vocabulary, the concept of coherence is thought to be more technical and thus to be less universally understood. This consideration seems to be true despite the fact that a great number of researchers have defined, and have explored the notion of discourse coherence (Abu Shawish, 2015; Berman & Slobin, 1994; Cook, 1989; Egg & Redeker, 2006; Fitzgerald & Spiegel, 1990; Halliday & Hasan, 1976). The reason might be that there is not such a thing as absolute indicator of coherence. A text may be *more* or *less* coherent (Sperber & Wilson, 1986). So it may need more or less processing effort to become a coherent unit. And if it does not, it would be regarded as incoherent.

Many propositions have been given regarding the analysis and evaluation of text coherence, and its relationship with other components of a written text such as grammar, vocabulary, punctuation, accuracy, content, etc. (cf. Fox, 1987; Halliday & Hasan, 1976; Polanyi, 1985; Reichman, 1978; van Dijk, 1980). The question, however, is how much these theories could account for the realization and/or resolution of what is known as *coherence breaks* or *coherence errors* in paragraphs and compositions written by EFL learners; knowing the fact that incoherence is a common problem and a predominant error in the students' writings.

Trying to find a theory of discourse coherence, one may end up with a large number of highly-distinguished propositions which claim they are able to address coherence relations (e.g., Lascarides & Asher's Segmented Discourse Representation Theory, 1993; Martin's Conjunctive Relations, 1992; Mann & Thompson's Rhetorical Structure Theory, 1988; Sperber & Wilson's Relevance Theory, 1986; etc.). Among these, however, RST outstands in several ways one of which is that in this approach to text coherence, identifying the writer's in-

tentions and the effect of the relation on the reader are as important as unfolding the rhetorical relations themselves. As mentioned by Taboada and Mann (2006) in their state-of-the-art article and confirmed by many practitioners in the field, the theory has much superiority over its counterparts, in that, it proposes a different and a more complete view of text organization, and it points to a tight connection between relations and coherence in texts.

As an approach to the study of text organization, RST gains its prominence from the way in which it explains coherence. Based on this model of text analysis, certain types of relations hold between text parts of a wide range, from clauses to groups of paragraphs. As a result, the overall coherence of the text is defined in accordance with the role each part plays with respect to other parts. As Mann (2005), one of the co-founders of the theory explains, a coherent text is the one which lacks any non sequitur; i.e. every part of the text has a reasonable presence, noticeable to readers. It is this characteristic of plausibility that RST analysis focuses on.

The present study intends to examine the types of coherence errors which Iranian FL learners make while writing in English. We decided on RST as the theory of analysis. The theory was chosen wisely having a number of superiorities over its major counterparts. According to RST, no analysis is complete until the whole text is brought within a single *global* structure. RST proposes a different and a more complete view of text organization, and it points to a tight connection between relations and coherence in texts (Taboada & Mann, 2006). In this respect, RST may stand high above theories like Segmented Discourse Representation Theory (Lascarides & Asher, 1991), Conjunctive Relations (Martin, 1992), or Centering Theory (Grosz, Joshi & Weinstein, 1995).

A study of textual organization cannot be comprehensive enough without the consideration of the variability between genres. For this reason, it would also be of interest to take into account the role that genre-dependent differences might play in constituting the written coherence mistakes of the same group of learners. Therefore, the research also aims at investigating the prospective differences between the kinds or number of coherence errors made by EFL learners across distinctive genres.

Research Questions

In line with objectives of the study, the following research questions were put forward:

1. What are the coherence errors in the writings of Iranian EFL learners?
2. How do the coherence errors differ in the descriptive and argumentative genres?

Coherence Relations

Whatever aspect of the coherence in their mind, almost all of the discourse analysts agree upon the fact that "an essential part of comprehending and creating discourse is the recognition of intended relations holding between component discourse segments" (Bateman & Rondhuis, 1997, p.3). Hovy (1990) refers to the rhetorical relations as the "basic building elements" of a coherent text.

Coherence relations are among those substantial characteristics which lead to the achievement of coherence in discourse. As Taboada (2006) makes it clear, when the hearer or reader recognizes coherence relations in a text, they can allocate coherence to that text. However, it is not obvious how much the reader is aware of the presence of such relations and their effect on his comprehension. What is unquestionable though is that the readers or hearers process a text incrementally; that is they add new information to the continuing string of text. Inarguably, when the hearers or readers are not able to understand why two discourse segments are placed next to one another or when they cannot find the relation between them, they judge the text as incoherent.

Different linguists have recognized different types of relations known as *rhetorical predicates* (Grimes, 1975), *conjunctive relations* (Halliday & Hasan, 1976), *paragraph types* (Longacre, 1977), *sequitutity relations* (Fillmore, 1974), etc. between clauses, sentences, or larger portions of discourse. A list of coherence relations in the form of taxonomies is what many of these linguists proposed while studying the ways in which utterances link together in a text.

The subject has always been appealing, albeit researchers are not unanimous in delivering a single set of relations. Major theorists of discourse coherence (e.g., Grosz & Sidner, 1986; Hobbs, 1985; Longacre, 1983; Mann & Thompson, 1988;, Martin 1992) have put forward dissimilar sets of relations, the number of which in each set varies from two (Grosz & Sidner, 1986) to over a hundred (Carlson, Marcu & Okurowski, 2002). The relation definitions also vary in terms of their association with semantics (Longacre, 1983; Hobbs, 1985), intentions (Grosz & Sidner, 1986), or a combination of both (Mann & Thompson, 1988).

The so-called "relational coherence" has been the focus of attention by many researchers. But much of the theoretical and empirical research in this area is based on the notions of Rhetorical Structure Theory by Mann and Thompson (1988). Moreover, various attempts to taxonomize RST relations and categorize them into different types or groups support the RST's appropriateness (Taboada & Mann, 2006).

Mann and Thompson (1988) warn that "no single taxonomy seems to be suitable" (p. 256). In the 1988 article, they proposed 24 relations which were then expanded to 30. They could be classified into two distinct groups: subject matter in which the relation would be intentionally recognized by the reader (e.g., Elaboration, Circumstance, Solutionhood, Cause, & Restatement) and presentational relations by which the reader's inclination would increase (e.g., Motivation, Background, Justify & Concession). (see Appendices A & B).

The broad domains of computational linguistics, functional linguistics, and formal discourse analysis have all produced rigorous reports of discourse structure. In most of these approaches, the coherence encompasses relational constructs that hold a text together. There is still no consensus concerning the nature of these relations but the general realm of relations prepares a suitable ground to carry out various comparisons, inclusions, or to yield new descriptions and propositions.

A Brief Introduction to RST

The notion of implicit relations was proposed for the first time in cross-linguistic studies (e.g., Ballard, Conrad & Longacre, 1971; Beekman & Callow, 1974; Grimes, 1975) and was later on developed within the framework of systematic grammar (e.g., Martin, 1992), computational linguistics (e.g., Hobbs, 1976; Grosz & Sidner, 1986; Mann & Thompson, 1987), and psycholinguistics (e.g., Sanders, Spooren, & Noordman, 1992). The notion has been justified by theories referred to as theories of coherence relations. Summarized below is the theory which inspired most research in discourse and coherence relations.

