

Author: Sophia Skoufaki

Department of Language and Linguistics

University of Essex

Wivenhoe Park

Colchester

CO4 3SQ

Essex, UK

Telephone: 01206 873754

Fax: -

E-mail: sskouf@essex.ac.uk

Article title: Rhetorical Structure Theory and coherence break identification

Short title: RST and coherence break identification

Word count (all inclusive): 8,611

Character count (with spaces): 56,121

Author bionote

Sophia Skoufaki is a Lecturer at the University of Essex. She specialises in vocabulary studies and discourse coherence. Her current research examines academic vocabulary teaching and learning, figurative language processing, and the application of Rhetorical Structure Theory to language teaching.

Correspondence address: Department of Language and Linguistics, University of Essex. Wivenhoe Park, Colchester, CO4 3SQ, Essex, UK, Email: sskouf@essex.ac.uk

Abstract: This article examines the claim of Rhetorical Structure Theory (RST) that violations of RST diagram formation principles indicate coherence breaks. In doing so, this article makes a significant contribution to the testing of RST. More broadly, it indicates that examining the coherence-break identification potential of coherence theories could help specify each theory's purview and, in the long term, lead to the creation of hybrid models of coherence. Moreover, it paves the way for the development of training resources on discourse (in)coherence for language teachers, exam markers and language learners. 84 paragraphs written by Taiwanese learners of English were analysed according to RST and coherence measures were calculated on the basis of this analysis. The results suggest that the violation of any diagram-formation principle indicates coherence breaks, thus corroborating this RST claim. Inter- and intrajudge agreement in terms of both RST coding and coherence measures calculated on the basis of coherence breaks are reported and discussed. The kinds of coherence breaks which are and are not located by RST analysis are discussed and exemplified. The paper concludes with a discussion of implications for pedagogy and future research.

Keywords: coherence, discourse analysis, English as a Foreign Language, Rhetorical Structure Theory, writing

Rhetorical Structure Theory and coherence break identification

Affiliation: University of Essex, UK

1 Introduction

Theories of discourse coherence, which has been broadly defined as ‘the sense of overall unity’ of a text (Mann et al. 1992: 6), are often used as descriptive tools in the analysis of coherence. Although exploring how a theory can account for a dataset can lead to the adaptation of the theory (Bechtel 1988: 37), such adaptations are usually small. Rhetorical Structure Theory (RST) (Mann and Thompson 1987, 1988) is a case in point. Although various studies corroborate RST (Taboada and Mann 2006) and some do not (e.g., Stede 2008), RST claims have not been tested systematically because its creators considered it a ‘descriptive linguistic approach’ (Mann et al. 1992: 270).

The present study tests the RST claim that violations of RST diagram-formation principles suggest coherence breaks, that is, parts of a text where coherence seems disrupted to the reader. So far, studies have assumed the validity of this claim except Skoufaki (2009) and Ahmadi and Parhizgar (2017). The study reported here avoids the methodological shortcomings of these studies and, consequently, makes a significant contribution to the testing of RST.

This study also examines which kinds of coherence breaks remain undetected by violations of RST diagram-formation principles. This investigation will help to delineate the purview of RST vis-a-vis other coherence theories.

In what follows, Section 2 introduces RST briefly. Section 3 reviews studies which applied coherence theories to coherence-break identification. Sections 4 and 5 summarise the method and results of this study, respectively. Section 6 draws conclusions and discusses implications for language pedagogy and future research.

2 Introduction to RST

RST ‘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’ (Taboada and Mann 2006: 425). In other words, coherence relations connect parts of a text. The analyst chooses the coherence relation which seems to have the function that the writer intended each ‘part of a text’ to have. The coherence relations originally posited by Mann and Thompson (1987, 1988) are not considered definitive.¹ Depending on the nature of the texts analysed, different sets of relations have been posited.

Some parts of text are nuclei and others are satellites. Satellites modify the meaning of the nuclei. The main idea of a text needs the nuclei to be put across. The nuclei, satellites and coherence relations in a text are presented in a hierarchical structure called *RST diagram* or *RST tree*. In RST diagrams, a coherence relation with only one nucleus is indicated by an arrow which starts from the satellite and points to the nucleus. A multinuclear coherence relation links two or more nuclei and is indicated by straight lines which connect the nuclei.

¹ Mann and Thompson’s (1987, 1988) coherence relations can be found at

www.wagsoft.com/RSTTool/RSTDefs.htm.

Figure 1 is the RST diagram of a text analysed in this study, which is provided here to illustrate RST diagrams. The topic rubric asked writers to describe their favourite exotic food and explain why they like it.

The text was first segmented into clauses, which constitute the unit of analysis. I considered unit 4 the main one in the diagram because it mentioned the writer's favourite exotic food. The fact that unit 4 expresses the most important concept in the text is evident from its appearance at the top level of the diagram. The units before unit 4 were considered its 'Preparation', that is, information aiming to make the reader more interested in or ready to read unit 4. The units following unit 4 were considered its 'Non-volitional cause', that is, the events described in these units caused the writer to have the exotic food preference mentioned in unit 4; and the 'Non-volitional cause' coherence relation was chosen over the 'Volitional cause' relation because food preferences are not intentional (in RST jargon, 'volitional') actions.

Figure 1 also shows how units below the top level of the RST diagram are connected. The structure of units 1-3 will be described here to further illustrate RST diagram construction. Unit 2 was tagged as the satellite of unit 3. They were linked with an 'Unconditional' coherence relation because the author seems to claim that the truth of unit 3 (seeing exotic food in Taiwan) does not depend on unit 2 (one's exact location in Taiwan). Units 2 and 3 together form the satellite of unit 1 in an 'Elaboration' relation. This relation was chosen because these units provide detailed information in support of the claim made in unit 1.

Figure 1 about here

Each of the nuclei of the diagram appears under a vertical line. As can be seen from the number of arrows in the diagram, it contains 13 mononuclear relations. Units 5-6 and 7-9 are connected via the only multinuclear relation in the diagram.

The creation of an RST diagram is limited by certain constraints. The main constraint 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. For example, Figure 1 includes instantiations of the first, third, and fourth schemas in Figure 2.

Figure 2 about here

According to Mann and Thompson (1988: 248-249), a well-structured RST diagram follows these principles:

- **Completedness:** At the top of the diagram all units should be linked together in one schema. This principle is observed in Figure 1 because the top row of the diagram is an instantiation of only schema 4 from Figure 2.
- **Connectedness:** With the exception of the top node, all nodes should either be leaf nodes or nodes inside a schema. This principle is observed in Figure 1 because except for the top node (node 4), nodes 1-3 and 5-16 are leaf nodes and the nodes below them are inside schemas.
- **Uniqueness:** Each schema in a diagram will consist of a different set of units. This principle is observed in Figure 1 because no unit belongs to more than one schema.
- **Adjacency:** Only adjacent functional elements can be grouped together to form larger functional elements. This principle is observed in Figure 1 because in every row of the diagram only adjacent functional elements are linked to each other.

Completedness, Connectedness, and Uniqueness are sufficient conditions for an RST diagram's well-formedness (Mann and Thompson 1988: 249). Coherence relations can exist between non-adjacent units or unit spans provided they have the same parent node (de Silva 2007: 37).

The constraints on diagram formation summarised above are at the heart of the explanation RST offers for coherence. If 'an RST diagram is a connected whole, with every unit of the text linked into the diagram somehow, then the analysis demonstrates how the text can be seen as coherent' (Taboada and Mann 2006: 428).

