

# Devising a Discourse Error Tagging System for an English Learner Corpus

Sophia Skoufaki  
University of Essex  
United Kingdom  
sskouf@essex.ac.uk

## Abstract

This paper reports on-going research aiming to result in the tagging of discourse errors in the Language Training and Teaching Center English Learner Corpus, a Taiwanese learner corpus of English constructed by the Graduate Institute of Linguistics and the Department of Foreign Languages and Literatures at National Taiwan University, the Department of English at National Chengchi University, and the LTTC, a language testing company. This study was motivated by the paucity of discourse error tags in other learner corpora (Díaz-Negrillo & Fernández-Domínguez, 2006) and by a desire to examine discourse errors made by intermediate-level Taiwanese learners of English.

In particular, this paper describes the pilot application of Rhetorical Structure Theory (RST) (Mann & Thompson, 1988) to the analysis of coherence relations in 45 written samples from the aforementioned corpus. This pilot study examines whether an RST coherence analysis can lead to tangible evidence for the existence of coherence breaks in the writing of the learners. The rationale of this study is that RST text analyses which violate some of the rules of RST diagram formation will point to coherence errors.

This study is useful for the discourse error tagging of the corpus because after listing the RST diagram abnormalities which characterize each kind of coherence error, the discourse error tagging will be done in a principled way. Moreover, the frequency of coherence error types offers some preliminary indications about which error types warrant further investigation through the analysis of more writing samples. The ultimate goal of this investigation is to identify the location of the most frequent coherence error types, so that the corpus can be tagged for them.

No validity and reliability tests have been conducted since this work was at an initial stage and only aimed to find out whether this method would be useful. Therefore, results are only preliminary. Results indicate that this approach can help locate coherence errors more reliably but that the tagger's intuition is needed for some coherence errors and, of course, it is also needed whenever one "problem" in an RST diagram points to more than one possible coherence error.

**Keywords:** corpus linguistics, writing assessment, error tagging, coherence, Rhetorical Structure Theory

## Introduction

Research findings indicate that English language learners produce various kinds of discourse errors in their writing. However, learner corpora have been tagged only for a few types of discourse errors (Díaz-Negrillo & Fernández-Domínguez, 2006). Moreover, no learner corpus has been tagged for coherence errors. This paper reports an on-going study aiming to result in the tagging of discourse errors in the Language Training and Teaching Center (LTTC) English Learner Corpus, a learner corpus of English cooperatively constructed by the Graduate Institute of Linguistics and the Department of Foreign Languages and Literatures at National Taiwan University (NTU), the Department of English at National Chengchi University (NCCU), and the LTTC, a language testing company.<sup>17</sup> Given the aforementioned lack of extensive discourse error tagging in similar corpora, this project may function as a springboard for the quantitative testing of relevant theoretical and pedagogical claims.

In particular, this paper reports on a pilot study where 45 written samples from the aforementioned corpus were analyzed according to Rhetorical Structure Theory (RST) (Mann & Thompson, 1988). This pilot study examined whether an RST coherence analysis can indicate coherence breaks in the writing of learners, given the subjectivity that judgments about text coherence usually entail. Provided that results indicate that RST coherence analysis can reveal coherence breaks, the frequency of different

---

<sup>17</sup> This corpus was compiled under the supervision of Professor Hintat Cheung, the director of the Graduate Institute of Linguistics at NTU. The co-directors were Professor Zhao-Ming Gao, from the Department of Foreign Languages and Literatures at NTU, and Professor Siaw-Fong Chung, from the Department of English at NCCU. I was the post-doctoral research associate working on the project. In the academic year 2009-10, the other project members were two PhD students, Ms. Sally Chen and Ms. Chi-Yi Wu, and the research assistant and administrator, Ms. Su-Mei Chen. In the academic year 2008-9, the research assistant and administrator was Ms. San-Ju Lin.

kinds of coherence errors will be measured. This analysis will provide some preliminary indications in regards to which errors should be tagged in the corpus, since corpus error tagging systems attempt to tag the most frequent errors.

This paper first reviews the motivation for the study. Then, it focuses on why RST was chosen to help locate coherence errors in the data as well as background information of this theory. Third, it provides an overview of the LTTC English Learner Corpus and provides a description of the data, the method of the project, and findings. Finally, the paper ends with conclusions about the usefulness of RST analysis for the location of discourse errors and a summary of future research interests.

### Corpus error tags and learner errors

The pilot study reported in this paper was partly motivated by the lack of coherence error tags in corpora. Although L1 Chinese learners of English and learners of English in general make many discourse errors, in a recent review of four of the best documented and most representative learner-corpus error tagging systems, Díaz-Negrillo and Fernández-Domínguez (2006) observed that very few discourse errors have been tagged. This fact is reflected in Table 1, which lists the discourse error types tagged in well-known learner corpora.

Table 1 also indicates that different discourse errors have been tagged in different corpora.