As the most powerful approach to coherence relations, Rhetorical Structure Theory is an important theoretical and practical advance which makes it possible to linguistically analyze discourse rhetorical structures and their relationships. The theory applies a *global* view of text and its main difference with other theories is that the emphasis is on the writer's *intentions* and the effect of the relation on the reader (van Dijk and Kintsch's, 1983).

According to RST a text is coherent when it is possible to construct at least one hierarchical structure for the text as a whole. To construct a hierarchy, the text is broken into units (mostly clauses) and each unit, then, becomes connected to another by adding a relation. An RST relation/coherence relation/discourse relation/conjunctive relation/rhetorical relation emerges when one part of a text plays a specific role relative to the other (see Table 1). Rhetorical relations are described in the theory and it has been claimed that the relations hold between individual clauses as well as larger segments of the text. The text segments (the nucleus and the satellite) are of two different weights; unlike the satellites, omitting the nucleus is considered to ruin the whole text. This is because nuclei are the most important parts of a text and satellites are the secondary parts which contribute to the nuclei, containing some additional information about them. As such, a text is not comprehensible without its nucleus, whereas it can be well understandable with no satellites being added.

The kind of analysis seen in RST normally approaches top-down and it highly necessitates the discourse analyst's intuition. In the first step, the text is segmented into the largest communicative units and the relations between them are characterized. The process then continues until the very minimal-sized units (mostly clauses) are recognized and the relations between them are found.

The single rhetorical structure must comply with RST organizational constraints. The four restrictions are: completedness (one schema application contains the entire text); connectedness (each span, except for the one that contains the entire text, is either a minimal unit or a constituent of another schema application); uniqueness (each schema application consists of a different set of text spans); and adjacency (text spans of each schema application constitute one text span). Any violation of coherence relations and/or their associated diagrams, then, will be considered as a coherence error. Having this in mind, it is no surprise to expect a coherence error wherever the whole or part of an RST diagram of an EFL learner's writing does not conform to the aforementioned conditions and constraints.

Table 1.
Categorization of RST relations; taken from Mann and Thompson (1988)

Presentational Relations	Subject Matter Relations	Multinuclear Relations
Antithesis	Circumstance	Conjunction
Background	Condition	Contrast
Concession	Elaboration	Disjunction
Enablement	Evaluation	Joint
Evidence	Interpretation	List
Justify	Means	Multinuclear Restatement
Motivation	Non-volitional Cause	Sequence
Preparation	Non-volitional Result	
Restatement	Otherwise	
Summary	Purpose	
	Solutionhood	
	Unconditional	
	Unless	
	Volitional Cause	
	Volitional Result	

The exalted properties of RST according to Taboada and Mann (2006) makes it possible to analyze and consequently understand a text, to detect "the conceptual structure of relations and coherence", and "as a conceptual starting point" to contribute to "a great diversity of work in several fields" (p. 448).

To put it briefly, RST is a fine, acceptable theory with a large number of practitioners in different domains. It helps the analyst to recognize the most important parts of a text and to make implicit relations more accessible. It attempts to explain the coherence of most texts applying rhetorical relations.

RST also includes two essential characteristics. On the one hand, a particular set of relations is specified by its developers. On the other hand, the location and the nature of its predefined units are explicitly stated. Unlike other theories, it provides a hierarchical rather than a linear presentation of coherence and also offers a systematic way of error analysis for the researcher. These features can make RST one of the highly accepted and most influential approaches to the analysis of coherence relations.

RST Applications

Rhetorical Structure Theory has a vast range of applications; from uses in computational linguistics (generation, parsing, summarization, argument evaluation, machine translation, and essay scoring) (e.g., Grosz & Sidner, 1986; Skoufaki, 2009) to ongoing research in the area of cross-linguistics (Cui, 1986; Péry-Woodley, 1998, 2001; Scott, Delin & Hartley, 1999; Taboada, 2001, 2004a, 2004b, etc.). The theory has been chosen as the theory of application in a number of studies on writing instruction and essay scoring (Bell, 2001; Bouwer, 1998; Kong, 1998; Pelsmaekers, Braeck & Geluykens, 1998; Torrance & Bouayad-Agha, 2001). Researchers interested in the application of RST in academic writing, mainly have analysis of discourse patterns at the heart of their studies.

In 1995 O'Brien examined a number of student essays in detail by comparing native-speaker undergraduate students' performances in two written texts using relational analysis by Mann and Thompson (1988). RST was explained as a suitable framework for this kind of case-study since it facilitated the precise deliverance of the sources of incoherence. The results showed that the way students handle the given materials in memory results in a coherent or an incoherent coursework text.

Anders Bouwer (1998) applied RST to describe a prototype Intelligent Teaching System to teach Dutch university students how to use punctuation in writing. The system was designed in a way that grammatical aspects and rhetorical structure of the texts were affected by punctuation marks. It was aimed at providing appropriate feedback and possible correction of the students' answers and it seemed to be advantageous over textbook-based rules instruction since its model of instruction was more understood by the students. The scope of the study, however, was only limited to punctuation.

In a cross-cultural study, Kong (1998) compared the rhetorical structure of Chinese and English business letters. The main part of his study was dedicated to the comparison of Chinese non-native English writers' request letters and native English counterparts. The results demonstrated a clear discoursal transfer from Chinese to English version of the texts. Despite some weaknesses, the findings exhibited good evidence for the applicability of RST to text analysis.

Still in the area of second language writing, Pelsmaekers, et al. (1998) tried to offer a rich examination of subordination and connectives in texts from a 'rhetorical structure' point of view. In general, the authors wished to know how they can "make the concepts of nuclearity and rhetorical relations operational for a discussion of L2-writing" (p. 196). Their results led to some interesting suggestions for both teachers and learners. It was recommended by the authors to employ in written English more integration and explication, lack of which claimed to be a result of transfer from the spoken mode.

In one year, two major articles were released each pointing to a new direction in the application of RST. The first one by Bell (2001) used the theory as a medium of instruction for writing composition with a focus on the structure of

argumentative essays. And the second one by Torrance and Bouayad-Agha (2001) was an in-depth discussion into the cognitive processes underlying the production of a coherent text by twenty native English writers using think-aloud protocols. The writers of the latter study claimed that the rhetorical structure of a text was mostly generated during pre-writing and drafting phases but further research is needed to confirm their result.

Lastly, in an innovative study, Skoufaki (2009) applied RST to the detection of coherence errors made by a group of low-intermediate learners of English. Her purpose was to make a list out of the found errors and also to compare the findings of the study to the errors located by *Criterion*; an Automated Writing Evaluation (AWE) software. However, as a preliminary study, the project lacked inter- and intra-judge reliability. The data was also inadequate which made the replica of such an investigation necessary.

The small number of research projects named above clearly shows the need for more examination and discussion regarding the ways in which RST can contribute to the current area of L2 writing analysis. For this reason, we decided to apply RST in a new and under-studied context; that is, the Iranian EFL context. Incoherence is a common writing problem among Iranian learners of English the sources of which have not been thoroughly probed into. Therefore, any attempt to make these sources clear is of a great prominence. Furthermore, when it comes to RST, no study, to the best of our knowledge, has been conducted on coherence in writing in the Iranian context. This was the incentive behind the current study.

Method

Participants

In order to illuminate the coherence breaks in the writing of Iranian EFL learners, a group of language learners participated in our study and RST was chosen as the method of analysis. The participants were a group of 32 Iranian male EFL learners participating in two intact classes at the Iran Language Institute, Shiraz, Iran. The students had the same first language, Persian, with ages ranging from 21 to 28. They were placed in upper-intermediate level based on their language proficiency assessed by a placement test (held by the institute) prior to the instruction.