3 Coherence theories and the identification of coherence breaks

3.1 Coherence theories other than RST and the identification of coherence breaks

Watson Todd and his collaborators (2004, 2007) have examined the potential of various theories of discourse to identify coherence breaks in the writing of English as a Foreign Language (EFL) learners. Watson Todd et al. (2004) examined how much the approach to topic identification (summarised in Watson Todd 1997, 1998) – an approach combining theme-rheme progression and Hoey's (1991) lexical analysis – matched teachers' assessment of EFL writing for coherence.² Correlation analyses were conducted between coherence measures stemming from the output of the topic-based analysis of 28 texts and the scores four EFL teachers gave these texts for coherence. The only coherence measure which correlated

² Watson Todd (2016) offers a thorough introduction to these and other text-linguistics approaches to discourse topics.

significantly with the teachers' scores was the number of moves per 10 T-units; moreover, it correlated significantly with three out of the four teachers' coherence scores.

In Watson Todd et al. (2007), lexical analysis, topical structure analysis and genre analysis were applied to eight academic essays written by first-year MA students at a Thai university. These essays were assessed by two academics according to their normal criteria. None of the coherence breaks located agreed significantly with the quantity or content of the academics' discourse-related comments. However, this study probably underestimates the coherence-break identification potential of these linguistic analyses because academics were not asked to identify coherence errors or rate the texts for coherence.

Miltsakaki and Kukich (2004) and Wang et al. (2012) examined the coherence-break identification potential of Centering Theory (Grosz et al. 1995). In Miltsakaki and Kukich (2004), the Automated Writing Evaluation (AWE) software *e-rater* (Burstein et al. 1998) was adapted to include a percentage of rough shifts, that is, abrupt transitions between topics, among the linguistic features it used to produce essay scores. This adaptation increased the software's ability to predict experts' holistic scores for quality of writing. Wang et al. (2012) created an AWE programme which would give EFL learners feedback on local coherence errors after analysing their essays in terms of Centering Theory. The coherence breaks the software located overlapped moderately with those the human experts located.

The low or moderate agreement between the breaks located intuitively and via these approaches to coherence applied in the studies reviewed in this section suggests that other approaches to coherence also need to be applied to coherence error identification.

3.2 RST and the identification of coherence breaks

This section will summarise studies which used RST to identify coherence breaks and will indicate relevant research gaps.

Research on whether specific coherence relations characterise certain text types is indirectly related to coherence-break identification. In Sanders (1997), when presented with texts with coherence relations ambiguous between being semantic (that is, relations due to the propositional content of the linked discourse units) and pragmatic (that is, relations due to a speech act that the reader thinks the author intended to perform), undergraduate students familiar with text analysis categorised these relations as semantic when reading descriptive texts and as pragmatic when reading argumentative texts. This finding suggests that a reader tends to expect coherence relations which are typical of a text type. Since a reader's sense of coherence in a text depends on the extent to which his 'expectations of what the discourse should be' are met (Williams 1985: 474, cited in Swales 1990), the aforementioned finding suggests a reader may consider unexpected coherence relations as the loci of coherence breaks. However, opinions vary on this issue. For example, Virtanen (1992) suggested that characteristics typical of narrative texts may appear in texts which have other discourse functions (e.g., argumentative) without seeming strange, while the opposite (i.e., characteristics typical of argumentative texts appearing in narrative texts) is not acceptable.

Mann and Thompson (1988: 260) mention missing coherence relations as a cause of coherence breaks. Pitler and Nenkova (2008) considered whether the number and kind of relations can predict ratings given by at least three college students to 30 Wall Street Journal articles in response to the question, 'On a scale of 1 to 5, how well written is this text?'. As Pitler and Nenkova (2008: 189) consider 'text readability and text coherence as equivalent properties', they view these ratings as equivalent to coherence ratings. These articles were annotated in terms of syntactic properties, vocabulary, cohesive devices, entity grids (Barzilay and Lapata 2008), number of coherence relations and number of explicit and

implicit coherence relations. The number – not the kind – of relations predicted most of the variance in the ratings.

Gruber (2006) and O'Brien (1995) RST-analysed native-speaker university students' essays. In Gruber (2006) relations among essay sections and in O'Brien (1995) relations among paragraphs were considered. Both studies found cases where text belonged to more than one unit span, called disrupted text spans by Gruber (2006: 114).

Although the findings in Gruber (2006) and O'Brien (1995) suggest that the violation of the Uniqueness principle of RST diagram formation can indicate coherence breaks, no study has systematically examined which coherence breaks are indicated by the violation of RST diagram-formation principles. The present study aims to fill this research gap.

This study is influenced by Skoufaki (2009) and Ahmadi and Parhizgar (2017). In Skoufaki (2009) RST analysis helped identify various coherence errors in paragraphs written on four topics in Writing Task 2 of the Intermediate General English Proficiency Test (GEPT) examination. The GEPT is a language proficiency examination administered by the Language Training and Testing Center (LTTC), a language testing company in Taiwan. Because paragraphs were analysed only by me and only once, the reliability of this data analysis could not be assessed.

Skoufaki (2009) was partially replicated by Ahmadi and Parhizgar (2017), which analysed essays written in English by male university students in Iran. Coherence errors were located 'whenever whole or part of an RST diagram of an EFL learner's writing does not conform to the aforementioned conditions', where the aforementioned conditions referred to the principles of RST diagram formation (Ahmadi and Parhizgar 2017: 14). Coherence break types overlapped between Skoufaki (2009) and Ahmadi and Parhizgar (2017) but some of the breaks the latter identified were not due to violations of RST diagram formation principles. For example, some relations were considered 'incorrect' although they did not violate any of

the RST diagram formation principles (Ahmadi and Parhizgar 2017: 23). Moreover, their RST analyses seem to go against Mann et al.'s (1992: 45) observation that 'all of RST is pre-realizational, since it makes statements about how such meanings and intentions are structured and combined, but not about how they are realized' [emphasis in original]. Ahmadi and Parhizgar (2017) posited relations which were linked to linguistic expressions in the texts (e.g., the relation 'Question' linked rhetorical questions to their nuclei). These methodological shortcomings notwithstanding, Ahmadi and Parhizgar (2017) is a valuable contribution to research on RST's potential for coherence-break identification because it i) assessed inter-judge agreement via Cohen's Kappa, ii) suggested that RST analysis can help locate coherence breaks not only in paragraphs but also in essays, and iii) examined essays written by EFL learners with an L1 other than Mandarin Chinese.

The findings of the present study are likely to be more valid than those of Ahmadi and Parhizgar (2017) because it a) avoids the methodological shortcomings summarised above, b) examines not only inter- but also intra-judge agreement and also via more than one statistical test (see the discussion in Section 4.2.2), and c) examines inter- and intra-judge agreement not only in terms of Marcu et al.'s (1999) criteria but also in terms of coherence measures calculated on the basis of coherence breaks in a text (see Section 5.5).

Although Wolf and Gibson (2005) did not locate coherence breaks through RST analysis, it will be reviewed here because its claims threaten the legitimacy of the present study. According to Wolf and Gibson (2005) crossed dependencies, that is, schemas which share units of analysis, and units with multiple parents occur naturally in text and, consequently, tree diagrams such as RST ones should be replaced by diagrams which allow such links among units. Wolf and Gibson's (2005) claim is relevant to the present study because it means abandoning the RST diagram formation principle of Uniqueness (see section 2). Egg and Redeker (2008, 2010) refuted Wolf and Gibson's (2005) argument by

indicating that the text segments Wolf and Gibson (2005) categorised as crossed dependencies and multiple parents can be analysed via tree diagrams.

4 Method

The method consists of me and the research assistant reading the paragraphs which were our data (see Section 4.1), keeping notes about coherence breaks we perceived intuitively in each paragraph, and analysing the paragraphs via RST. Before this main phase of the study, the research assistant was training in the use of RST.

4.1 Data

This paper reports on research using examination data provided by the LTTC, namely, 84 paragraphs written by Taiwanese lower-intermediate learners of English in Writing Task 2 of the Intermediate GEPT examination. 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 (LTTC-ELC) (Huang et al. 2010; Cheung et al. 2011).³

Examinees read topic rubrics in Chinese. Table 1 presents their English translations.

