The handle http://hdl.handle.net/1887/69810 holds various files of this Leiden University dissertation.

**Author:** Brown, J.D.
**Title:** Using Rhetorical Structure Theory for contrastive analysis at the micro and macro levels of discourse: An investigation of Japanese EFL learners' and native-English speakers' writing
**Issue Date:** 2019-03-12
Chapter 3 | Methodology & Results

Introduction
As has been shown in Chapter 2, the majority of contrastive studies conducted between Japanese and English writing range from the objective, quantifiable measurements of sentence-level (micro) features (see Bryant, 1984; Casanave, 1994; Moriya, 1997; Nishigaki & Leishman, 2001; Oi, 1986) to the more interpretive observations of discourse analysis, i.e., macro-level (see Achiba & Kuromiya, 1983; Hirose & Sasaki, 1994; Ito, 2004; Matsumoto, 1995; Miyake, 2007; Sasaki & Hirose, 1996). Yet none thus far have been able to explicitly identify what could account for the incongruities in Japanese English writing outside grammatical and idiomatic errors. These studies have made it quite apparent that Japanese English writing is regularly regarded as illogical, ambiguous, and less coherent, especially in comparison to texts written by L1 English speakers (NESs), demonstrating issues of cohesion in Japanese English writing. Nevertheless, understanding why this is the case has proven to be more difficult.

Those that have looked at overall rhetorical/organizational patterns at the macro-level, tend to lean heavily on the hypothesis of different rhetorical expectations and conventions between English and Japanese in their explanations, frequently making overgeneralizations. Furthermore, these studies are often limited to relatively subjective approaches of discourse analysis, making their findings less reliable. Though there is plenty of anecdotal testimony from L2 learners suggesting that differences in values and expectations between languages and cultures exist (see, for example, Li, 1996; Matsuda, 2001; Shen, 1989), discrepancies in the literature and lack of empirical evidence make it difficult to come to any concrete conclusions or to pinpoint exactly with what conventions Japanese L2 writers struggle beyond the grammatical and idiomatic.

Other studies, which have attempted to analyze both micro- and macro-level features and provide more objective and quantifiable findings, have, in many cases, unfortunately, been guilty of inaccurate hypotheses, such as investigations of cohesive devices to describe coherence. These studies have failed to sufficiently explain incongruences in cohesion beyond the types and frequency of cohesive
devices used. More importantly, however, as discussed in Chapter 2, cohesion cannot easily account for coherence as cohesive devices at the micro-level do not necessarily play a role in shaping and constructing the coherence of a text at the macro-level.

We are therefore left with a dilemma: What exactly is causing Japanese English writing to be experienced as “illogical” and “incoherent” by NESs? Despite the numerous contrastive studies conducted to address this question, none have been able to provide an adequate answer. Accordingly, what is needed is a study that provides empirical and quantifiable evidence for a diagnosis that could plausibly account for why Japanese L2 English writing is simply not as “good” as that of NESs’ beyond grammatical accuracy, idiomatic errors, and macro-organization by comparing features in English texts written by Japanese L2 writers (specifically, Japanese English as a Foreign Language learners, JEFL) with those in texts written by NESs. We must then ask ourselves: What conventions are slipping through—not being addressed or focused on enough—in English education in Japan? What “un-articulated” values are buried beneath the language, culture, rhetoric, or whatever it may be that is preventing JEFL writers from achieving native-like competence in English writing beyond grammatical accuracy and idiomaticity? In other words, in addition to identifying differences, it is vital to pinpoint the specific conventions JEFLs need explicated.