The data on the Chinese Learner of English Corpus are especially relevant to this study. The Chinese Learner of English Corpus has the most extensive list of discourse errors, but these errors were considered syntactic rather than discursal, so they are tagged as syntactic errors. These facts can

Table 1 : Types of discourse errors tagged in four well-known learner corpora

| Learner corpus                                 | Types of discourse errors tagged   |
|--|--|
| Cambridge Learner Corpus (CLC)                 | Pronoun reference  |
| FreeText                                       | 'Cohesion' (i.e., subordination and coordination errors)<br>Style (unclear or heavy)<br>Register |
| International Corpus of Learner English (ICLE) | Unnatural discourse  |
| Chinese Learner English Corpus (CLEC)          | Coordination<br>Subordination<br>Anaphoric reference<br>Sentence fragment<br>Run-on sentences    |

be seen as indicating that discourse errors occur in the writing of Chinese learners of English and that these errors have not received the proper attention since they have not been tagged as such.

The other main reason for the desire to create discourse error tags for the aforementioned corpus is that the discourse errors of L1 Chinese learners of English have not been examined sufficiently through quantitative methods. The research testing the claim that the paragraphs and essays of L1 Chinese learners of L2 English have inductive/non-linear rather than deductive/linear order is the case in point. It has been claimed that these learners present the main point of their writing only at the end of a paragraph or essay, whereas in L1 English writing the main point is presented first (e.g., Kaplan, 1966; Matalene, 1985). The claim for the use of an inductive order only by L1 Chinese learners of English (and not by native speakers of English) has been challenged. For example, Scollon and Scollon (1995) used ethno methodology to show that inductive and deductive patterns both exist in the

writing of both native speakers of English and native speakers of Chinese but they are used for different pragmatic purposes. However, their analysis relates only to spoken discourse, so one cannot draw any conclusions about the existence of inductive patterns in written native English. This research gap is filled by Chen (2008). In a quantitative study, he found, among other things, that a minority of the native speakers of English preferred essays written with an inductive rather than deductive pattern and nearly half of them preferred paragraphs written in an inductive rather than a deductive order. This finding shows that inductive patterns can be used in written native English. Finally, Mohan and Lo (1985) review Chinese writing textbooks and analyze Classical Chinese texts to show that the deductive pattern is the most usual and prescribed essay writing pattern in Chinese. Moreover, when inductive order is found in the writing of L1 Chinese learners of English, there is controversy over the cause of this pattern. For example, one possible reason is the influence from L1 rhetorical structure, as in the contrastive rhetoric literature (e.g., Chen, 2001; Kaplan, 1966; Matalene, 1985). Another is the lack of relevant or useful feedback and instruction from teachers (e.g., Gonzalez, Chen, & Sanchez, 2001; Mohan & Lo, 1985). Yet another possible reason is the inability to properly structure an essay not only in the L2 but also in the L1 because one has not reached the right developmental stage in his/her writing ability (e.g., Mohan & Lo, 1985). It was considered that tagging the aforementioned learner corpus for errors of coherence would point to possible errors of inductive order and would, therefore, lead to more research on this controversial issue.

### Why was RST used in this study?

It is difficult to reliably identify coherence errors because readers of the

same text may form different interpretations of the coherence relations among elements of the text, as Mann and Thompson (1988) have pointed out. It is also more difficult to identify the exact location of coherence errors than that of other discourse errors. For example, let us consider the two extracts from the LTTC English Learner Corpus which appear in Figure 1. If we compare them, we will see that the discourse error in the first extract is easier to spot than the error in the second one. The first extract is an example of a sentence fragment and it can be located easily because our knowledge of syntax helps us in this task and because the error spans only one sentence. In the second extract, the error is more diffused in the paragraph, so it is more difficult for one to realize and specify which sentences are involved. This is an example of an inductive order error. The last sentence should have been the third, following the sentence “There are many reason to this problem.” for the paragraph to have a deductive organisation.

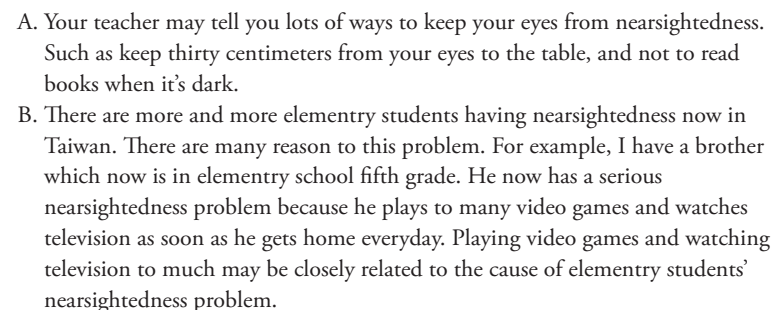
- 
- A. Your teacher may tell you lots of ways to keep your eyes from nearsightedness. Such as keep thirty centimeters from your eyes to the table, and not to read books when it's dark.
  - B. There are more and more elementary students having nearsightedness now in Taiwan. There are many reason to this problem. For example, I have a brother which now is in elementary school fifth grade. He now has a serious nearsightedness problem because he plays to many video games and watches television as soon as he gets home everyday. Playing video games and watching television to much may be closely related to the cause of elementary students' nearsightedness problem.