Table 1 about here

³ The LTTC-ELC contains approximately 2 million words and consists of language samples by Taiwanese learners of English who have sat the Intermediate and High-Intermediate GEPT. A portion of the LTTC-ELC is freely searchable by the public via an online concordancer at www.lttcelc.org.tw/index.php.

If we categorise topics in terms of their communicative purpose and posit Welrich's (1976) distinction among 'descriptive', 'narrative', 'expository', 'instructive', and 'argumentative' text types, most topics require the creation of mixed text types (see Virtanen 1992) as they cannot fall neatly under a single category. All topics except 'Nearsighted' and 'Helping others' ask for a description of a preferred thing or person and an explanation for this preference. Therefore, all topics except two require partly descriptive and partly expository texts. 'Nearsighted' requires an expository text because it asks for the reasons for nearsightedness among Taiwanese children and ways to avoid nearsightedness; although it can be associated with a simple text type, even this topic asks for two kinds of content, because test-takers have to first answer a 'why' and then a 'how/what' question. 'Helping others' requires a partly argumentative, partly narrative text because it requires an argument for or against an opinion and then a description of an event.

Texts were equally distributed between those which had achieved one of the two highest score bands to ensure that the data that would be analysed would vary in terms of coherence error types. In this examination texts were scored on a scale from 0 to 5. Band A includes texts which received scores between 4 and 5 and band B those which received scores between 3 and 3.9. Texts that received band C (2 to 2.9) or D scores (0 to 1.9) were not analysed in this study because its pilot study indicated that such texts were likely to include many linguistic errors and, consequently, would be too difficult to analyse.

To keep text length as constant as possible, texts that were between 100 and 140 words long were used. The cut-off of 100 words was adopted because the pilot study suggested that shorter texts yielded too few RST units for coherence breaks to occur in them.

To examine coherence breaks in texts written on more than one topic, the 84 texts were equally distributed across the six topics. In other words, 7 band A and 7 band B texts were analysed per topic.

4.2 RST analysis

The method involved the analysis of all 84 texts by me and only 25 of them by a research assistant, independently. We used RSTTool (O'Donnell 1997, 2000), a free software programme designed for RST text annotation.⁴ The units of analysis were, as in most studies which analysed texts via RST, clauses except when they were complements of prepositions and verb objects. Specifically, the unit of analysis was, as in Carlson and Marcu (2001), the clause with the exception of clauses which are complements of prepositions or verb objects.

32 relations were used to tag texts in this study on the basis of the pilot study. They include a) the original 23 posited by Mann and Thompson (1987, 1988) except 'Volitional result'; b) five suggested by Mann in his former website (namely, 'Joint', 'Preparation', 'Unconditional', 'Unless' and 'List');⁵ and c) seven from Carlson and Marcu's (2001) set (namely, 'Comment', 'Conclusion', 'Topic-shift', 'Manner', and 'Explanation-argumentative').

4.2.1 RST analysis: training phase

The main analysis was preceded by the RST familiarisation of the research assistant, a recent PhD graduate specialising in EFL writing but unfamiliar with RST. She was first provided

⁴ This software can be downloaded from www.wagsoft.com/RSTTool/section2.html.

⁵ The relations adopted from Mann and Thompson (1987, 1988) and from Mann's former website can be found at www.wagsoft.com/RSTTool/RSTDefs.htm.

with training materials (i.e., a document summarising the aims and method of the project, electronic copies of Skoufaki 2009, Mann and Thompson 1988, Taboada and Mann 2006, and a table listing the coherence relations used in the current study and constraints on their use). Then she was asked to RST-tag electronic texts pre-segmented into RST units by me.

She and I independently tagged four texts per week. The texts used for training were different to those in the main phase of the study but were selected from the Intermediate level of the LTTC-ELC and were answers to the same topic rubrics as the main-phase texts. The assistant was also provided with the English translation of the rubrics (see Table 1).

At the end of each week she was informed of the inter-judge agreement results for each text and of cases where she had used a relation inappropriately, had not added a span to the RST diagram where it should have occurred, or had not taken into account a topic rubric while analysing a text.

4.2.2 Inter-judge agreement in the training phase

Inter-judge agreement, which the training phase aimed to increase to an acceptable level, was assessed via Marcu et al.'s (1999) approach.⁶ Marcu et al. (1999) compared inter-judge agreement via Siegel and Castellan's (1988) version of the Kappa test, a generalisation of Scott's pi test (Scott 1955) for more than two judges. In this test the probability that an item is assigned to a particular category is the same among judges, a computation considered

⁶ Although I am aware of the benefits that the new qualitative comparison by Iruskieta et al. (2013, 2015) has over Marcu et al.'s (1999) quantitative comparison, inter- and intra-judge agreement was assessed only quantitatively. Iruskieta et al. (2013, 2015) had not appeared when this study was conducted nor was I otherwise aware of their novel qualitative approach to RST inter- and intra-judge agreement.

inappropriate for discourse coding tasks (e.g., Wiebe et al. 1999). In this study Cohen's Kappa was used as the main agreement measure because it computes the probability of chance agreement between judges based on each judge's own distribution of categories. Scott's pi was also used so that Kappa scores in this study would be comparable with those of studies which calculated RST tagging agreement using Siegel and Castellan's Kappa.

Both tests suffer from the 'prevalence problem': if one/more categories occur much more often than the rest, the probability of chance agreement increases and, in turn, the coefficient value decreases for a given $P(A)$. Since some coding categories were used much more frequently than others, the agreement measure corrected for prevalence and suggested by Byrt et al. (1993) was also used.

Following Marcu (2000), inter-judge comparisons were made in terms of

- a) whether a unit or set of units constituted a 'functional element', that is, whether a unit or a set of units functioned as a nucleus or a satellite;
- b) the 'nuclearity' of a unit or set of units, that is whether a unit or set of units formed a nucleus or a satellite or did not form a functional element; and
- c) 'relations', that is, which, if any, out of the 32 relations used in this study applied to the span of units.

As data distributions were not normal for any week of training, Table 2 gives the median score for each inter-judge agreement measure per level of comparison and week of training. All Cohen's Kappa and Scott's pi tests were significant at $p < .001$.

Table 2 about here

Table 2 suggests that inter-judge agreement was quite stable across weeks for functional elements and nuclearity. Every week agreement was lower for relations than for functional

elements or nuclearity. Because other RST studies (e.g., Carlson et al. 2001) suggest that even months of training do not result in high inter-judge agreement for relations coding, the level of agreement in week 5 was considered satisfactory and training stopped then.

4.2.3 RST-analysis: main phase

RST analyses of the data summarised in Section 4.1 were conducted. To assess inter-judge agreement, the research assistant (J2) analysed 25 of the 84 texts I (J1) analysed. To assess intra-judge agreement, a month after finishing the initial text analysis J1 reanalysed 17 texts and J2 reanalysed 5 texts.

4.3 Measures of coherence according to RST analysis

The RST diagrams J1 had constructed were examined for indications of violations of RST diagram-formation principles. Because the number of RST units varied among paragraphs, a quantitative analysis of the number of violations of RST-diagram formation principles per paragraph would not be valid. Two measures of coherence were calculated.

Density of diagram-formation principle violations was a calculation of the number of RST-diagram formation principle violations in a text, normalised by 10 RST units. It was calculated as the total number of RST-diagram formation principle violations in a text divided by its total number of RST units and multiplied by 10.

Units affected by diagram-formation principle violations indicated how many RST units were affected by these violations in a text normalised by 10 RST units. It was calculated as the total number of units affected by these violations in a text divided by its total number of RST units and multiplied by 10.

5 Results

5.1 Inter- and intra-judge agreement

Table 3 summarises the results of interjudge agreement measures. All Cohen's Kappa and Scott's pi tests were significant at $p < .001$.

Table 3 about here

Inter-judge agreement was at the same level as during the training phase; in both periods median agreement was approaching .8 on functional elements, .7 on nuclearity, and .6 on relations.