This chapter reports on a comparison of 44 written texts between JEFLs (N = 22) and NESs (N = 22). By adopting a quantifiable measurement of discourse features taken from Rhetorical Structure Theory (RST), I attempt to identify the distinctive features of JEFL writers’ texts in comparison to those in the NESs’ texts to explicate features that have not been identified in past research but that could reasonably explain the lack of coherence in English texts written by JEFL writers. A Japanese text is also used as a point of reference to qualitatively observe and discuss similarities and differences of Japanese writing and JEFL English writing (to be done in Chapter 4). What follows is a description of the data, participants, analytical framework, and statistical methods used to collect and analyze the corpora. The descriptive statistical results and quantitative comparisons based on the analytical method used will be presented at the end.
Participants & Data

Source of Data
Four hundred texts (200 NES texts and 200 JEFL texts) were randomly selected from the International Corpus Network of Asian Learners of English or ICNALE (Ishikawa, 2015). ICNALE is “a collection of controlled essays and speeches by learners of English in 10 countries and areas in Asia as well as L1 productions by 350 native-English speakers” (Ishikawa, 2015). The corpus is divided into two categories: ICNALE-Spoken and ICNALE-Written. For the purposes of this study, I only selected data from the ICNALE-Written corpus and limited this search to texts written by NESs and JEFLs. How the native-English speakers’ texts were collected for the ICNALE-Written corpus is unclear, but the texts produced by Asian learners of English (ALEs) were collected following a four-point protocol: 1) Entry of personal information, 2) completion of questionnaire to measure motivation, 3) completion of a vocabulary test, and 4) submission of two essays.

Participants of ICNALE
Before submitting their writing to the ICNALE-Written Corpus, ALEs are asked to download an Excel file from the ICNALE website that consists of four pages, and, on the first page, enter their personal information, including name, country, age, sex, name of school, grade, major, years of studying English, and score and type of language proficiency test (see Image 3.1). On the next page, they are asked to complete a questionnaire to measure their motivation (see Image 3.2), followed by a vocabulary test consisting of 50 questions in order to assess the participant’s current vocabulary repertoire (see Image 3.3). The test is not timed, but dictionaries are not permitted.

Once these preliminary sections are completed, the participant is asked to write two 200-to-300-word essays based on the following two prompts:

Do you agree or disagree with the following statements? Use reasons and specific details to support your answer.

- It is important for college students to have a part-time job.
- Smoking should be completely banned at all restaurants in the country.
In an effort to create a reliable corpus and control the writing conditions (Ishikawa, 2015), the participant is asked to follow a set of instructions as listed on the ICNALE Project website:

1) Clarify your opinions and show the reasons for them and some examples.
2) You can use 20 to 40 minutes for each essay. This means that you have 40 to 80 minutes to complete two essays. Do not finish too early or spend too much time. Spend 20 to 40 minutes for each.
3) Never use any dictionary or reference tools.
4) Do not plagiarize anyone else's essays.
5) The length of your single essay should be from 200 to 300 words (not 200 to 300 letters). Too short or too long essays cannot be accepted. Please follow the rules rigidly. You can check the length using the word count function of MS Word.
6) You must conduct a spell check before completing your writing.
7) After conducting the spell check and word count, both of which are standard functions of MS Word, please copy your essays from MS Word to the ninth sheet of the Excel file. Also, you must record the word count of each essay at the bottom.

**Figure 3.1.** Screenshot of ICNALE-Written personal info page.
Figure 3.2. Screenshot of ICNALE-Written questionnaire page.

Figure 3.3. Screenshot of the first page of five of ICNALE vocabulary test.
Pre-determined Parameters
As mentioned earlier, it is not clear if the NESs were expected to follow a similar protocol to that of the ALEs when submitting their writing. The main issue I was concerned with, however, was ensuring that the texts to be compared are in fact comparable as this is an important factor to take into consideration when comparing corpora (Connor & Moreno, 2005; James, 1980; Moreno, 2008). James argued that this is of particular importance for contrastive purposes:

The first thing we do is make sure that we are comparing like with like: this means that the two (or more) entities to be compared, while differing in some respect, must share certain attributes. This requirement is especially strong when we are contrasting, i.e., looking for differences—since it is only against a background of sameness that differences are significant. We shall call this sameness the constant and the differences variables. (p. 169)

Accordingly, should any differences between corpora be identified, it is necessary to take a number of measures that can ensure “a background of sameness” on which the differences can be contrasted and thereby revealed to be significant or not. Therefore, though the texts in my study were randomly selected from ICNALE-Written, I set a few parameters for participants and their texts before incorporating them into my corpora.