Figure 1 : Extracts from two paragraphs from the LTTC English Learner Corpus

Therefore, a bottom-up method of coherence relations should be used so that coherence errors will be identified as reliably and objectively as possible.

RST (Mann & Thompson, 1988) was chosen because the output of other methods of location coherence breaks, such as topical structure analysis and genre analysis in Watson-Todd et al. (2007), has been shown to have little relationship with English teachers' judgments. Second, strong correlations have been found between RST analyses which show that a text is coherent and subjective judgments that a text is coherent (Taboada & Mann, 2006a). Finally, other studies have assessed coherence in written discourse through RST analyses (e.g., Kong, 1998; Pelsmaekers et al., 1998), so the study reported here could benefit from reviewing this literature. Finally, RST was chosen because it has not been applied to the location of coherence errors, so an evaluation of its application for this purpose is interesting from a methodological perspective.

In their review of theoretical work on RST, Taboada & Mann (2006a, p. 425) give a simple definition of RST: "RST addresses text organization by means of relations that hold between parts of a text. It explains coherence by postulating a hierarchical, connected structure of texts, in which every part of a text has a role, a function to play, with respect to other parts in the text." The connections which are posited between parts of a text and which show the function of each "part of text" in the text are called "coherence relations." Coherence relations show the function that the analyst thinks that the writer intended each "part of text" to have in relation to other parts of text.

Some units are called "nuclei" and others "satellites." In RST jargon, nuclei are units of analysis which are necessary parts of a text and satellites are units of analysis which modify the meaning of the nuclei. The main idea of a text needs the nuclei to be put across but if the satellites were deleted, the same main idea, more or less, would be expressed.

For example, an analyst would say that there is an elaboration coherence relation between two units of analysis, if (s)he thinks that the aim of the

author is that the reader recognizes the satellite as providing additional information for the nucleus. Figure 2 shows an extract of a paragraph from the LTTC English Learner Corpus. Each unit of analysis appears within square brackets. The second and third clauses are linked through the relationship of Joint<sup>18</sup> because one is added to the other and jointly modify the first sentence by elaborating its meaning ("Elaboration").

[Your teacher may tell you lots of ways to keep your eyes from nearsightedness.]  
[Such as keep thirty centimeters from your eyes to the table,] [and not to read books when it's dark.]

Figure 2 : Extract from a paragraph from the LTTC English Learner Corpus

As mentioned previously, the coherence relations in a text are usually presented in a hierarchical structure. Figure 3 shows the structure of the extract in Figure 2. The software used to produce Figure 2 was the RST Annotation Tool by Daniel Marcu,<sup>19</sup> which is an improvement on Marc O'Donnell's RST Tool. In RST diagrams, coherence relations are indicated by arrows. An arrow starts from a satellite and points to a nucleus. However, there are also some coherence relations which link units of the same kind. The relation Joint is such a "multinuclear" relation.

As mentioned earlier, the analyst chooses the coherence relation which

<sup>18</sup> In recent publications on RST, Joint is not considered a coherence relation but a schema (see the paragraph below Figure 3 for a definition of "schema") because there are no conditions on its application (e.g., Taboada & Mann, 2006b, p. 569). As it is explained in the "Method" section of this paper, the coherence relations posited in this study were those posited by Bill Mann and found at the website <http://www.wagsoft.com/RSTTool/RSTDef.htm>, together with some additional ones posited by Carlson, Marcu and their collaborators (Carlson & Marcu, 2001). "Joint" is among the coherence relations in Bill Mann's list, so it is considered a coherence relation in the current study.

<sup>19</sup> This software was downloaded from <http://www.isi.edu/licensed-sw/RSTTool/index.html>.

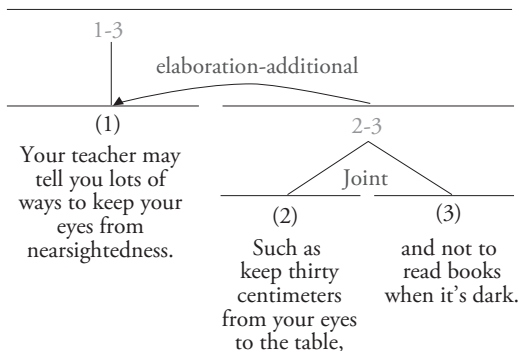


Figure 3 : RST diagram indicating the coherence relations in the extract presented in Figure 2

seems to have the function that the writer intended each 'part of text' to have in relation to other parts of text. There are certain constraints on the analyst's choice of a coherence relation, but this paper describes only one of them because, although they guided RST data analysis, the rest are not directly related to the drive of this study. The constraint which has helped to form the method of this project is that each text should have the structure of a coherence-relation schema. Such a schema is an abstract representation of coherence relation diagrams. The analyst tries to fit a whole text into one schema and to fit sub-schemas under this schema. Figure 4 shows the schemas which have been posited by Mann and Thompson (1987, 1988).