Table 4 summarises the results of intra-judge agreement measures. All Cohen's Kappa and Scott's pi tests were significant at $p < .001$. According to all measures, intra-judge agreement was at the same level as inter-judge agreement.

Table 4 about here

5.2 Coherence breaks indicated by the violation of RST-diagram formation principles

The comparison between the RST analyses and the coherence breaks intuitively located by me and the research assistant suggests that the violation of any principle of RST-diagram formation indicates a coherence break. Figure 3 summarises these findings.

Figure 3 about here

‘Dangling’ (groups of) units were not linked to the earlier section of an RST diagram because their content seemed unexpected given the writing topic of a text. The resulting RST diagrams violate the Connectedness and Completedness principles. Figure 4 gives an example of a dangling schema.

Figure 4 about here

The extract in Figure 4 comes from an ‘Invention’ paragraph. Units 12-14 were considered irrelevant because although the topic was not argumentative, these units address the reader. In fact, in four of the 17 occurrences dangling schemas, the dangling schema arose from an address to the reader. This finding is consistent with the finding that EFL learners use first and second personal pronouns significantly more frequently than native English speakers (e.g., Behnam et al. 2014).

Dangling units/schemas which were not the last or first in a text brought about other dangling units/schemas. For example, in Figure 5 units 7-11 dangle because urging the reader to do something – through the ‘Motivation’ relation – is not expected given the writing topic; this topic asked test-takers to state whether they agree that helping others is the source of happiness and explain why/why not.⁷ Because these dangling units occurred in the middle of the text, units 1-6 and 12-17 dangled too.

Figure 5 about here

⁷ The diagram in Figure 5 is presented horizontally instead of vertically due to space limitations.

A relation between units which belong to different schemas violates the Uniqueness principle. This kind of relation cannot be illustrated diagrammatically because RSTTool autocorrects such diagram abnormalities. However, one can consider Figure 6. It shows the first lines of a paragraph on the 'Exotic food' topic; the units appear within square brackets and are numbered for ease of reference to them in the discussion that follows.

Figure 6 about here

Unit 2 restates information given in unit 1, so 2 is the satellite and 1 the nucleus of a 'Restatement' relation. Unit 3 expresses an 'Evaluation' of the situation presented in the schema formed by units 1 and 2. Unit 4 exemplifies the countries whose food the Taiwanese can eat in Taiwan, so it is the satellite of an 'Elaboration' relation and 1 is the nucleus. This relation is problematic because unit 4 intrudes the sub-diagram already formed by units 1 and 2.

Relations between schemas with intervening schemas constitute violations of the Adjacency principle. For example, the extract in Figure 7 comes from a 'Future invention' paragraph. Units 16-17 are the 'Summary' of units 7-13. This Summary relation violated the Adjacency principle because units 14-15 are between the linked 7-13 and 16-17 units.

Figure 7 about here

Table 5 shows the total instances of each kind of violation of RST-diagram formation principle across the texts. It is evident that the vast majority of these violations are those of the principles of Connectedness and Completedness.

Table 5: Total instances of the violation of each RST-diagram formation principle in all texts

Completedness/ Connectedness violations	Uniqueness violations	Adjacency violations
17	1	2

5.3 Potential coherence breaks indicated in RST diagrams but not due to the violation of RST-diagram formation principles

The RST analysis also indicated inductive content order in five texts. In the top row of typical RST diagrams arrows in satellite relationships point to a unit or schema in the beginning of the diagram; this right-to-left directionality of the arrows is due to the deductive content order typical in English paragraphs, where a sentence or group of sentences at the beginning of a paragraph expresses the paragraph's main statement and the rest provide detailed information evidence or arguments to support it.

Since the current study does not contain a control group of paragraphs written on the same topics by native speakers of American English (the dialect taught to Taiwanese learners), it does not examine whether the paragraphs have inductive content order significantly more frequently than those written by native speakers. It only points to the potential of RST diagrams to suggest this order in a paragraph.

5.4 Coherence breaks that could not be identified through RST analysis

In the RST literature, relations atypical of specific text types are considered problematic for coherence (e.g., Azar 1999). Although while analysing the paragraphs I considered some relations problematic for coherence, RST diagrams did not flag them as coherence breaks. For example, the extract in Figure 8 is from an ‘Exotic food’ paragraph. In units 6-15, the author narrates a positive dining experience at an Italian restaurant. Therefore, an ‘Elaboration’ relation links units 6-15 to units 3-5. The topic rubric’s requirement to ‘give reasons to explain why you like the food’ would have been met more directly through relations such as ‘Explanation-argumentative’ or ‘Evidence’. However, one should keep in mind that not all researchers consider the inappropriate use of a coherence relation as problematic for coherence (see Section 3.2).

Figure 8 about here

5.5 Inter- and intra-judge agreement in terms of coherence measures

Table 6 summarises the descriptive statistics for the coherence measures used in this study for the 25 texts analysed independently by the judges. The data were positively skewed because according to either judge only six texts contained violations of RST-diagram formation principles. The median score was 0 in both measures for both judges.

Table 6 about here

Cronbach's α was negative for both coherence measures (-.62 for break density; -.58 for units covered). This finding can be attributed to the lack of agreement on which texts had coherence breaks between the two judges; although for each judge six texts contained coherence breaks, all of these texts differed between judges. This finding is reminiscent of Watson Todd et al.'s (2004) finding that agreement about the location of coherence breaks varied among teachers. Both studies suggest that coherence break identification is highly subjective.

Cronbach's α suggested that intra-judge agreement was high for J1 – .83 for break density and .97 for units covered. For J2 Cronbach's α could not be estimated because all the RST diagrams she created followed all RST-diagram formation principles. This finding may be due to the fact that J2 reanalysed only five texts.

5.6 Coherence measures according to RST analyses for all texts

Table 7 presents descriptive statistics for the coherence measures which resulted from RST-diagram formation violations in the RST diagrams I had made. Both measures were positively skewed with a median of 0.

Table 7 about here

Descriptive statistics for both measures are very low. Given the observations in Sections 4.3 and 4.4, these findings suggest that RST analysis can point to only a few kinds of coherence breaks.

6 Conclusion, future research and pedagogical implications

This study examined the potential of RST to indicate coherence breaks by virtue of the violation of its diagram-formation principles. The findings suggest that the violation of any diagram-formation principle leads to coherence breaks.

The coherence breaks discussed in this paper were located only by me, hence these findings should be compared with coherence breaks identified by others as well. This gap in the study was filled by Skoufaki (in preparation). It examined how much the coherence breaks suggested by the violation of RST-diagram formation principles in 36 paragraphs in the present study map onto i) the coherence scores and ii) comments about coherence breaks that EFL teachers added to these paragraphs. The results suggest a significant positive relationship between coherence measures based on teachers' comments on coherence breaks and those based on the RST analysis for one third of the teachers. Irrelevant content and inductive order were identified by some of the teachers as well as through RST analysis. However, the findings also suggested a lack of a relationship between the teachers' scores for coherence and RST-based coherence-break identification. This finding could be interpreted as indicating the highly subjective nature of coherence judgments (e.g., Watson Todd et al. 2004).

The highly subjective nature of coherence judgments is also evident from the fact that in this study J1 and J2 produced RST diagrams with violations of RST-diagram formation principles for *different* paragraphs. The fact that RST analysis relies on the analyst's intuition

about which coherence relations exist in the text has the advantage of being realistic about the nature of coherence but this also means that different analysts may produce RST diagrams which differ in terms of the coherence breaks they suggest. This study should be replicated to assess the generalisability of the findings.

The present study should be replicated also because it analysed only paragraphs and a reader may be satisfied with the coherence of a text at paragraph level but not at whole-text level (e.g., see Schiffner 2014). Research locating coherence breaks in whole texts (see Ahmadi and Parhizgar 2017) and research using larger units of analysis such as paragraphs (see O'Brien 1995) or text sections (see Gruber 2006) *and*, as the current study did, examining inter- and intra-judge agreement is necessary.