Parameter 1
First, all texts in both corpora were to be written by undergraduates so that the ages and experiences of the NES and JEFL writers would be similar. Since the purpose of this study is to look at what could plausibly account for why Japanese English writing is regarded as less coherent and more illogical when compared to the writing of native-English speakers, it did not make sense to compare their writing to writing that had been revised, polished and/or published. After all, any piece of poorly organized and illogical writing can become more coherent and logical after multiple revisions. Furthermore, there are different expectations of “good” writing depending on the level of the writer. For example, what would be considered good writing at the high school level would not necessarily be considered good at the college level, and, at the same time, what would be considered good at the graduate level should not be expected from those at the undergraduate level. Accordingly, by limiting JEFL and NES texts written by undergraduates from the ICNALE-Written, I was able to reasonably ensure that in fact they were representative of a typical writer of that age, educational background, and experience. In this way, when I
compared the JEFL texts to the NES texts I was not setting the bar exceedingly and unrealistically high but establishing what the standard expectations are of “good” writing at their comparable level, i.e., the undergraduate level.

Parameter 2
Another important parameter was the content of the texts. As Hinkel (2002), Kegley (1986), Brown, Hilgers, and Marsella (1991) note, topics and rhetorical modes can cause a great deal of variance in text production. This is especially true in the case of L2 writers (Tsai & Cheng, 2009; Li, 2011). In an effort to avoid such extraneous variability, I chose to limit my search to the “part-time jobs” prompt. This is because the JEFL writers were more likely to perform better on this task due to the likelihood of familiarity with the topic, as well as the narrative-like rhetorical mode this task tends to lend itself towards. This belief is in accordance with Li (2011) who found that narrative tasks that are related to a writer’s personal experience elicit a higher-quality written product.

Furthermore, the more familiar the topic is to the writer the more likely the writer will be able to focus on the writing itself. According to Hinkel (2002), prompts can have a heavy influence on the textual and linguistic features of a text. In other words, if the topic is overly complicated, the text will demand more complex syntactic and grammatical structures. If not properly constructed, these could certainly affect the readability and, hence, the quality of the text. As the purpose of this study is to look beyond such surface-level errors, I thought it best to avoid circumstances where there would be a higher chance of such errors occurring. An additional concern, as will be discussed later in this chapter, is that texts with poor English are difficult if not impossible to annotate within Rhetorical Structure Theory (RST), since it is not possible to know the writer’s plausible intentions if the annotator cannot understand what the writer wrote. Selecting texts that were written on a topic that is simple and relatable helped to ensure that the writers were writing within their linguistic capacity. This parameter was also applied to the NES writers. This was done not for proficiency purposes but because comparing texts of the same topic would further ensure the comparability of those texts to one another as their structure is likely to be most similar.

Parameter 3
Because poor English language skills may negatively impact how effectively RST annotations can be applied to the text, it was also crucial to make sure that each JEFL text was produced by a writer with a relatively proficient command of the English language, as studies have shown that linguistic competence is positively
correlated to writing performance (Reed, 1992; Sasaki & Hirose, 1996; Tillema, 2012). Thus, another parameter set specifically for JEFL texts was that the writer had to demonstrate a B1 level or above in the Common European Framework (CEFR), at which level the writer should demonstrate the ability to “write straightforward connected texts on a range of familiar subjects within his field of interest, by linking a series of shorter discrete elements into a linear sequence” (Council of Europe, 2001). According to CEFR, writers below this level may “show coherence problems which makes the text hard to understand.” Therefore, limiting my corpus to JEFL texts that were written by writers who could effectively produce longer, more coherent texts assisted in providing me with the data I needed to investigate the rhetorical structure of a text beyond local features. CEFR band scores for individual writers are provided by ICNALE.

Two additional parameters were set for NES texts. First, since college level information is not available for native-English speakers in the ICNALE, I determined the age of the writer must be between 18 to 24 years of age, which is the standard age range for an undergraduate student (NCES, 2015). Based on this criterion, 100 of the 200 NES texts available in ICNALE-Written were selected.