Materials

The materials of the study were 64 paper-based samples of writing produced by the participants of the study. The writings were in two different genres – 32 descriptive and 32 argumentative. The descriptive texts were written on the participants' favorite places and the argumentative ones on the preference of books over TV.

Data Collection

At the end of the term, the participants were given descriptive and argumentative writing assignments two weeks apart. The topics for the two compositions were chosen according to the two genres; the first one wanted the students to describe their favorite place and in the second one the participants had to argue whether reading books is better than watching TV. They were similar to the assignments they had been subjected to since the beginning of the semester. In fact, during the semester they had received instruction on how to write in the two genres. Different writing samples of the two genres had also been analyzed to make the participants better conceptualize the genres.

The participants did their assignments at home and brought them to the teacher. To make sure that the students would take the homework seriously and would do it themselves, the teacher informed them that they would receive feedback on their writing problems and that the teacher would help them in class to improve their writing. The writing samples were later copied to be used for the current research before they were given back to the students.

Data Analysis

In order to see what problems EFL learners had with regard to the text coherence, we analyzed 64 pieces of writing. The students' writings were rendered into the electronic versions using Microsoft Word. Then, the text files were imported to the computerized tool. The RSTTool345 for Windows was used to analyze the data. The first step was to segment a whole text at clause boundaries. The second step was structuring; that is, to connect the segments into a rhetorical structure tree. The third step was to assign predefined coherence relations. The tool also provided an editing option which allowed the user to add, delete, or rename the previously annotated relations. Finally, frequencies of different relations were counted using the automatic statistical analysis from the RSTTool.

In addition to the above quantitative analyses, a qualitative enquiry was also employed. In the qualitative analysis, the final rhetorical structure tree diagrams were examined carefully in order to specify and make a list of coherence errors which were commonly made by the participants. Such an analysis ultimately gave us some clues about the areas of difficulty for the Iranian EFL learners. Therefore, we tried to find the sources of coherence breaks and to explain the factors which may have led to the occurrence of these errors.

The above-mentioned steps of data analysis through RST are explained in more detail below.

Segmentation

Discourse segments should be non-overlapping spans of text. For the database in this study, a clausal definition of discourse segments was chosen. We adopt-

ed this method of segmenting discourse because it was easy to use. The segmentation procedure eventually provided the researchers with a total of 1634 clauses.

In segmenting the writings into their constituent clauses, a number of points were taken into consideration. First, we did not classify *and* as delimiting discourse segments if it was used to conjoin nouns or verbs in a conjoined noun phrase like *tea and juice and coffee* in example (1) or a conjoined verb phrase, like *laugh and play* in example (2):

(1) I can drink tea and juice and coffee.

(2) My cousins and I laugh and play all day and night.

Second, we classified periods, semicolons, and commas as delimiting discourse segments. However, in cases like example (3) in which they conjoin a complex noun phrase, commas were not classified as delimiting discourse segments:

(3) There are many different kinds of birds such as pigeon, parrot, partridge and nightingale.

The RSTTool is capable of automatically producing a segment boundary after each end of sentence. But the algorithm to spot ends of sentences is not perfect in cases of unusual punctuation. It does, however, handle phenomena such as "e.g.", "i.e.", "etc.", "...". When the system produces wrong segmentation, the user can delete or insert segmentation by clicking at the appropriate place in the text. In the same way, the user can segment the text at clause-unit points.

As a result, the typical process after importing a plain text file was to initially hit the *Sentences* button and then add in segment marks where it was fit, and finally delete bogus marks. A problem occurred here with embedded elements – where a rhetorically dependent stretch of text occurred within a main clause. For example, in the sentence below:

(4) My favorite place that I like it very much is Haftkhan restaurant complex.

Since the interface does not handle such cases, the solution was for us to move the embedded clause outside the enclosing sentence for analysis:

(5) My favorite place is Haftkhan restaurant complex that I like it very much.

Structuring

The second interface of RSTTool allowed the user to connect the segments into a rhetorical structure tree. This is called structuring. Initially, all segments were unconnected. The annotator could drag the mouse from one segment (the satellite) to another (the nucleus) to link them.

Relations

It was necessary to define a set of relations to describe the rhetorical structure of our database. A fixed relation set (ExtMT.rel) provided by the tool was chosen to link portions of texts to one another. The relations determine, for instance, if one span of text elaborates on the span which precedes it, or gives some justification for it, or expresses a conclusion which follows from it.

The system allows both mononuclear and also multi-nuclear RST relations. Mononuclear relations such as *Elaboration* and *Solutionhood* consist of a single nucleus and a single satellite. Multinuclear relations such as *Joint* and *Sequence* link together a set of entities of equal status (see Appendix C).

In total, five relations were added to the original set of relations. The first one, *conclusion*, was defined as indicating a relation in which the satellite was a reasoned judgment, inference, necessary consequence or final decision. The second one, in the form of a *question*, was delivered in two of the compositions and was immediately answered in one of them by the writer himself; therefore led to the formation of a third type of relation which was *answer*. The next relation was *wish*. This relation was seen once as a writer wished his previous statement would happen. Finally, some learners had written a number of noun phrases in the middle of their paragraphs used as *topic*s for the next part of their writings. Topics later were connected to their satellites applying elaboration relations.

Descriptive Statistics

A popup menu in the RSTTool provides some options (buttons). We used the Statistics button in the software to count the number of each relation along with the mean score for each. The calculation considers how a text element links to its parent, so a specific relation over three satellites is counted as three instances of that relation.

Annotators

The annotators for the database were two MA students of Teaching English as a Foreign Language (TEFL). One of the authors of this paper provided training for the other annotator. For training, the second annotator was explained the background of the project, discourse segmentation, coherence relations and how to recognize them, and the way to use the annotation tool. He was also given example texts to annotate. After completing the training, both annotators worked on the whole data.

Reliability

The definitions of relations place constraints on the kinds of meanings that the nucleus and satellite may express and the consequences of recognizing the rela-

tion (the effect). These constraints and conditions of the relations are often sufficient to produce good inter-coding reliability. Nevertheless, accounting for the reliability of our analysis, the two annotators worked on the same data at the same time. Their results were then compared to make sure they matched together.

For discourse segmentation, a pilot study on 10 texts was implemented. The agreement was never below 95% as calculated by *number of common segments* divided by the *number of common segments + number of differing segments*. Therefore, all the 64 texts were segmented by the two annotators together, resulting in segmentations that both annotators could agree on.

To determine inter-annotator agreement for coherence relations' annotation, we also calculated kappa statistics (Carletta, 1996). The kappa coefficient (K) measures agreement among a set of annotators making category judgments:

$$K = \frac{P(A) - P(E)}{1 - P(E)}$$

P(A) is the proportion of times that the annotators agree and P(E) is the proportion of times that we would expect them to agree by chance. The resulting kappa coefficient ranges from zero to one. K > .80 is considered a good reliability while .67 < K < .80 allows tentative conclusions to be drawn.

For all the annotations of the 64 texts in our study, the agreement was .82, per chance agreement was .24, and kappa was .77 which showed an acceptable level of reliability.