The rare occurrences of violations of RST-diagram formation principles in the data may suggest that no single theory of coherence can predict all kinds of coherence judgments. According to this interpretation of the findings, the potential of various discourse theories to coherence-break prediction and their relationship to coherence judgments should be compared. The ultimate goal of this enterprise would be to combine elements of different theories to create a hybrid theory with more explanatory power. Although Watson-Todd et al. (2004, 2007) conducted such research, all the coherence approaches they put to the test were related to coherence marking and topic development. A principled way of examining various theories' potential to predict coherence breaks would test the predictions of theories functioning at different discourse levels, namely, theories which deal with how coherence is signalled via cohesive devices (e.g., Centering Theory), theories which examine topic development (e.g., topical structure analysis), theories which examine relations between propositions (e.g., RST), and theories which examine overall text structure (e.g., genre analysis).

This study has implications for language assessment and pedagogy. The Common European Framework of Reference for Languages level descriptors for coherence and cohesion in writing are vague (see Council of Europe 2017: 141); rating scales for coherence used in classroom and examination writing assessment are more vague than those for other criteria (e.g., Burstein et al. 2013); exam markers find marking writing scripts for coherence harder than for other criteria (Cotton and Wilson 2008); and teachers (e.g., Gonzáles et al. 2001) and AWE software (e.g., Skoufaki 2009) face difficulties in providing learners with feedback on coherence errors. The present study's finding that RST can predict coherence breaks may inform relevant training for language teachers, exam markers, and language learners.

Ryan (1984) and O'Brien (1995) suggest that writers' understanding of coherence affects how coherent their writing is. Therefore, the findings of this study can inform teaching aiming to improve EFL learners' understanding of coherence. Because of their visual impact, RST diagrams could help raise EFL learners' awareness of some kinds of coherence breaks. Ahmadi and Parhizgar (2017) argue that language teachers should be trained in RST analysis because this training would improve their assessment of students' writing for coherence and their teaching about coherence. However, since teachers' free time is limited, perhaps they could just discuss with their students RST diagrams which exemplify coherence breaks, such as those in the present study.

References

- Ahmadi, Alireza and Salma Parhizgar. 2017. Coherence Errors in Iranian EFL Learners' Writing: A Rhetorical Structure Theory Approach. *Journal of Language Horizons* 1(1). 9-37.

- Azar, Moshe. 1999. Argumentative Text as Rhetorical Structure: An Application of Rhetorical Structure Theory. *Argumentation* 13(1). 97–114.
- Barzilay, Regina & Mirella Lapata. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics* 34(1). 1–34.
- Bechtel, William. 1988. *Philosophy of science: An overview for cognitive science*. Hillsdale, NJ: Erlbaum.
- Behnam, Biok, Fathemeh Mirzapour & Mohammad Amin Mozaheb. 2014. Writer's presence in English native and non-native speaker research articles. *Procedia – Social and Behavioral Sciences* 98(6). 369–374.
- Burstein, Jill, Karen Kukich, Susanne Wolff, Chi Lu, Martin Chodorow, Lisa Braden-Harder, & Mary D. Harris. 1998. Automated essay scoring using a hybrid feature identification technique. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1, Montreal, Canada, 10–14 August 1998*, 206–210. Stroudsburg, PA: Association for Computational Linguistics.
- Burstein, Jill, Joel Tetrault & Martin Chodorow. 2013. Holistic Annotation of Discourse Coherence Quality in Noisy Essay Writing. *Dialogue and Discourse* 4(2). 34–52.
- Byrt, Ted, Janet Bishop & John B. Carlin. 1993. Bias, Prevalence, and Kappa. *Journal of Clinical Epidemiology* 46(5). 423–429.
- Carlson, Lynn & Daniel Marcu. 2001. *Discourse Tagging Manual*. ISI Tech Report ISI-TR-545.
- Carlson, Lynn, Daniel Marcu & Mary E. Okurowski. 2001. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue – Volume 16, Aalborg, Denmark 1–2 September 2001*, 1–10. Stroudsburg, PA: Association for Computational Linguistics.

- Cheung, Hintat, Jessica Wu, Zhao-ming Gao & Siaw Fong Chung. 2011. The construction of the LTTC English Learner Corpus: Progress report of NTU-LTTC Joint Research Project on Language Teaching and Language Testing.
- Cotton, Fiona & Kate Wilson. 2011. An Investigation of Examiner Rating of Coherence and Cohesion in the IELTS Academic Writing Task 2. *IELTS Research Reports* 12. 1–76.
- Council of Europe. 2017. *Common European Framework of Reference for Languages: learning, teaching and assessment. Companion volume with new descriptors*. <https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/168074a4e2> (accessed 29 November 2017).
- De Silva, Nishadi H. 2007. *A narrative-based collaborative writing tool for constructing coherent technical documents*. Southampton: University of Southampton PhD thesis.
- Egg, Markus & Gisela Redeker. 2008. Underspecified discourse representation. In Anton Benz & Peter Kühnlein (eds.), *Constraints in Discourse* (Pragmatics and Beyond New Series 172), 117–138. Amsterdam & Philadelphia: John Benjamins.
- Egg, Markus & Gisela Redeker. 2010. How complex is discourse structure? In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner & Daniel Tapias (eds.), *Proceedings of LREC2010, Malta, 17–23 May 2010*, 1619–1623. Valletta, Malta: European Language Resources Association.
- González, Virginia, Chia-Yin Chen & Claudia Sanchez. 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.
- Grosz, Barbara, Aravind Joshi & Scott Weinstein. 1995. Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics* 21(2). 203–226.
- Gruber, Helmut. 2006. Rhetorical Structure Theory and quality assessment of students' texts. *Information Design Journal* 14(2). 114–129.

- Hoey, Michael. 1991. *Patterns of Lexis in Text*. Oxford: Oxford University Press.
- Huang, Chu-Ren., Winnie Cheng, Hintat Cheung, Yasunari Harada, Huaqing Hong, Skoufaki, Sophia, and Helen K.Y. Chen. 2010. English Learner Corpus: Global Perspectives with an Asian Focus. In Tien-En Kao and Yaofu Lin (eds.) *A New Look at Language Teaching and Testing: English as Subject and Vehicle*, 85-117. Taipei: Language Training and Testing Center.
- Iruskieta, Mikel, Iria da Cunha & Maite Taboada. 2015. A qualitative comparison method for rhetorical structures: identifying different discourse structures in multilingual corpora. *Language Resources and Evaluation* 49(2). 263–309.
- Iruskieta, Mikel, Arantza Diaz de Ilarraza & Mikel Lersundi. 2013. Establishing criteria for RST-based discourse segmentation and annotation for texts in Basque. *Corpus Linguistics and Linguistic Theory* 11(2). 302–334.
- Mann, William C., Christian M. I. M. Matthiessen, & Sandra A. Thompson 1992. Rhetorical Structure Theory and text analysis. In William C. Mann and Sandra A. Thompson (eds.), *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*, 39–78. Amsterdam & Philadelphia: John Benjamins.
- Mann, William C. & Sandra A. Thompson. 1987. Rhetorical Structure Theory: A Theory of Text Organization (No. ISI/RS-87-190). Marina del Rey, CA: Information Sciences Institute.
- Mann, William C. & Sandra A. Thompson. 1988. Rhetorical Structure Theory: Towards a functional theory of text organisation. *Text* 8(3). 243–281.
- Marcu, Daniel. 2000. The Rhetorical Parsing of Unrestricted Texts: A Surface-Based Approach. *Computational Linguistics* 26(3). 395–448.
- Marcu, Daniel, Estibaliz Amorrortu & Magdalena Romera. 1999. Experiments in constructing a corpus of discourse trees. In Marilyn Walker (ed.), *Proceedings of the*