The second parameter was that the texts must be regarded as “good” writing by other native-English speakers. To identify which texts are “good,” a questionnaire with all 100 texts and a one-to-five scale (one being poor and five being excellent) was designed on Google Forms and emailed to three native-English speakers, which implies that maximum score was 15. They were asked to intuitively evaluate the texts and rate each text according to the provided scale. “Good” texts had scores of 1.8 standard deviations or more above the mean (the mean for the 100 texts was 10.2). In other words, to be regarded as “good,” the texts must have received a 12 or better.

Data Collected for JEFL Corpus
Based on the above parameters for JEFL texts, 22 texts that met the criteria were pulled from the 200 JEFL texts that had been initially selected from the ICNALE-Written corpus. This resulted in a total number of 5763 words (median per text = 265) in the JEFL corpus. The total number of segments, i.e., EDUs, was 1173 (median per text = 26).
Data Collected for NES Corpus
The application of the above parameters to the NES corpus resulted in 44 texts, 25 of which were identified as “good” by native English speakers (12 out of 15 or above). For this study, however, I only ended up using 22 of those texts, which resulted in 4544 total number of words (median per text = 212.5) and a total of 673 EDUs (median per text = 15).

Japanese Texts
In order to have a point of comparison to which to refer back and observe whether the features and phenomena observed in the JEFL texts could be plausibly traced back to the L1, a small number of Japanese texts written on the same topic were also collected.

A short questionnaire was designed in Japanese on Google Forms (see Appendix F) and, with the help of two colleagues, distributed by email to 27 Japanese undergraduate students at three different universities in Japan. This means that the Japanese texts collected were written by different students than those who had contributed texts to the ICNALE. By having different writers of a homogenous group (i.e., Japanese university students) complete the same task, it was possible to avoid an order effect (Sasaki & Hirose, 1996).

Students were asked to anonymously complete the questionnaire online, which covered questions about their English language proficiency (i.e., formal test scores, like TOEFL, TOEIC, or IELTS), academic level, field of study, age, sex, year of studying English, and experience abroad (essentially adopting the questions found in the personal info section of the ICNALE questionnaire). Once students provided this information, they were then asked to write a short essay in Japanese based on the part-time job prompt used in ICNALE.

To further improve comparability between the L1 and L2 it was important to avoid possible external influences on the written products. Since a number of studies have shown that different topics may affect both the quality and length of writing (see Carlson, Bridgeman, Camp, & Waanders, 1985; Hamp-Lyons, 1990; Reid, 1990), the same topic (part-time jobs) was chosen as the prompt for the L1 and L2 texts (i.e., the Japanese text and the JEFL texts). Respondents completed the questionnaire and submitted their writing samples on a voluntary basis. In total, nine texts were collected.
After reviewing the nine texts, however, I noticed that three of the texts consisted of three or fewer sentences, which therefore had to be deemed insufficient and dismissed due to their length, so that six texts remained for analysis.

Next, as with the NES texts, these six texts were rated by three native-Japanese speakers on a scale of one to five (one being poor and five being excellent). To serve as a representation of “good” Japanese writing, the text was selected on the basis of a score of 2.3 standard deviations or more above the mean (the mean was 10.5), that is, a 12.8/15 or better. One text received a score of 13 by the raters. This text acted as my point of comparison to the L2 texts, that is, the JEFL texts, and is representative of “good” Japanese writing in this study. The text and the results from the analysis of the text will be presented and discussed in greater detail in Chapter 4.

Analytical Framework
Rhetorical Structure Theory
Before moving on to the methodology, it is first necessary to provide a description of the methodological framework on which this study is conducted. In the next several sections, I will elaborate on the theory behind Rhetorical Structure Theory (RST), its concept, and approach.