Results

Descriptive Statistics

A total number of 2069 coherence relations were identified with 950 (45.9%) pertaining to the descriptive genre (a favorite place) and 1119 (54.1%) belonging to the argumentative genre (TV or books). The most frequent coherence relation in the whole data was *joint* (443) with the percentage of 21.4. The next most frequent type was *conjunction*, constituting 15.4% of the whole relations. In addition to these two, *elaboration* (14.2%), *list* (8.3%), and *contrast* (5.7%) were the most frequent rhetorical relations. As for the least frequent, *wish* (<0.1%), *volitional result* (<0.1%), *answer* (<0.1%), *question* (0.1%), *interpretation* (0.2%), *summary* (0.3%), *motivation* (0.3%), *solutionhood* (0.4%), and *disjunction* (0.4%) stood out from the total of 31 types of relations.

Table 2 presents the frequency and percentage of coherence relations across the two different genres. According to the Table, the most frequent relation used by participants for description of their favorite place pertains to *joint* (24.9%). The same is true for the other genre; in the argumentative texts *joint* was used more than any other relation (18.4%).

Comparing the two genres, *joint* (24.9%), *elaboration* (18.8%), *conjunction* (12.2%), and *list* (10.5%) relations were used more frequently in the descriptive writings. The same was true for the argumentative writings. However, in the argumentative genre, contrast (10.5%) was also very frequent (the third most frequent) which is expected given the nature of argumentative genre which demands contrasting ideas.

Table 2.
Frequency and percentage of coherence relations in descriptive and argumentative texts

Type of Relation	Descriptive Texts		Argumentative Texts	
	Frequency	Percent	Frequency	Percent
Joint	237	24.9	206	18.4
Elaboration	179	18.8	115	10.3
Conjunction	116	12.2	203	18.1
List	100	10.5	71	6.3
Non.V.Cause	55	5.8	48	4.3
Justify	38	4.0	35	3.1
Sequence	36	3.8	0	0.0
Evaluation	33	3.5	11	1.0
Circumstance	31	3.3	39	3.5
Background	26	2.7	10	0.9
Non.V.Result	13	1.4	51	4.6
Preparation	11	1.2	29	2.6
Purpose	9	0.9	20	1.8
Evidence	8	0.8	10	0.9
Restatement	8	0.8	7	0.6
Topic	7	0.7	17	1.5
Conclusion	6	0.6	12	1.1
Enablement	6	0.6	11	1.0
Antithesis	5	0.5	29	2.6
Concession	5	0.5	20	1.8
Condition	5	0.5	35	3.1
Summary	5	0.5	2	0.2
Motivation	4	0.4	3	0.3
Disjunction	2	0.2	6	0.5
Interpretation	2	0.2	2	0.2
Solutionhood	1	0.1	7	0.6
V.Result	1	0.1	0	0.0
Wish	1	0.1	0	0.0
Contrast	0	0.0	117	10.5
Question	0	0.0	2	0.2
Answer	0	0.0	1	0.1
Total	950	100.0	1119	100.0

However, the least frequent relations in the two genres were not that much similar. In the descriptive data, on the one hand, *disjunction* (0.2%), *interpretation* (0.2%), *solutionhood* (0.1%), and *volitional result* (0.1%) were rarely seen. In the argumentative data, on the other hand, *interpretation* (0.2%) and *summary* (0.2%) were among the infrequent relations; along with *question* (0.2%) and *answer* (0.1%) which were both absent from the first set of data. Overall, the second set of data demonstrated more usage of coherence relations.

If we look at Table 2 once again, we recognize that in both genres *conjunction* and *elaboration* are associated with high percentages of occurrence; 12.2%

and 18.8% in descriptive texts and 18.1% and 10.3% in argumentative ones. The difference is in the *contrast* relation which has a high percentage of 10.5 in the argumentative writings (about the difference between TV and books) but is not being used in the descriptive writings at all.

The relation which was absent in argumentative texts was *sequence.* In descriptive texts, on the other hand, it could be observed more or less regularly (3.8%). Two of the most important rhetorical connections in texts were *result* and *conclusion* which were more seen in the argumentative than in the descriptive writings. The presentational relations of *antithesis* and *concession* in argumentative writings also outnumbered their counterparts in descriptive manuscripts.

In what follows, the results of the study are presented in line with the research questions.

Coherence errors in the writings of Iranian EFL learners

Table 3 summarizes the coherence breaks indicated by the main abnormalities found in the RST diagrams. The first significant error (*irrelevant content*) which is more associated with the text content was when parts of a text talked about something significantly far from the main content of the whole text.

Table 3.
Frequency and percentage of diagram abnormalities

Diagram Abnormalities	Frequency	Percent
Irrelevant Content	36	19.1%
Violation of Completedness	34	18.5%
Violation of Connectedness	31	16.5%
Incorrect Place	25	13.3%
Incorrect Relation	21	11.2%
Crossed Dependency	20	10.6%
Scattered Units	17	9.0%
Topic	4	2.1%
Total	188	100%

Consider the following examples:

(1) My uncle is living in Dubai. He really wants to return to Shiraz but he can't. Because he is working in Dubai.

(2) But now majority of people have smart phones and they can do a lot of jobs with it.

(3) I'm writing my second writing for my dear teacher and friend.

The first example was in a writing talking about the writer's favorite place and it was not related to the rest of the text. In the second example, the writer

talked about the advantages of smart phones while he was discussing the advantages and disadvantages of watching TV and its difference with reading books. Example (3) is one of a relatively high number of clauses which did not add any new information to the topic of the compositions. Other examples of this type included clauses such as "OK", "finished", "the end", etc., which only present some unnecessary content.

The second and third coherence errors in the list were closely related to the constraints which were set by Mann and Thompson (1988). *Violation of connectedness* happens when a text span does not have any relation with other text spans. As a result, the entire text does not fit into a single schema which in turn leads to the *violation of completedness* which constituted 18% of the total errors in this study. Further analysis indicated that this type of error evidently appeared in more than half of the writing samples (53%) written by the Iranian participants.

The fourth type of errors happened when a specific unit in the form of a single clause or a text span appeared in a place where it was not allowed to. Figure 1 clearly shows an instance of such a coherence break.
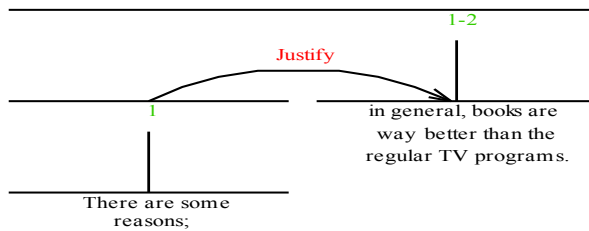

Figure 1.
Example of an incorrect place error

If unit 1 was a background for the content of 1-2; then, the order would be true. However, since the writer is going to justify unit 1-2 by presenting a number of reasons, the order of the above diagram will be corrected when unit 1 comes after unit 1-2.

Incorrect relations constituted the fifth type of coherence breaks. These errors occurred when a writer related two parts of the text with a wrong relation. In the example which is shown below (Figure 2), the writer has related unit 23 to 22 with a cause relation. However, the correct relation is *elaboration*.

The sixth error took place when a unit intruded in the sub-diagram which had already been formed. As an example, consider the following text:

"11. [In the garden we have a dog,] 12. [and its name is Peter.] 13. [It's a genteel dog] 14. [and never attack to us.] 15. [The garden was very big] 16. [and has one corner for 5 persons to sleep.] 17. [At fall when we go to garden] 18.