- ACL Workshop 'Towards Standards and Tools for Discourse Tagging', Maryland, 21 June 1999*, 48–57. Stroudsburg, PA: Association for Computational Linguistics.
- Miltsakaki, Eleni & Karen Kukich. 2004. Evaluation of text coherence for electronic essay scoring systems. *Natural Language Engineering* 10(1). 25–55.
- O'Brien, Teresa. 1995. Rhetorical structure analysis and the case of the inaccurate, incoherent source-hopper. *Applied Linguistics* 16(4). 442–482.
- O'Donnell, Michael. 1997. RSTTool: An RST Analysis Tool. In *Proceedings of the 6th European Workshop on Natural Language Generation*, 24–26 March, 1997 Gerhard-Mercator University, Duisburg, Germany.
- O'Donnell, Michael. 2000. RSTTool 2.4 – A Markup Tool for Rhetorical Structure Theory. *Proceedings of the International Natural Language Generation Conference (INLG'2000)*, Mitzpe Ramon, Israel, 13-16 June 2000, 253-256.
- Pitler, Emily & Ani Nenkova. 2008. Revisiting Readability: A Unified Framework for Predicting Text Quality. In Sebastian Padó (ed.), *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 186–195. Edinburgh, Scotland.
- Sanders, Ted. 1997. Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes* 24(1). 119–147.
- Schiftner, Barbara. 2014. (Non-)signalling of coherence structures in English learner writing. In Helmut Gruber & Gisela Redeker (eds.) *The pragmatics of discourse coherence: theory and applications*, 243–266. Amsterdam: John Benjamins.
- Scott, William A. 1955. Reliability of content analysis: The case of nominal scale coding. *Public Opinion Quarterly* 19(3). 321–325.
- Siegel, Sidney & N. John Castellan. 1988. *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

- Skoufaki, Sophia. (in preparation) Looking for common ground between theory and practice: coherence breaks according to Rhetorical Structure Theory and teachers in the writing of EFL learners.
- Skoufaki, Sophia. (2009) An exploratory application of Rhetorical Structure Theory to detect coherence errors in L2 English writing: possible implications for Automated Writing Evaluation software. *International Journal of Computational Linguistics and Chinese Language Processing* 14(2). 181-203.
- Stede, Manfred. 2008. Disambiguating rhetorical structure. *Research on Language and Computation* 6(3). 311–332.
- Swales, John. 1990. Nonnative speaker graduate engineering students and their introductions: Global coherence and local management. In Ulla Connor & Ann M. Johns (eds.) *Coherence in writing. Research and pedagogical perspectives*, 187-207. Alexandria, Virginia: Teachers of English to Speakers of Other Languages.
- Taboada, Maite & William C. Mann. 2006. Rhetorical Structure Theory: looking back and moving ahead. *Discourse Studies* 8(3). 423–459.
- Virtanen, Tuija. 1992. Issues of text typology: Narrative – a ‘basic’ type of text? *Text* 12(2). 293–310.
- Wang, Yongqing, Michael Harrington & P. White. 2012. Detecting Breakdowns in Local Coherence in the Writing of Chinese English Speakers. *The Journal of Computer Assisted Learning* 28(4). 396–410.
- Watson Todd, Richard. 1997. Textual patterns in teachers’ eliciting. *RELC Journal* 28(1). 1–14.
- Watson Todd, Richard. 1998. Topic-based Analysis of Classroom Discourse. *System* 26(3). 303–318.

- Watson Todd, Richard. 2016. *Discourse topics* (Pragmatics and Beyond New Series 269). Amsterdam & Philadelphia: John Benjamins.
- Watson Todd, Richard, Patteera Thienpermpool & Sonthida Keyuravong. 2004. Measuring the Coherence of Writing using Topic-based Analysis. *Assessing Writing* 9(2). 85–104.
- Watson Todd, Richard, Somreudee Khongput, & Pornapit Darasawang. 2007. Coherence, Cohesion and Comments on Students' Academic Essays. *Assessing Writing* 12(1). 10–25.
- Werlich, Egon. 1976. *A text grammar of English*. Heidelberg: Quelle & Meyer.
- Wiebe, Janyce M., Rebecca F. Bruce & Thomas P. O'Hara. 1999. Development and use of a gold-standard data set for subjectivity classifications. In *ACL99, Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 246–253. College Park, MD.
- Williams, James D. 1985. Coherence and cognitive style. *Written Communication* 2(4). 473–491.
- Wolf, Florian & Edward Gibson 2005. Representing Discourse Coherence: A Corpus-based Analysis. *Computational Linguistics* 31(2). 249–287.