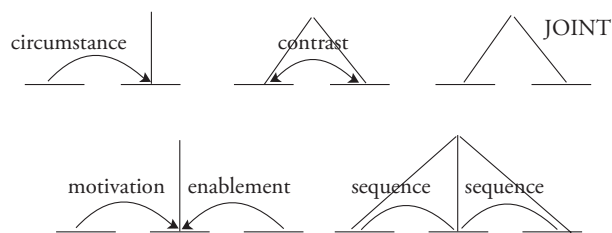


Figure 4 : Figure taken from Mann and Thompson (1987, p. 7)

The aforementioned schema application constraints have some consequences for the location of coherence errors. Since all these requirements must be met for a text to be considered coherent in RST, their violations indicate coherence errors. Therefore, coherence errors are expected to be indicated by diagrams which:

- a) do not comply with the structure of any schema,
- b) include sub-diagrams which do not comply with the structure of any schema, or
- c) include schemas which include shared units of analysis ("crossed dependencies").

This conclusion leads us to the rationale of this study: each kind of coherence error is indicated by one of these abnormalities in the diagram. By listing the abnormalities which characterize each kind of coherence error, the corpus can later be tagged for coherence errors in a principled way.

## Data

The data are 45 paragraphs written by Taiwanese lower-intermediate learners of English in Writing Task 2 of the Intermediate General English Proficiency Test (GEPT) examination, a language proficiency examination administered by the LTTC. In this task, test-takers are asked to write a 120-word paragraph. These files form part of the written section of the LTTC English Learner Corpus. The corpus consists of language samples by Taiwanese learners of English who have taken the GEPT. In order to examine coherence errors in paragraphs written on more than one topic, the 45 paragraphs were equally distributed across topics. Topics were presented to test-takers in Chinese. Two of these topics were about personal preferences (favorite food and idol) and the third asked test-takers to explain why many

elementary-school children in Taiwan are nearsighted and to propose effective ways of preventing nearsightedness.

To ensure analyzed data would vary in terms of coherence error types, samples were equally distributed across score bands in each topic. In other words, in each topic five files had low scores (ranging from 1 to 2), five files had medium scores (ranging from 2.5 to 3.5) and five had high scores (ranging from 4 to 5).

## Method

The method involved the analysis of the aforementioned paragraphs by the author according to RST and using the RST Annotation Tool software.

The units of analysis were defined in the same way as in the tagging of 385 documents of American English selected from the Penn Treebank (Carlson & Marcu, 2001). Broadly speaking, clauses were the units of analysis, except when they were complements of prepositions and verb objects; those clauses belonged to the same unit of analysis as the clauses where the prepositions and verbs occurred. However, because the tagset that Carlson, Marcu and their collaborators used was specific to the nature of the texts which they analysed (that is, Wall Street Journal articles), I preferred to use the more neutral coherence relation categories by Bill Mann.<sup>20</sup> Since I combined the units of analysis from the Penn Treebank corpus and Bill Mann's categories, I had to compromise the unit-of-analysis segmentation in cases where the units of analysis warranted a coherence relation which was not among those in Bill Mann's list. This happened when the coherence

<sup>20</sup> These are the original categories posited by Mann and Thompson (1987, 1988) with some additions and can be found at this website: <http://www.wagsoft.com/RSTTool/RSTDefs.htm>.

relation of Attribution was posited by Carlson and Marcu to link speech and thought verbs with their complements. In these cases, I considered the verb and its complement clause as one unit of analysis.

As the analysis of the texts was progressing, it became obvious that Bill Mann's list of relations could not cover all the coherence relations in the data, so they were supplemented with eight relations from the tagset by Carlson, Marcu and their collaborators (Carlson & Marcu, 2001). These additional coherence relations were: Same-unit, Comment, Conclusion, Topic-shift, Manner, Explanation-argumentative.

## Results and discussion

### Qualitative results

Table 2 summarizes the coherence breaks indicated by the main abnormalities found in the RST diagrams.

Table 2 : Abnormalities found in the RST diagrams of the 45 paragraphs

| Diagram abnormalities                      | Coherence breaks indicated by diagram abnormalities  |
|--|--|
| Dangling units of analysis                 | Irrelevant content<br>Incomprehensible content<br>'Self-sufficiency'   |
| Crossed dependencies                       | Although a sub-diagram has already been formed for one part of the text, a coherence relation arises between another text part and a unit which is a member of the first sub-diagram |
| Unexpected relation                        | Motivation   |
| Relations in unexpected parts of a diagram | Conclusion<br>Background   |

The first problem was dangling units of analysis, that is, clauses or larger elements which seem unrelated to the content of the rest of the text, so they cannot be linked to it through a coherence relation. Such dangling units mostly indicated irrelevant content. One unit was left dangling because it was impossible to understand its meaning. Finally, there was one instance of a self-sufficient clause, which gave the reason why the writer liked a specific foreign food in a postscript.

Figure 5 gives an example of a dangling unit with irrelevant content.