[all of trees' leaves are yellow or orange.] 19. [But in the winter twin of the dog died.]"

Unit 13-14 is an elaboration for unit 11-12 which is a nucleus. Unit 19 also is another satellite which elaborates more about the content in 13-14; but when it wants to connect to it, it crosses over the text span number 15-18 in the middle and constitutes a *crossed dependency* coherence error.
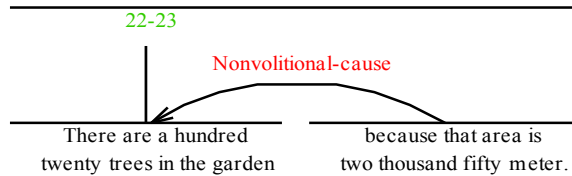

Figure 2.
Example of an incorrect relation error

Similar to the *incorrect place* coherence error, was the *scattered units* error. In this type of breakdown, sub-parts of two or more spans of text intermingled and established a scattered set of diagrams. Figure 3 is an example of the scattered units coherence error.

The text in Figure 3 reads as "11. [The news and information that you get from TV is update and newer than the books.] 12. [Overall reading a book is boring for me.] 13. [I do not to enjoy reading a book] 14. [and I do not like it very much.] 15. [TV has different programs for different styles and years.] 16. [We can see movies, sports, news, and etc in TV,] 17. [but books has many papers] 18. [that most of them do not have beautiful pictures to attract people.]" Units 11-16 and 12-18 represent two conjunction relations which are themselves sub-parts of a contrast relation (unit 11-18).
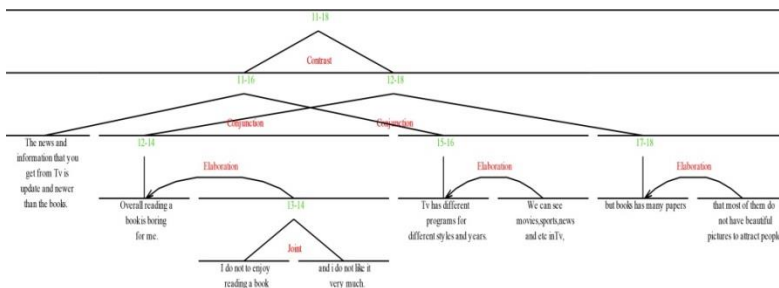

Figure 3.
Example of scattered units error

Finally, the last error (*topic*) was seen in only four manuscripts. Three of them were in the writings on the differences between TV and books; and they came from the students' tendency to put a topic before a list of advantages and/or disadvantages of each instrument. The *topic* error from the first writing assignment appeared in a student's description of a restaurant complex in which he separated the explanation of each section of the restaurant with a single topic (Figure 4).
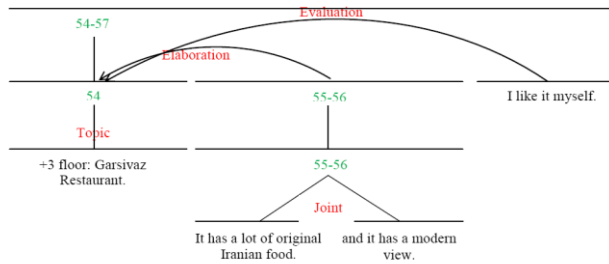

Figure 4.
Example of a topic error

In addition to the aforementioned coherence errors, the authors also found 15 instances of incomprehensible content. They were excluded from the calculations because they were impossible to be explained or understood and they could not be linked to other sub-diagrams. In other words, the researchers could not make sense of the utterances produced, for example because the utterance produced was structurally too problematic and vague to be easily interpreted in relation to a specific diagram. The researchers also discovered that out of 64 writing samples, eight were free of errors.

Coherence errors in descriptive vs. argumentative genres

In order to see if there was any relationship between the given writing genres and the coherence errors, different types of errors in each genre were examined separately and then compared. Table 4 presents the results.

The analysis showed that types of coherence errors were the same in both descriptive and argumentative writing samples. All eight diagram abnormalities were found in both genres.

The difference, however, was in the number of each type of error in each genre. It is necessary to point out that the writings were approximately the same in terms of their length; descriptive samples consisted of 803 clauses with the mean score of 25.09 and argumentative samples consisted of 831 clauses with the mean score of 25.96.

Table 4.
Frequency and percentage of coherence errors across the two genres

	Descriptive		Argumentative	
Coherence Error	Frequency	Percent	Frequency	Percent
Irrelevant Content	14	15.5	22	22.4
Violation of Completedness	14	15.5	20	20.4
Violation of Connectedness	9	10.0	22	22.4
Incorrect Place	14	15.5	11	11.2
Incorrect Relation	15	16.6	6	6.1
Crossed Dependency	15	16.6	5	5.1
Scattered Units	8	8.9	9	9.2
Topic	1	1.1	3	3.1
Total	90	100	98	100

According to Table 4, the most frequent types of coherence errors in argumentative samples were "violation of connectedness" and "irrelevant content" (each 22.4%). However, in descriptive samples, "incorrect relation" and "crossed dependency" breakdowns had the highest percentage of total errors (each 16.6%). This is while one of the lowest percentages of errors in argumentative texts belonged to "crossed dependency" (5.1%).

The argumentative genre also had a significantly higher amount of "violation of completedness" (20.4%). This amount for the descriptive genre was equal to 15.5% which is also very high. The error with the least percentage of occurrence was "topic" in both samples (1.1% for descriptive and 3.1% for argumentative).

Discussion

Application of RST in distinguishing coherence errors of Iranian EFL learners resulted in a number of findings that are discussed below.

Given that subparts of texts are related to one another in various ways, understanding these relations is an important part of understanding a text as a whole. The results of the current study showed that some coherence relations appeared more than the others in the written texts provided by our participants.

More than any relation, the participants made use of *joint* to connect different parts of their texts together. This overuse of *joint* relation was intelligible while the researcher was analyzing the data, primarily because of the high observation of the discourse marker "and". The connector not only was used to join or conjoin the clauses together; but also was applied to build coherence relations of *contrast*, *result*, *elaboration*, *evaluation*, *conclusion*, and even *antithesis*. This diversity in the utilization of discourse markers for any relation some-

times leads to the establishment of incorrect relation and as a result to the in-coherency in texts. The issue will be talked about later.

The high percentage of *conjunction* relation also confirms that the students were obliging to connect text segments to each other. But it did not necessarily guarantee the coherence of their texts. The researchers did not expect *summary*, *restatement* and *conclusion* relations to be used rarely, because part of the coherence of a text is its writer's ability to sum up the writing by giving a conclusion or restating the previous contents. Iranian EFL learners, however, were weak in this respect.

In addition to these, the background relation also had a low frequency. This is another sign of probable incoherence in a text. Background helps the reader to better communicate with the text. Beside *joint* and *conjunction* relations which were observed a lot in both writing samples, in descriptive writing the participants elaborated more but in the argumentative assignment they mostly contrasted two contents as they elaborated on each one. The amount of *list* re-lation use indicates that the participants were comfortable presenting what they wanted to say in the form of lists. They also provided more *result*s for what they wanted to argue in the argumentative writings.