RST: An overview
RST, developed by Mann and Thompson (1988), is a descriptive linguistic approach that analyzes the organization of discourse. Unlike other theories of text structure (e.g., Grosz & Sidner, 1986), Mann and Thompson eliminate the necessity of linguistic devices as indicators of relations, and, alternatively, offer a systematic way for texts to be annotated by modeling the rhetorical structure of a text into a hierarchical discourse tree, or RST tree, in which relations between spans of texts are identified. These spans of texts are usually made up of smaller units called the minimal building blocks of discourse or Elementary Discourse Units (EDUs).

The coherence structure of the texts is described in terms of RST (Mann & Thompson 1988; http://www.sfu.ca/rst). The smallest units (elementary discourse units or EDUs) in this analysis are clauses, except for intraclausal constituents, restrictive relative clauses, or independent fragments, that function as complete utterances. The functional relations between the propositions in a text are defined in terms of semantic constraints on the constituent units and the analyst’s plausibility judgments about the writer’s purpose in producing those units. The unit in a relation that is most central to the writer’s purposes is called the nucleus; a less central supporting or expanding unit is called a satellite. The relations apply
recursively to yield a hierarchical structure (so there are nuclei and satellites at all levels, embedded in each other). RST trees combine subject matter relations (relating states of affairs) and presentational relations (relating illocutions or text parts) in a single representation. As Berzlánovich, Egg, and Redeker (2012) explain, “This conflation of content structure and intentional structure allows the analyst to choose the contextually most salient relation, that is, the one that maximizes the relevance of a unit to the local or global discourse purpose at hand” (p. 144).

**Nucleus-Satellite**

Perhaps the most noteworthy feature of RST is the concept of nucleus and satellite, which make up the spans within the text. RST relations are defined in terms of the nucleus-satellite notion, and texts are structured through the rhetorical relations between these two components. Text coherence is attributed to the presence of those rhetorical relations, which are defined in terms of the writer’s intention, that is, the desired effect the writer intended by presenting two spans together. Though the concept of nuclearity is rather novel, the classification in which rhetorical relations occur appears to be comparable to that of other discourse structure theories.

**Asymmetric-Symmetric relations**

According to Mann and Thompson (1988) rhetorical relations can either be asymmetric (nucleus-satellite) or symmetric (multi-nuclear). This resembles the hypotactic-paratactic distinction made by Grimes (1975). However, as has been pointed out, while these classifications were expressed in semantic and syntactic terms, RST’s nucleus-satellite notion is concerned with the (interpreted) purposes of the writer and the effect he/she presumably intends to bring about, which is more in line with the function of rhetoric (Mann & Thompson, 1987; Matthiessen & Thompson, 1988). Therefore, RST provides a way by which to describe the coherence of a text beyond linguistically observable phenomena and as deliberate choices made by a writer seeking to bring about a specific reaction in his/her reader. In fact, one of the four constraints of RST in which rhetorical relations may be defined revolves around this effect (see the list of constraints below as identified by Taboada & Mann, 2006), which is achieved through “a mixture of propositional and intentional language” (Knott, 1996, p. 39).
RST Constraints

1. Constraints on the nucleus;
2. Constraints on the satellite;
3. Constraints on the combination of the nucleus and satellite;
4. Effect (achieved on the reader)

An asymmetric relation is made up of two spans: the N and the S. The nucleus acts as the more important span as it is essential to the writer’s goal and purpose at that point in the text and is independent from the other span, whereas the satellite is less important and serves to support the nucleus; it cannot act independently (see Figure 3.4). Asymmetric relations can be further subcategorized into two categories: Presentational (their effect is to increase some inclination in the reader) and Subject Matter (their effect is that the reader recognize the relation between two spans of texts).

Unlike asymmetric relations, symmetric relations can be made up of two or more spans, each of which is of equal importance to the writer’s intention. In other words, each node in a symmetric relation is a nucleus. An example of a symmetric relation from the RST website (http://www.sfu.ca/rst/) is given below:

Texts are structured through the relationships between these two components, i.e., the nucleus and satellite. EDUs serve as either a nucleus or a satellite and act as spans within the RST tree. A span may be made up of a single EDU or more. The way by which one EDU, or span, is connected to another is by addition of either one of the 25 asymmetric relations or one of the seven symmetric relations (for the full list of relations and their categories used in this study see Table 3.1 in Appendix A; see

Figure 3.4. RST tree representing asymmetric relations.
Table 3.2 in Appendix A for a comprehensive list of the RST relations’ definitions), which can only be assigned by following four constraints: completeness, connectedness, uniqueness, and adjacency.