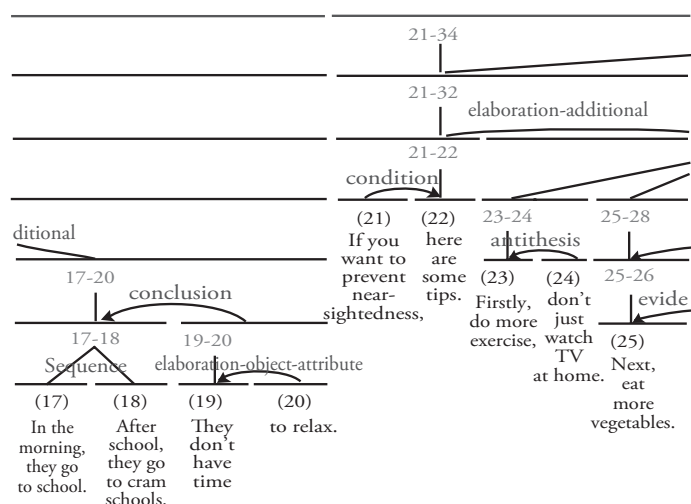


Figure 5 : Extract from a diagram with a dangling structure (units 21-34)

This extract comes from a paragraph written on the topic about the nearsightedness of elementary school children in Taiwan. Given that the topic asked test-takers to propose effective methods of preventing nearsightedness in general, the advice which the writer gives to the reader on how to avoid nearsightedness in the sub-diagram which consists of units 21 to 34 was tagged irrelevant.

An example of crossed dependencies cannot be illustrated diagrammatically because the RST Annotation Tool automatically corrects such abnormalities in a diagram. However, one can consider the coherence relations among the units of the extract in Figure 6. This figure shows the first lines written in a paragraph on the favorite exotic food topic. In this figure, the units are numbered for ease of reference in the discussion that follows.

1. [Taiwan is a special country.]
2. [We can eat a lot of foods from other countries.]
3. [They are gathered in this small island.]
4. [Like Japan, America, Thailand and more.]

Figure 6 : Extract from a paragraph on the favorite exotic food topic

Unit 3 restates information given in unit 2, so 3 is the satellite and 2 the nucleus of a Restatement coherence relation. Together, they express a result which stems from the fact that Taiwan is a special country, expressed in unit 1. Therefore, units 2 and 3 together form the satellite of a Result coherence relation, where unit 1 is the nucleus. Unit 4 exemplifies the countries whose food the Taiwanese can eat in Taiwan, so it is the satellite of an Elaboration coherence relation and 2 is the nucleus. This coherence relation is problematic because unit 4 intrudes in the sub-diagram which has already been formed by units 2 and 3.

Coherence relations which should not have been there form the next problem in the RST diagrams. The only such coherence relation which was found in the pilot was Motivation. In this coherence relation, the satellite unit offers a reason why the reader should do what is expressed in the nucleus unit. This relation is found in argumentative discourse (Azar, 1999) and not in expository and narrative discourse, which the GEPT test-takers were expected to produce. Figure 7 gives two examples of this error in an extract from a paragraph on nearsightedness of elementary students in Taiwan.



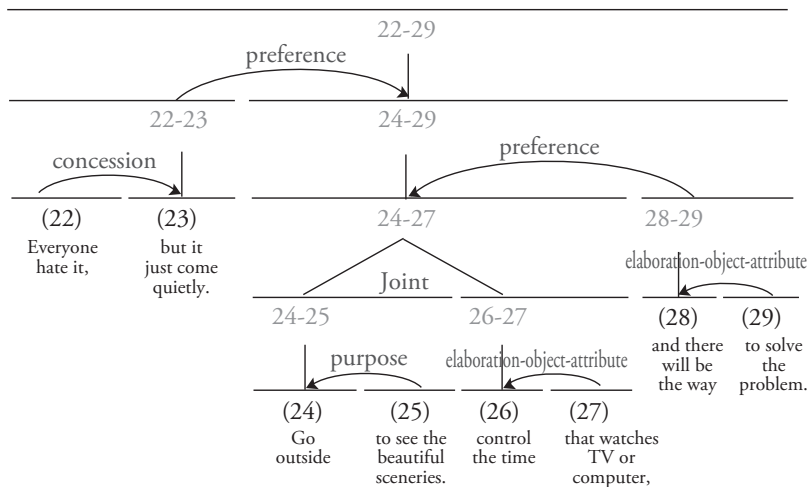


Figure 7 : Extract from a diagram with an unwarranted Motivation coherence relation

In units 22 and 23, ‘it’ refers to nearsightedness. These units jointly form a sub-diagram which serves as the satellite in a Motivation relation because they give a reason why someone should do the actions described in the units 24-29. Units 28 and 29 have the same function for units 24-27, so they are the satellite in a Motivation relation as well.

Finally, coherence relations in inappropriate parts of a text are problematic as well. These coherence relations are acceptable if they occur in the right parts of a text but there were cases where their location was inappropriate and indicates inductive ordering of content. The Conclusion coherence relation indicates a relation where the satellite is a reasoned judgment, inference, necessary consequence or final decision. For example, a student explained why Taiwanese elementary pupils are near-sighted by giving the example of what happened to her younger brother and concluded that “playing video games and watching television too much may be closely related

to the cause of elementary students’ nearsightedness problem.” The Background coherence relation usually appears in introductions or briefly in later parts of a text but when students use it to structure their reasoning it may lead to inductive content. Another student writing on the nearsightedness topic gave a lot of background information on Taiwanese students’ habits and this structure made the content seem inductive.