Figure 1

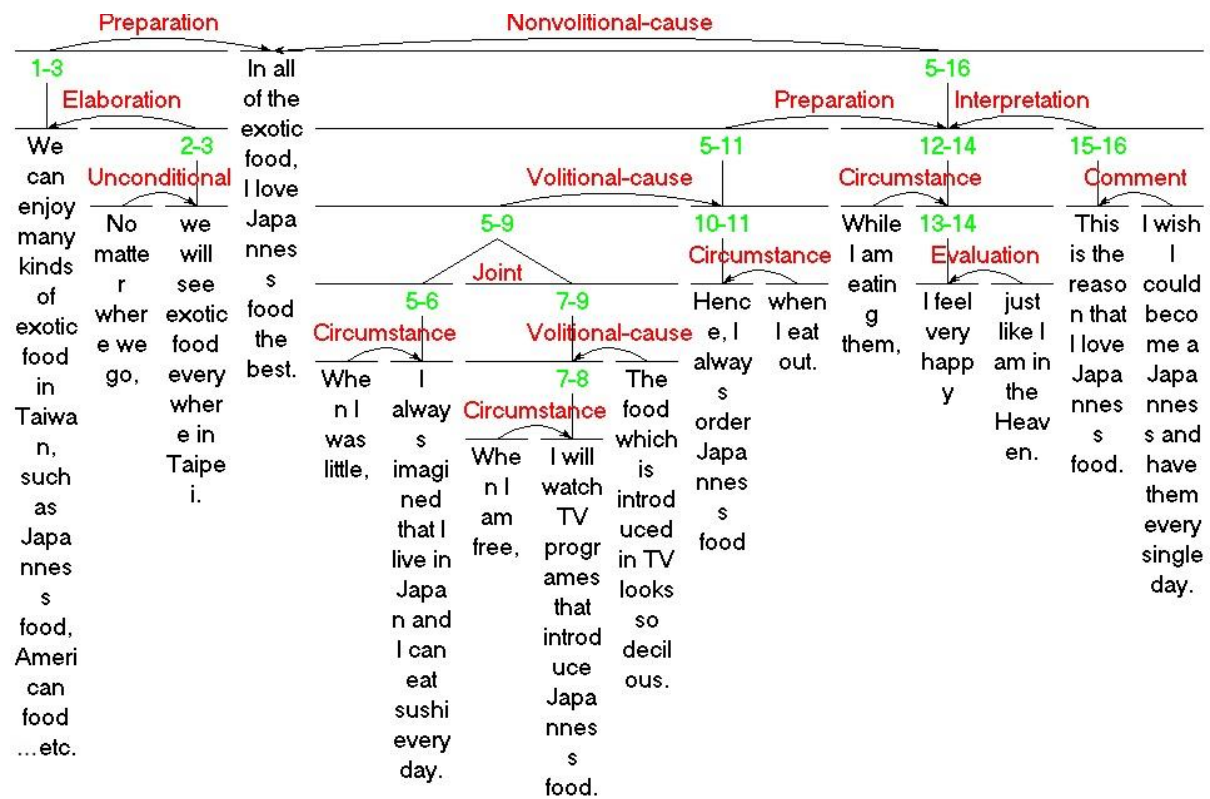


Figure 2

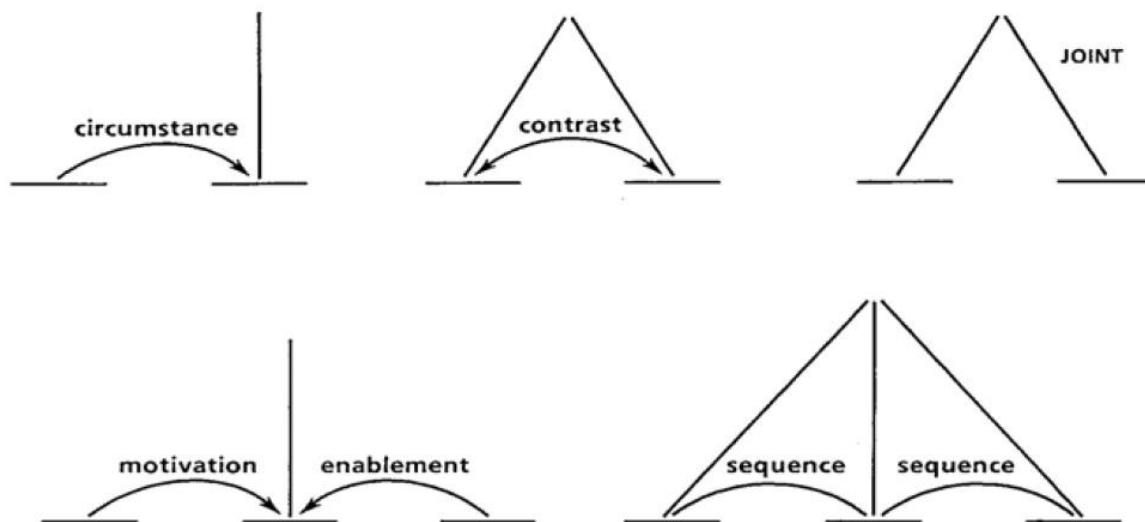


Figure 3

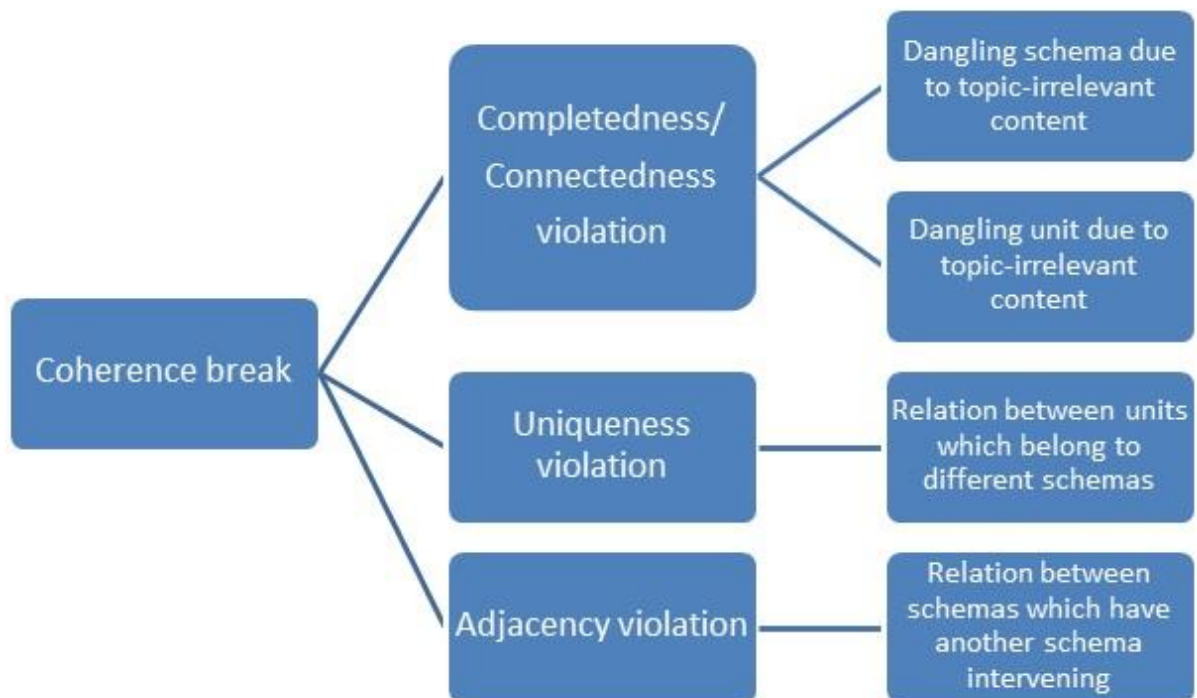


Figure 4

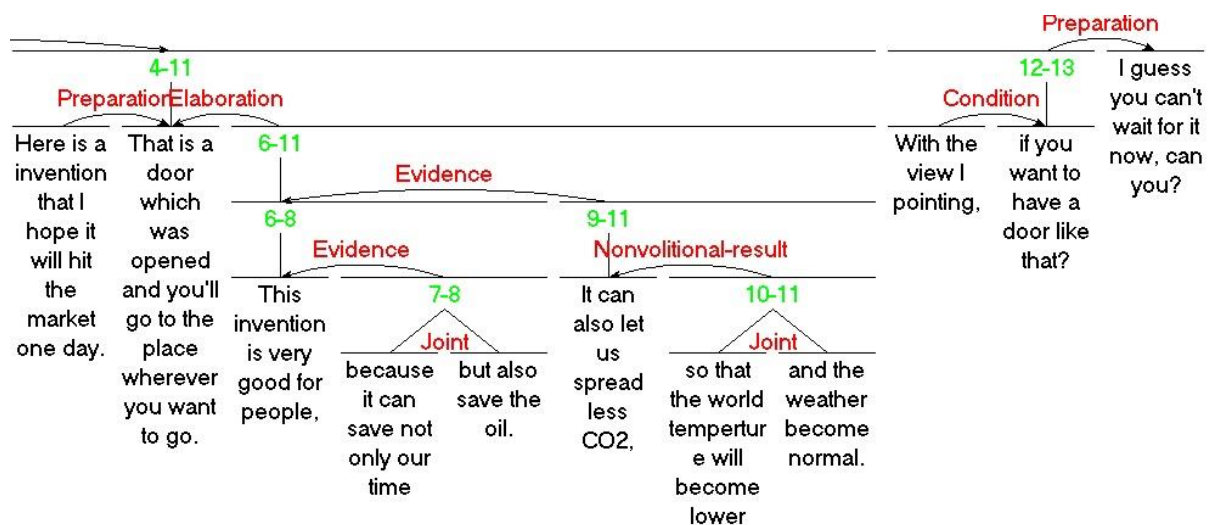


Figure 5



Figure 6

[1. We can taste a lot of foods from other countries.] [2. They are gathered in this small island.] [3. That's so amazing.] [4. Like Japan, America, Tailand and more.]