On the one hand, *evaluation* was more observable in descriptive texts which showed that the writers wanted the reader to recognize the value that they as-signed to their contents. On the other hand, argumentative texts included more instances of *antithesis* and *concession*; this way the reader's positive regard for the writer's idea would increase and result in the easier acceptance of the ar-gument.

Considering the above-mentioned differences between relations found in the two genres, the data truly reveals that the participants have written two distinctive compositions in terms of genre.

The most important objective of this research study was to find the coher-ence errors in EFL learners' writings. The results demonstrated eight different error breaks. Interestingly, the most common coherence error discovered in our data was the presence of those contents which were totally irrelevant to the content of the whole text.

There were also 31 instances of unconnected units. It is clear that any single writer should prevent distracting the reader from the intended message by cre-ating a continuity between one part of the text and another, the quality which was extremely violated here. As mentioned before, under RST a text is consid-ered coherent if it is possible to construct a single "rhetorical structure" cover-ing all the surface units of the text. The two errors mentioned above, however, resulted in 53% of the texts to become incoherent because they did not con-form to the completedness constraint.

Delivering a textual segment in inappropriate place, mixing textual seg-ments of distinct text spans with each other and crossed dependency were the errors which mainly stem from students' willingness to apply inductive and

somehow choppy content order. This may mean that in writing a composition, EFL students did not decide on what they wanted to write beforehand (no writing outline) and also they did not organize their ideas. They started to write from the outset and continued to the point when they felt they had nothing more to say.

All *incorrect relation* errors were those which were wrongly signaled. For instance, a student wanted to bring out a *cause* for a statement but instead of *because* he used *and*; hence, he produced a *joint* relation. The use of transitional devices such as *yet* – to introduce a contrast relation – or *therefore* – to set up a conclusion – helps to accomplish better coherence in texts. Mann and Thompson (1987) maintain that relations between satellite and nucleus need not to be made explicit by linguistic markers. However, they also point out that there are connectors that can be used to make a relation specific.

Finally, errors regarding the appearance of *topics* in the middle of the paragraphs were very few but emphasized that some students were not aware of the structural organization of a good writing.

Textual organization cannot be studied without consideration of the variability between text genres. In case of our study, the descriptive writing samples were organized around a central theme (the writer's favorite place); while the argumentative samples were organized around a central claim to be argued for (the priority of books over TV).

Based on our results, descriptive samples demonstrated coherence more than argumentative ones. So we can say that the overall coherence of a text depends on the presence of a conventional scheme (recognized as genre) which fulfils a particular communicative purpose. Overall, participants had more coherence errors in argumentative writing. This may be due to the fact that the complexity of argumentative essays is greater than that of descriptive. According to Mosenthal (1985), the difference between these two genres is highly noticeable, in a way that in a continuum, he places descriptive texts at one end and argumentative texts at the other.

The students, moreover, produced more irrelevant contents and more unconnected units in argumentative texts. But they displayed deductive content order in this genre more than in the descriptive one as if they were more aware of the order in which they wanted to present their argument in the texts about *TV or books*. They mostly had a topic sentence which said what their idea was and then they developed it. In contrast, in descriptive paragraphs they presented their ideas more inductively.

Comparing this study with that of Skoufaki (2009), the researcher can conclude that in general more coherence errors in terms of type and frequency were found in writing essays written by Iranian EFL learners. Participants of that study who were Taiwanese EFL learners produced *irrelevant content*, *incomprehensible content*, *self-sufficiency*, *crossed dependency*, *unexpected relation*, and *incorrect place* coherence errors. Similar to the present study, inductive content order was detected resulting from a high frequency in "dangling struc-

ture" RST diagrams. Skoufaki aimed to see if there exists a relationship between coherence errors and topic and not genre of writings. Her quantitative analysis indicated that all coherence errors located in the data varied depending on the topic.

Turning to studies which have examined the effect of genre on writing, we realize that relatively little work has been done in this area. For example, Crowhurst (1987) found that narratives had a higher use of cohesive ties than did argumentative writings. Knudsen (1992) studied the difference in coherence in student writing between a complex narrative task and a simple one, and came up with mixed results. But theorists, particularly in Australia, have written about the importance of genre in instruction (e.g., Grabe & Kaplan, 1996) asserting that language use is determined by social context and that genre can be a relevant method of varying student's language output. Their claims have been confirmed by Strid (1998) who suggests that genre affects coherence. Therefore, theory holds that differences exist between genres, and one of these differences may be the level of coherence found in descriptive and argumentative writings.

Conclusion

When one talks about coherence, the focus primarily is on the links between sentences and paragraphs and the logical connection between all the units in a text. Coherence refers to the less tangible ways of connecting discourse, properties such as rhetorical relations which are covert and which reside in the way people interpret texts rather than in the texts themselves (Yule, 1996).

Despite its centrality to language interpretation, discourse coherence is a subject that has only been sketchily addressed. Therefore, we intended to bring to light – by way of analyzing a number of writing compositions – the necessity and functionality of coherence studies. Incoherence is a continuing problem in the students' writing which can become a major obstacle to their success. The reason might be that coherence errors in comparison to grammatical errors are more difficult to deal with as they involve strings of sentences or paragraphs. In addition, judgments about coherence also include a reader, a writer, the stock of world knowledge which they share and a communication situation.

Some practitioners believe that the problems with evaluating coherence stem from the fact that it is by nature subjective. However, the previously successful applications of RST for analyzing discourse coherence have proven that it is possible to detect coherence errors in a relatively reliable manner. While the coherence of a text can be measured, it can often involve considerable effort.

Our attempts to unveil the coherence errors of EFL learners' written discourse brought about some stimulating results. One of the outstanding outcomes of this study was that almost half of the learners' writings were subject to the *incompletedness* error. Even if it was the only error they had in their writ-

ing, it made the whole text incoherent. We know that a piece of text is not coherent unless all of its constituent parts are placed under a single text span. In comparison, if a text includes only one *incorrect relation* error, the incoherence is limited to a small part of that text and not to all of it.

Thereupon, students must make sure that the content portrayed in their writings holds together into a coherent whole, that all the sentences, clauses, and phrases contribute to the meaning of the whole piece, that the text is organized and it makes sense moving from one idea to another.

Another significant outcome of this study was the coherence difference between descriptive and argumentative texts. The descriptive compositions appeared to be more coherent than the argumentative ones. It is obvious from what we observed that the coherence structure of an individual text is a description of its coherence relations; and that the organization of these relations varies between different genres. In other words, coherence was susceptible to genre. The difference could stem from the difficulty of writing in one genre than the other; that is, the higher cognitive load involved in argumentative writing may lead to less successful production on the part of language learners. However, the difference may also be partly due to the topic of the writings. It may be the case that the topic for one assignment was easier than the other one or that it was more tangible for the participants.

Obviously the effect of each type of error is different from the other. For instance, consider the error of *scattered units*. Sometimes two or three *irrelevant contents* cannot affect the coherence to the extent that two spans of text with *scattered units* can. Therefore, it is possible that the coherence errors found in descriptive genre had produced higher levels of incoherency than the ones found in argumentative genre; or vice versa.

RST serves as an appropriate tool to conduct research in different fields, especially in writing instruction and essay scoring. For all educators, RST is a credible and effective method of text-based research in writing. This study is probably the first one in the Iranian context which represents a means to provide insights into student compositions from the local structure (sentence coherence) issues to the global structure (paragraph coherence) ones. Most of the previous studies used cohesion as a yardstick to measure coherence. But cohesion is only one part of probing into the sources of incoherence, and studying coherence from the global viewpoint can lead to useful results and lighten the possible sources of coherence.