Although the RST analysis yielded a wealth of diagram problems which indicate coherence errors in the data, some coherence errors did not show up as problems in the RST diagrams. In other words, the aforementioned diagram errors are not enough to pinpoint all the coherence errors in the data. There are cases where the writer inappropriately addresses the reader but this does not lead to a structural error in the diagram and cases where a topic sentence is missing or scattered in different parts of the text without affecting the diagram. Therefore, the intuition of the error tagger is always necessary for the location of coherence errors.

## Quantitative results

The frequency of each kind of diagram abnormality was examined to see which of these errors are more likely to be frequent and, therefore, worthy of further investigation. The rationale is that if some of these errors are frequent, they should be tagged in a larger number of writing samples so as to examine whether they are indeed frequent; if they prove to be frequent, they warrant to be tagged in the LTTC English Learner Corpus.

The coherence error frequencies in this study should be interpreted with caution because they are based on an RST analysis which has been conducted only by one person and only once. In other words, they are based on data which have not been checked for their validity and reliability. Moreover, the number of writing samples analyzed was small, so no inferential statistics have

been conducted on the data.

As it has been mentioned in the overview of the qualitative results, dangling structures usually indicated irrelevant content but there was also one case where I could not link a structure to a preceding sub-diagram because this structure was incomprehensible and another where it appeared in a post-scriptum. Because these two errors occurred only once each, I have excluded them from the calculations which resulted in the figures in Table 3. In this table, because written samples varied in terms of their length, the number of occurrences of dangling structures was divided by the total number of units of RST analysis in each sample. In this way, the frequency of this diagram abnormality was normalized in a way which was appropriate to the way texts were analyzed. The last column is a coarser estimation of the frequency of dangling structures per topic; it is the count of the texts which included at least one dangling structure.

Table 3 : Irrelevant content instances across topics according to the RST analysis of 45 paragraphs

| Topic           | Cumulative 'dangling' structures normalized per RST unit of analysis | Mean 'dangling structures' normalized per RST unit of analysis | Writing samples with at least one dangling structure; percentage of texts per topic is given within parentheses |
|-----------------|--|--|---|
| Nearsightedness | 0.377  | 0.021  | 5 (33.33%)  |
| Idol            | 0.059  | 0.004  | 1 (6.66%)   |
| Favorite food   | 0.111  | 0.012  | 2 (13.33%)  |

All three measures of frequency agree with each other. They show that in the Nearsightedness topic there are more dangling structures than in the other

topics and that the Favorite food topic contains more dangling structures than the Idol topic. This finding can be seen as indicating that topic affects the occurrence of irrelevant content. Especially in terms of the last frequency measure, it is impressive that in one topic one third of the samples contained irrelevant content. All the frequencies are small, but it should be kept in mind that these samples were only single short paragraphs, so learners did not have many “opportunities” to write irrelevant content.

These differences in the frequency of dangling structures were also examined to see whether they seem related to the score band (low, medium, or high) under which the samples fall. In Table 4, the cumulative percentages of the dangling structures are presented in terms of essay topic and score band.

Table 4 : Cumulative percentage of “dangling” structures per topic and score band

| Score band       | Essay topic     |      |               |
|------------------|-----------------|------|---------------|
|                  | Nearsightedness | Idol | Favorite food |
| Low score        | 37.92%          | 0%   | 50%           |
| Mid score        | 38.65%          | 100% | 0%            |
| High score       | 23.42%          | 0%   | 50%           |
| Total percentage | 100%*           | 100% | 100%          |

*Note.* The total number from the percentages in this column is 99.99% because these numbers are rounded. The exact total number is 100%.

The breakdown of samples which contain dangling structures in the nearsightedness topic is expected, since one would expect that learners with low and mid scores would be more likely to include irrelevant content in their writing than the high-performing learners. The data for the other two topics

are more complicated, since all cases of irrelevant content in the idol topic occurred in the middle-score paragraphs and half of them in the low- and the other half in the high-score paragraphs in the food topic. However, this finding can be easily explained by the very few occurrences of dangling structures for the idol and food topics. There was only one occurrence of a dangling structure in the idol topic and it was in a middle-score paragraph and there were only two occurrences in the food topic, one in a low- and the other in a middle-score paragraph.

The discussion of results about dangling structures indicates that in most cases they indicate irrelevant content, but that for Taiwanese learners of low-intermediate proficiency level, whether such coherence errors occur frequently depends on the essay topic.

The coherence error which stems from crossed dependencies seems to be very rare since it occurred only once. There is some controversy over whether such diagrammatic structures should be considered erroneous, because it has been claimed that crossed dependencies occur in the productions of native speakers as well (Wolf & Gibson, 2004, 2005). Therefore, such errors probably do not warrant further investigation or error tagging.