Figure 7

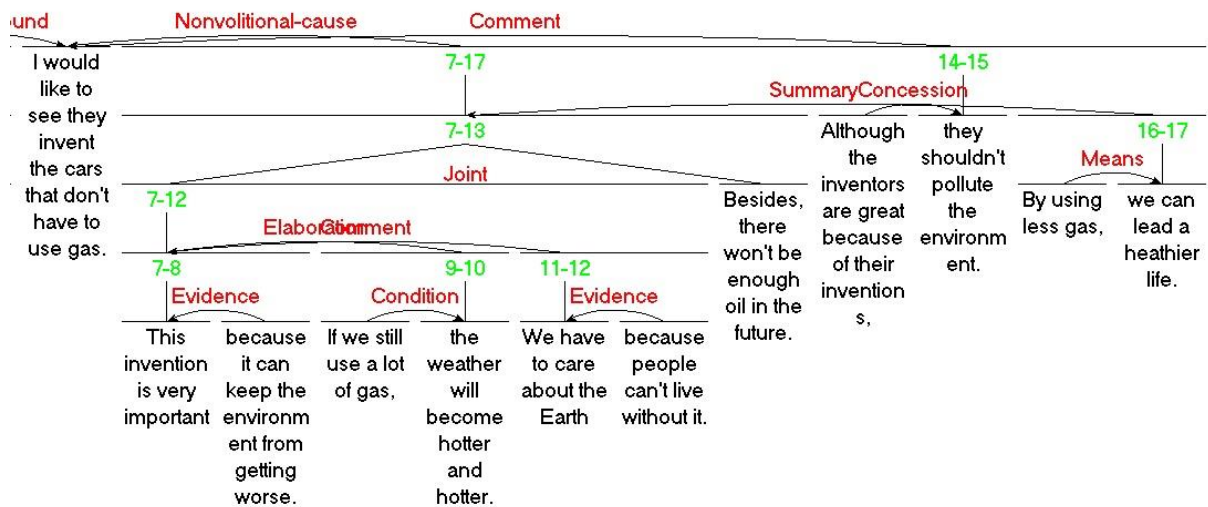


Figure 8

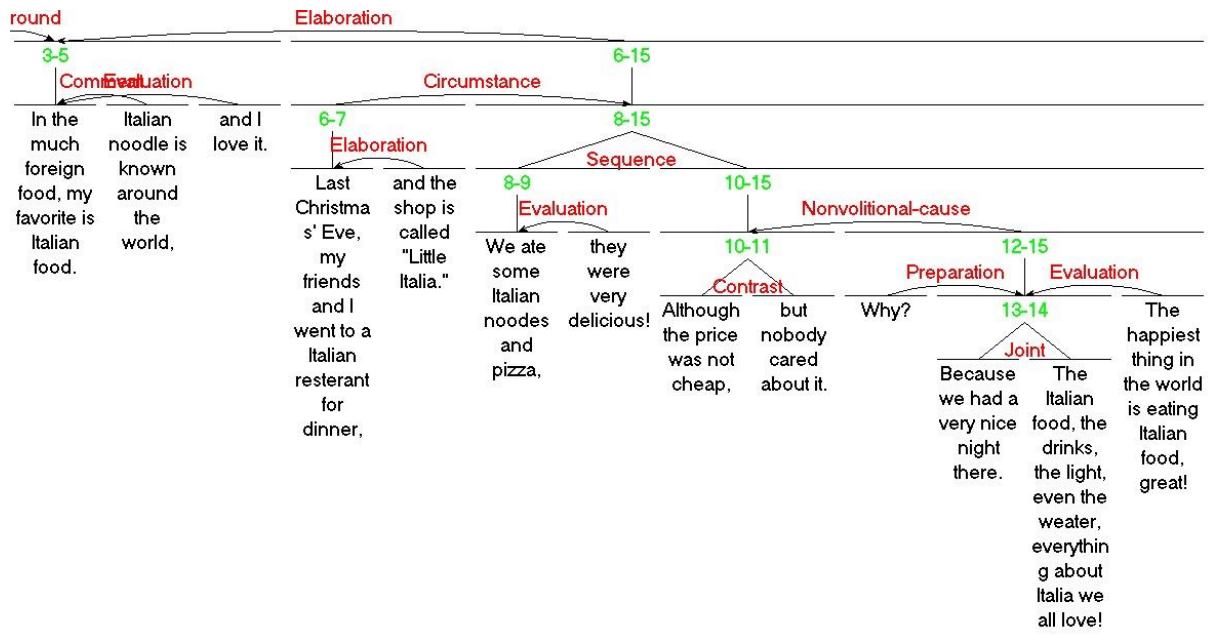


Figure 1: RST diagram of a text

Figure 2: RST schemas posited by Mann and Thompson (1987, 1988); figure adapted from Mann and Thompson (1987: 7)

Figure 3: Violations of RST diagram-formation principles which map onto coherence breaks

Figure 4: Extract from an RST diagram where the last schema is dangling

Figure 5: RST diagram of a text with units expressing irrelevant content in its middle

Figure 6: Extract from a paragraph illustrating the violation of the Uniqueness principle

Figure 7: Extract from an RST diagram illustrating the violation of the Adjacency principle

Figure 8: Extract from an RST diagram which includes an unexpected relation given the task requirements specified by the topic rubric

Table 1: English translations of the topic rubrics in the Intermediate GEPT Writing Task 2(adapted with permission from www.lttcelc.org.tw/docs/topics.pdf)

Topic code	English translations of the Chinese topic rubrics
Nearsighted	More and more elementary school students in Taiwan suffer from nearsightedness. Write an essay in which you: (1) explain the possible reasons for this phenomenon, and (2) provide some ways to prevent nearsightedness.
Idol	Many people have an idol who can be a film and television celebrity, an athlete, a writer, or even someone beside you. Write an essay in which you: (1) describe the idol you like or liked, and (2) give reasons to why you like(d) him/her.
Exotic food	We can taste a lot of exotic food in Taiwan. Write an essay in which you: (1) describe your favourite exotic food, and (2) give reasons to explain why you like the food.
Helping others	One Chinese proverb says, 'Helping others is the source of happiness.' Write an essay in which you: (1) express whether you agree with this proverb and provide specific reasons, and (2) describe one experience when you helped others.
Invention	Many inventions in history have changed the life of mankind. Write an essay in which you: (1) describe one invention that you wish to see in the future, and (3) explain the importance of this new invention.
Music	Music is an essential part of many persons' lives. Write an essay in which you: (1) explain the importance of music to you and provide specific reasons, and (2) describe your favourite type of music.

Table 2: Median scores for Cohen's Kappa, Scott's pi, and Byrd et al.'s formula for the comparison of functional elements, nuclearity, and relations between judges per week of RST coding training

Week	Functional elements			Nuclearity			Relations		
	K_{Co}	π	K_{cor}	K_{Co}	π	K_{cor}	K_{Co}	π	K_{cor}
1	.78	.78	.84	.62	.62	.74	.47	.46	.61
2	.71	.71	.81	.69	.69	.79	.53	.53	.66
3	.77	.70	.86	.74	.67	.83	.62	.58	.74
4	.80	.77	.89	.75	.75	.85	.69	.68	.81
5	.77	.77	.87	.69	.69	.79	.58	.58	.73

Table 3: Median scores for Cohen's Kappa, Scott's pi, and Byrd et al.'s (1993) formula for the comparison of functional elements, nuclearity, and relations between judges for 25 texts

Functional elements			Nuclearity			Relations		
K_{Co}	π	K_{cor}	K_{Co}	π	K_{cor}	K_{Co}	π	K_{cor}
.79	.79	.87	.69	.69	.87	.57	.57	.74

Table 4: Median scores for Cohen's Kappa, Scott's pi, and Byrd et al.'s (1993) formula for J1 and J2's intrajudge comparison of functional elements, nuclearity, and relations

Rater	Functional elements			Nuclearity			Relations		
	K_{Co}	π	K_{cor}	K_{Co}	π	K_{cor}	K_{Co}	π	K_{cor}
J1	.81	.82	.91	.78	.79	.89	.71	.72	.85
J2	.80	.80	.86	.74	.74	.84	.69	.69	.79

Table 5: Total instances of the violation of each RST-diagram formation principle in all texts

Completedness/ Connectedness violations	Uniqueness violations	Adjacency violations
17	1	2

Table 6: Mean and standard deviation (SD) for coherence measures in the judges' analyses of 25 texts

Judge	Density of diagram-formation principle violations	Units affected by diagram-formation principle violations
J1	.25 (.52)	1.4 (2.88)
J2	.18 (.35)	1.17 (2.61)

Table 7: Mean and standard deviation (SD) for coherence measures based on J1's RST analysis of all 84 texts

Units affected by diagram-formation principle violations	Units affected by diagram-formation principle violations
.36 (.6)	1.1 (2.26)