Teachers today should be aware that a good prose must communicate effectively and meaningfully by means of its unity, coherence, and emphasis on its prospective readers. Training in RST analysis may provide a way to raise teachers' awareness of coherence and coherence breaks. Results of such analyses may guide teachers in the assessment of coherence. RST also provides language learners with a general strategy to analyze their own text in a conceptual and tangible way.

Not everything described by discourse analysts may have any immediate application in language teaching but the more one can learn from text organization – from small units to large – the more likely it is to create authentic materials and activities for the classroom.

Our study depicts a route for several future research projects. Since this study is the first one in Iranian context, replications of it seems to be necessary. Research studies can be constructed with different participants from different levels of language proficiency. Studies could also be handled analyzing the relation between writing scores and coherence errors. Other genre categories can also be examined in terms of written coherence errors. The effect of writing topics can also be added to the role of genre in written coherence. It would also be the subject of another study to see if marked and unmarked relations affect text coherence. Finally, a full study is definitely required with a larger corpus of student essays to see whether the findings can be generalized to other samples.

References

Abu Shawish, J. (2015). Discoursal coherence breaks experienced by AL-Quds Open University EFL Majors. *International Journal on Studies in English Language and Literature (IJSELL), 3*(8), 1-15.

Ballard, D., Conrad, R. & Longacre, R. (1971). The deep and surface grammar of interclausal relations. *Foundations of Language*, *4*, 70–118.

Bateman, J., & Rondhuis, K. J. (1997). Coherence relations: Towards a general specification. *Discourse Processes, 24*, 3–49.

Beekman, J., & Callow, J. (1974). *Translating the word of God*. Grand Rapids, MI: Zondervan Publishing House.

Bell, M. A. (2001). Online notes on the structure of argument essays. Retrieved from http://www.users.bigpond.com/m487396/Argument/essay_notes.htm.

Berman, L., & Slobin, D. I. (1994). *Relating events in narrative: Across linguistic developmental study*. Hillsdale, NJ: Lowrence Erlbaum Associates.

Bouwer, A. (1998). An ITS for Dutch punctuation. *Intelligent Tutoring Systems*, *1452*, 224–233.

Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, *22*(2), 249–254.

Carlson, L., Marcu, D., & Okurowski, M. E. (2002). *RST discourse Treebank*, LDC2002T07 [Corpus]. Philadelphia, PA: Linguistic Data Consortium.

Cook, G. (1989). *Discourse*. OUP: Oxford.

Crowhurst, M. (1987). Cohesion in argument and narration at three grade levels. *Research in the Teaching of English*, *21*, 185–201.

Cui, S. (1986). *A comparison of English and Chinese expository rhetorical structures*, (Unpublished master's thesis). UCLA, USA.

Egg, M., & Redeker, G. (2006). Underspecified discourse representation. In A. Benz, & P. Kühnlein (Eds.), *Constraints in discourse* (pp. 137-163). Amsterdam: Benjamins. Available from http://www.let.rug.nl/~egg/Papiere/eggredeker06a.pdf.

Fillmore, C. (1974). Pragmatics and the description of discourse. In C. Fillmore, G. Lakoff, & R. Lakoff (Eds.), *Berkeley studies in syntax and semantics*: Vol. 1 (pp. 1–21). University of California, Berkeley, California.

Fitzgerald, J., & Spiegel, D. L. (1990). Children's writing. *Research in the Teaching of English*, *20*, 263–280.

Fox, B. A. (1987). *Discourse structure and anaphora: Written and conversational English*. Cambridge: Cambridge University Press.

Grabe, W., & Kaplan, R. (1996). *Theory and practice of writing*. London and New York: Longman.

Grimes, J. (1975). *The thread of discourse*. The Hague: Mouton.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, *12*(3), 175–204.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, *21*(2), 203–225.

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

Hobbs, J. R. (1976). *A computational approach to discourse analysis* (Research Rep. No. 76-2). New York: City University of New York.

Hobbs, J. R. (1985). *On the coherence and structure of discourse* (Tech. Rep. No. CSLI-85-37). Stanford: Stanford University, Center for the Study of Language and Information.

Hovy, E. H. (1990). Unresolved issues in paragraph planning. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation* (pp. 17–45). London: Academic Press.

Knudsen, R. (1992). Effects of task complexity on narrative writing. *Journal of Research and Development in Education*, *26*, 7–14.

Kong, K. C. C. (1998). Are simple business request letters really simple? A comparison of Chinese and English business request letters. *Text*, *18*(1), 103–141.

Lascarides, A., & Asher, N. (1991). Discourse relations and defeasible knowledge. *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics* (pp. 55–63). Morristown, NJ: Association for Computational Linguistics.

Lascarides, A., & Asher, N. (1993). Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, *16*(5), 437–493.

Longacre, R. E. (1977). The paragraph as a grammatical unit. *Symposium on Discourse*. UCLA.

Longacre, R. E. (1983). *The grammar of discourse*: *Notional and surface structures*. New York: Plenum Press.

Mann, W. C. (2005). RST website. Retrieved from http://www.sfu.ca/rst.

Mann, W. C., & Thompson, S. A. (1987). *Rhetorical structure theory: A theory of text organization* (Tech. Rep. No. RS-87-190). Los Angeles, CA: USC/Information Sciences Institute.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, *8*(3), 243–281.

Martin, J. R. (1992). *English text: Systems and structure*. Amsterdam: John Benjamins.

Mosenthal, P. B. (1985). Defining the expository discourse continuum. Towards a taxonomy of expository text types. *Poetics*, *14*, 387–414.

O'Brien, T. (1995). Rhetorical structure analysis and the case of the inaccurate, incoherent source-hopper. *Applied Linguistics*, *16*(4), 442–482.

Pelsmaekers, K., Braecke, C., & Geluykens, R. (1998). Rhetorical relations and subordination in L2 writing. In A. Sánchez-Macarro, & R. Carter (Eds.), *Linguistic choice across genres: Variation in spoken and written English* (pp. 191–213). Amsterdam and Philadelphia, PA: John Benjamins.

Péry-Woodley, M. P. (1998). Signaling in written text: A corpus-based approach. In M. Stede, L. Wanner, & E. Hovy (Eds.), *Proceedings of COLING-ACL Workshop on Discourse Relations and Discourse Markers* (pp. 79–85). Montréal, Canada.

Péry-Woodley, M. P. (2001). Modes d'organisation et de signalisation dans des texts procéduraux. *Languages, 141*, 28–46.

Polanyi, L. (1985). *The structure of discourse.* Norwood: Ablex.

Reichman, R. (1978). Conversational coherency. *Cognitive Science*, *3*, 283–327.

Sanders, T. J. M., Spooren, W. P. M., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, *15*, 1–35.

Scott, D., Delin, J., & Hartley, A. (1999). Identifying congruent pragmatic relations in procedural texts. *Languages in Contrast, 1*(1), 45–82.

Skoufaki, S. (2009). An exploratory application of rhetorical structure theory to detect coherence errors in L2 English writing: Possible implications for automated writing evaluation software. *Computational Linguistics and Chinese Language Processing, 14*(2), 181–204.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition* (2nd ed.). Oxford: Blackwell.