As explained in the previous section, the coherence relation of Motivation was not expected to be found because it occurs in argumentative text types whereas the essay topics were expository. This coherence relation occurred only in the paragraphs written on the nearsightedness topic. This finding is congruent with the previous finding that irrelevant content made manifest by dangling structures was much more frequent in the nearsightedness than in the other texts. Indeed, it seems that there is some interrelation between dangling structures and the existence of a Motivation relation in nearsightedness texts, as shown in Table 5.

Table 5 : Motivation relation instances in the paragraphs on the Nearsightedness topic

| Cumulative instances of Motivation coherence relation normalised per RST unit of analysis | Mean instances of Motivation relation normalized per RST unit of analysis | Percentage of paragraphs with Motivation relation which also have 'dangling' structures |
|---|---|---|
| 0.112   | 0.007   | 66.67%  |

Nevertheless, the very low mean occurrence of this Motivation relation and the fact that it occurred only in one topic perhaps indicates the limited utility of looking for such errors in the data of intermediate-level Taiwanese learners of English.

In terms of the coherence errors due to the occurrence of a coherence relation in an inappropriate part of a paragraph, Table 6 below presents the same kinds of normalized data as Table 3 but for the inappropriate occurrences of the Background coherence relation.

Table 6 : Irrelevant content instances across topics according to the RST analysis of 45 paragraphs

| Topic           | Cumulative inappropriate uses of the "background" coherence relation normalized per RST unit of analysis | Mean inappropriate uses of the "background" coherence relation normalized per RST unit of analysis | Number of writing samples with at least one instance of an inappropriate use of the "background" coherence relation; percentage of texts per topic is given within parentheses |
|-----------------|--|--|--|
| Nearsightedness | 0.059  | 0.004  | 1 (6.67%)  |
| Idol            | 0.184  | 0.012  | 3 (20%)  |
| Exotic food     | 0  | 0  | 0 (0%)   |

As can be seen, the majority of cases occur in the Idol topic, so it seems that the occurrence of such errors also depends on topic. To see whether there was a score-band effect as well, in Table 7, the cumulative percentages of the dangling structures are presented in terms of essay topic and score band. This table does not present data for the Favorite food topic because the Background coherence relation did not appear in those texts.

Table 7 : Cumulative percentage of cases of Background coherence relation per topic and score band

| Score band       | Essay topic     |        |
|------------------|-----------------|--------|
|                  | Nearsightedness | Idol   |
| Low score        | 0%              | 0%     |
| Mid score        | 0%              | 42.83% |
| High score       | 100%            | 57.17% |
| Total percentage | 100%            | 100%   |

This table indicates that the Background coherence relation occurred in the wrong part of the text for paragraphs which achieved medium and high scores. This finding may not be significant in the Nearsightedness topic since this error was only found in one paragraph, but it seems to be more important in the Idol topic since this error occurred in one fifth of these paragraphs. Therefore, the frequency of this kind of error seems to warrant further investigation.

The last coherence error indicated by the RST analysis is the use of the Conclusion coherence relation in an inappropriate part of the text. This error occurred only twice and only in two middle-score paragraphs, so it seems that this error occurs rarely. Moreover, it occurred only in the Nearsightedness topic, so a topic effect is also possible for this error. Given the low frequency

and the occurrence only in one topic, this error may not warrant further examination.

## Conclusion

The results of this pilot study indicate that problems in RST diagrams point to various coherence errors related to both global and local coherence. In particular, dangling units and unexpected coherence relations in the diagrams are indications of irrelevant content. Thus, they point to global coherence errors. Coherence relations in inappropriate parts of the text indicate inductive content order. Therefore, they highlight a kind of a global coherence error as well. Finally, the crossed dependencies indicate local coherence errors because they apply to coherence relations within, rather than across, sub-diagrams. Consequently, this method of textual analysis seems promising.

This categorization of indices of coherence errors can be done based on the diagrammatic properties of these errors. Global coherence errors occur between sub-diagrams whereas local ones occur within sub-diagrams. This finding indicates another reason why this method merits further examination. RST analysis is more data-driven and constrained than other methods, so it can locate and categorize coherence breaks more reliably.

The finding that not all coherence errors were manifested as problems in the RST diagrams indicates that for some coherence errors, the tagger's intuition is needed. Of course, it is also needed whenever one "problem" in the diagram is an index of more than one possible coherence errors. Therefore, although RST analysis helps to constrain the subjectivity of the analyst because of the constraints imposed by the definitions of the coherence relations and the coherence-relation schemas, the detection of coherence

errors still partly relies on the analyst's subjective judgments.

The quantitative results of this study are only indicative, since the RST analysis was conducted by only one researcher and for a small number of writing samples. These results show that the coherence errors of crossed dependencies and the inappropriate occurrence of the Conclusion coherence relation occur very infrequently. Therefore, such errors probably do not warrant any further investigation and perhaps need not be tagged in the LTTC English Learner Corpus. The relatively frequent occurrences of the Background coherence relation in inappropriate parts of a paragraph indicates that low-intermediate level Taiwanese learners of English make inductive-order errors. Therefore, such analyses should be conducted with more writing samples and tested for validity and reliability in order to test whether they are indeed frequent and, consequently, should be tagged in the corpus.