Strid, J. E. (1998). *Coherence in the narrative and persuasive writing of adolescents*. Talk presented at Illinois TESOL, Chicago, Il.

Taboada, M. (2001). *Collaborating through talk: The interactive construction of task oriented dialogue in English and Spanish*, (Unpublished doctoral dissertation). Universidad Complutense, Madrid.

Taboada, M. (2004a). *Building coherence and cohesion: Task-oriented dialogue in English and Spanish*. Amsterdam and Philadelphia, PA: John Benjamins.

Taboada, M. (2004b). Rhetorical relations in dialogue: A contrastive study. In C. L. Moder, & A. Martinovic-Zic (Eds.), *Discourse across languages and cultures* (pp. 75–97). Amsterdam and Philadelphia, PA: John Benjamins.

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

Taboada, M., & Mann, W. C. (2006). Rhetorical structure theory: Looking back and moving ahead. *Discourse Studies*, *8*(3), 423–459.

Torrance, M., & Bouayad-Agha, N. (2001). Rhetorical structure analysis as a method for understanding writing processes. In L. Degand, Y. Bestgen, W. Spooren, & L. van Waes (Eds.), *Multidisciplinary approaches to discourse*. Amsterdam: Nodus.

van Dijk, T. A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York: Academic Press.

Yule, G. (1996). *The study of language* (2nd ed.). Cambridge: Cambridge University Press.

Appendix A

Definitions of presentational relations;
taken from Mann (2005)

Relation Name	Constraints on either S or N individually	Constraints on N + S	Intention of W
Antithesis	on N: W has positive regard for N	N and S are in contrast (see the Contrast relation); because of the incompatibility that arises from the contrast, one cannot have positive regard for both of those situations; comprehending S and the incompatibility between the situations increases R's positive regard for N	R's positive regard for N is increased
Background	on N: R won't comprehend N sufficiently before reading text of S	S increases the ability of R to comprehend an element in N	R's ability to comprehend N increases
Concession	on N: W has positive regard for N; on S: W is not claiming that S does not hold;	W acknowledges a potential or apparent incompatibility between N and S; recognizing the compatibility between N and S increases R's positive regard for N	R's positive regard for N is increased
Enablement	on N: presents an action by R (including accepting an offer), unrealized with respect to the context of N	R comprehending S increases R's potential ability to perform the action in N	R's potential ability to perform the action in N increases
Evidence	on N: R might not believe N to a degree satisfactory to W; on S: R believes S or will find it credible	R's comprehending S increases R's belief of N	R's belief of N is increased
Justify	None	R's comprehending S increases R's readiness to accept W's right to present N	R's readiness to accept W's right to present N is increased
Motivation	on N: N is an action in which R is the actor (including accepting an offer), unrealized with respect to the context of N	Comprehending S increases R's desire to perform action in N	R's desire to perform action in N is increased
Preparation	none	S precedes N in the text; S tends to make R more ready, interested or oriented for reading N	R is more ready, interested or oriented for reading N
Restatement	none	on N + S: S restates N, where S and N are of comparable bulk; N is more central to W's purposes than S is.	R recognizes S as a restatement of N
Summary	on N: N must be more than one unit	S presents a restatement of the content of N, that is shorter in bulk	R recognizes S as a shorter restatement of N

Appendix B

Definitions of subject matter relations; taken from Mann (2005)

Relation Name	Constraints on either S or N individually	Constraints on N + S	Intention of W
Circumstance	on S: S is not unrealized	S sets a framework in the subject matter within which R is intended to interpret N	R recognizes that S provides the framework for interpreting N
Condition	on S: S presents a hypothetical, future, or otherwise unrealized situation (relative to the situational context of S)	Realization of N depends on realization of S	R recognizes how the realization of N depends on the realization of S
Elaboration	none	S presents additional detail about the situation or some element of subject matter which is presented in N or inferentially accessible in N in one or more of the ways listed below. In the list, if N presents the first member of any pair, then S includes the second: ⬜ set :: member ⬜ abstraction :: instance ⬜ whole :: part ⬜ process :: step ⬜ object :: attribute ⬜ generalization :: specific	R recognizes S as providing additional detail for N. R identifies the element of subject matter for which detail is provided.
Evaluation	none	on N + S: S relates N to degree of W's positive regard toward N.	R recognizes that S assesses N and recognizes the value it assigns
Interpretation	none	on N + S: S relates N to a framework of ideas not involved in N itself and not concerned with W's positive regard	R recognizes that S relates N to a framework of ideas not involved in the knowledge presented in N itself
Means	on N: an activity	S presents a method or instrument which tends to make realization of N more likely	R recognizes that the method or instrument in S tends to make realization of N more likely
Non-volitional Cause	on N: N is not a volitional action	S, by means other than motivating a volitional action, caused N; without the presentation of S, R might not know the particular cause of the situation; a presentation of N is more central than S to W's purposes in putting forth the N-S combination.	R recognizes S as a cause of N
Non-volitional Result	on S: S is not a volitional action	N caused S; presentation of N is more central to W's purposes in putting forth the N-S combination than is the presentation of S.	R recognizes that N could have caused the situation in S
Otherwise	on N: N is an unrealized situation; on S: S is an unrealized situation	realization of N prevents realization of S	R recognizes the dependency relation of prevention between the realization of N and the realization of S
Purpose	on N: N is an activity; on S: S is a situation that is unrealized	S is to be realized through the activity in N	R recognizes that the activity in N is initiated in order to realize S
Solutionhood	on S: S presents a problem	N is a solution to the problem presented in S;	R recognizes N as a solution to the problem presented in S
Unconditional	on S: S conceivably could affect the realization of N	N does not depend on S	R recognizes that N does not depend on S
Unless	none	S affects the realization of N; N is realized provided that S is not realized	R recognizes that N is realized provided that S is not realized
Volitional Cause	on N: N is a volitional action or else a situation that could have arisen from a volitional action	S could have caused the agent of the volitional action in N to perform that action; without the presentation of S, R might not regard the action as motivated or know the particular motivation; N is more central to W's purposes in putting forth the N-S combination than S is.	R recognizes S as a cause for the volitional action in N
Volitional Result	on S: S is a volitional action or a situation that could have arisen from a volitional action	N could have caused S; presentation of N is more central to W's purposes than is presentation of S;	R recognizes that N could be a cause for the action or situation in S

Appendix C

Definitions of multinuclear relations; taken from Mann (2005)

Relation Name	Constraints on each pair of N	Intention of W
Conjunction	The items are conjoined to form a unit in which each item plays a comparable role	R recognizes that the linked items are conjoined
Contrast	No more than two nuclei; the situations in these two nuclei are (a) comprehended as the same in many respects (b) comprehended as differing in a few respects and (c) compared with respect to one or more of these differences	R recognizes the comparability and the difference(s) yielded by the comparison is being made
Disjunction	An item presents a (not necessarily exclusive) alternative for the other(s)	R recognizes that the linked items are alternatives
Joint	None	none
List	An item comparable to others linked to it by the List relation	R recognizes the comparability of linked items
Multinuclear Restatement	An item is primarily a reexpression of one linked to it; the items are of comparable importance to the purposes of W	R recognizes the reexpression by the linked items
Sequence	There is a succession relationship between the situations in the nuclei	R recognizes the succession relationships among the nuclei.