Further data analysis is also necessary to see whether the topic effects on coherence errors occur when more samples are analyzed. If these topic effects are confirmed through further research, attempts should be made to explain them. This research would be beneficial for English language proficiency examination design, since it would indicate what factors need to be standardized in order to prevent higher frequencies of coherence errors for some essay topics. This research could also benefit the design of Automatic Writing Evaluation (AWE) software, such as *Criterion* (e.g., Burstein, Chodorow, & Leachock, 2004) since, if the topic-related factors shown to influence these errors could be detected by AWE software, essay scoring and feedback could be refined (Skoufaki, 2009).

As mentioned already, because the RST analysis in this study was conducted by only one researcher, the validity and reliability of the analysis was not tested. In future studies, the coherence breaks pinpointed by the method will be compared against the intuitions of another linguist who will

be trained to do RST analysis, and against the judgments of experienced English language teachers and native speakers of American English.

From a practical point of view, this pilot study has shown that this method is labor-intensive, so only a small part of the corpus can be fully tagged for coherence relations. A broader set of coherence relations should be created for the discourse error tagging of the whole corpus.

In conclusion, this pilot study has shown that RST analysis of second language writing can lead to RST diagram abnormalities which seem to indicate specific kinds of coherence errors but that it also fails to locate some kinds of coherence errors. The quantitative results suggest that some coherence errors are more frequent than others and, therefore, the former warrant further examination more than others. The quantitative analysis also indicated some topic effects in the occurrence of some kinds of coherence errors, a finding which also merits further examination because it may have various pedagogical applications. Therefore, this study opens many avenues for further research.

## References

- Azar, M. (1999). Argumentative Text as Rhetorical Structure: An Application of Rhetorical Structure Theory. *Argumentation*, 13(1), 97-114.
- Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion Online Writing Evaluation service. *AI Magazine*, 25(3), 27-36.
- Carlson, L., & Marcu, D. (2001). *Discourse Tagging Manual*. ISI Tech Report ISI-TR-545.
- Chen, J. P. (2001). *Markedness in intercultural discourse: a study of Chinese EFL students' discourse patterns*. Ph.D. thesis, Guangdong University of Foreign Studies, in ERIC, RIE, June 2001.
- Chen, J. P. (2008). An investigation into the preference for discourse patterns in the Chinese EFL learning context. *International Journal of Applied Linguistics*, 18(2), 188-211.
- Díaz-Negrillo, A., & Fernández-Domínguez, J. (2006). Error tagging systems for learner corpora. *Revista Española de Lingüística Aplicada*, 19, 83-102.
- González, V., Chen, C.-Y., & Sanchez, C. (2001). Cultural Thinking and Discourse Organizational Patterns Influencing Writing Skills in a Chinese English-as-a- Foreign-Language (EFL) Learner. *Bilingual Research Journal*, 25(4), 417-442.
- Kaplan, R. (1966). Cultural thought patterns in intercultural education. *Language Learning*, 16(1), 1-20.
- Kong, K. C. C. (1998). Are Simple Business Request Letters Really Simple? A Comparison of Chinese and English Business Request Letters, *Text*, 18(1), 103-41.
- Mann, W. C., & Thompson, S. A. (1987). *Rhetorical Structure Theory: A Theory of Text Organization* (No. ISI/RS-87-190). Marina del Rey, CA: 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.
- Matalene, C. (1985). Contrastive rhetoric: an American writing teacher in China. *College English*, 47(8), 789-808.
- Mohan, B. A., & Lo, W. A.-Y. (1985). Academic writing and Chinese students: transfer and developmental factors. *TESOL Quarterly*, 19(3), 515-34.
- 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.
- Scollon, R., & Wong-Scollon, S. (1995). *Intercultural communication: A discourse approach*. Blackwell: Oxford, U.K. and Cambridge, MA, U.S.A.
- 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. In C. L. Liu & Z.-M. Gao (Eds.), *International Journal of Computational Linguistics and Chinese Language Processing* [Special Issue on Computer Assisted Language Learning], 14(2), 181-203.
- Taboada, M., & Mann, W. C. (2006a). Rhetorical Structure Theory: looking back and moving ahead. *Discourse Studies*, 8(3), 423-459.
- Taboada, M., & Mann, W. C. (2006b). Applications of Rhetorical Structure Theory. *Discourse Studies*, 8(4), 567-588.
- Watson-Todd, R., Khongput, S., & Drasawang, P. (2007). Coherence, cohesion and comments on students' academic essays. *Assessing Writing*, 12(1), 10-25.
- Wolf, F., & Gibson, E. (2004). Representing Discourse Coherence: A Corpus-based Analysis. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING)*, Geneva, Switzerland.
- Wolf, F., & Gibson, E. (2005). Representing Discourse Coherence: A Corpus-based Analysis. *Computational Linguistics*, 31(2), 249-287.