![RST tree example](image)

**Figure 3.5.** RST tree representing a contrast relation (one of the seven types of symmetric relations).

Completeness refers to the fact that an RST tree must cover the entire text. In other words, annotation is very holistic in nature—it takes the entire text into account and a portion of a text cannot be considered separate from the rest. Connectedness regards how the sub-trees within the tree that represents the entire texts are connected to one another. They must act as a constituent to another until the minimal unit is identified. In a well-structured text, no tree can function outside the hierarchy of the entire text. Uniqueness refers to the concept of how each set of spans must be made up of different segments. For example, in Figure 1, [Span 1-3] is made up of EDU 1, EDU 2, and EDU 3. That span is then broken down into two smaller spans on the following level in the hierarchy of the RST tree: [Span 1] (EDU 1) is the N and [Span 2-3] (EDU 2 and EDU 3) is the S of [Span 1]. It would not be possible, however, for the two separate spans on the same level to be made up of the same segments, e.g., [Span 1] [Span 2-3] and a third span [Span 3], since segment 3 already occupies [Span 2-3] at that level of the tree. This is what uniqueness refers to. And, finally, adjacency denotes the limitation of spans. In RST, only adjacent spans can be connected to one another to form larger spans. So, for example, referring again back to Figure 1, [Span 1] is connected to [Span 3] only by incorporating [Span 2] to create [Span 1-3]. There cannot be a [Span 1&3]—again: in a well-formed text—as one span cannot cross over another without connecting with it according to RST principles. These constraints are employed in this study to produce structural diagrams, i.e., RST trees, of each text in the NES and JEFL corpora respectively.
RST & coherence

RST is designed to represent the discourse structure of a text through coherence relations. When annotating an entire text within the RST framework, the analyst seeks an annotation that includes every part of that text in one connected whole (Taboada & Mann, 2006). The whole text is broken into EDUs and the way by which one unit is connected to another is by addition of one of 32 predefined RST relations, which are defined in terms of four fields: 1) Constraints on the nucleus; 2) Constraints on the satellite; 3) Constraints on the combination of the nucleus and satellite; and 4) Effect (achieved on the text receiver) (Taboada & Mann, 2006). Furthermore, when marking relations the annotator must work within four guidelines: completeness, connectedness, uniqueness, and adjacency. This presupposes that the text is in fact well-structured and every unit is related to one another and/or plays a coherent role in the whole structure of the text. Accordingly, in less coherent texts, these constraints cannot be applied. The question then remains: How does one apply RST to a less well-structured text?

Though this may appear to be an issue in contrastive studies that regularly investigate less well-structured texts (that often being the point of such studies), the fact of the matter is that it is for this very reason that RST is exceedingly suited for contrastive studies. Recent studies have found RST an effective approach for identifying breaks in coherence of the text for the very reason that it is designed to describe the rhetorical relations between segments and spans of texts and thereby the coherence formed within the hierarchy of the text (see Skoufaki, 2009; Yamashita, 2015). Accordingly, should an abnormality occur when annotating a text within the RST framework, the likely cause is that a coherence break has been unearthed within the text. Skoufaki (2009) identified a number of such abnormalities.

Dangling units refer to a unit that is seemingly unrelated to the rest of a text but is linked to it anyway (Skoufaki, 2009). This can occur for a variety of reasons. Often times the content is irrelevant to the main idea of the text. It may also be that the content itself is incomprehensible due to semantic and syntactic errors in the writing. In other cases it may appear as a narrative postscript that distracts or digresses from the topic—similar to didactic remarks as discussed in Chapter 2.

Crossed dependencies occur when a unit intrudes on another span of texts that already hold a rhetorical relation. An example of such an anomaly is given below. Boundaries between segments are marked with the number sign (#). Each segment is subsequently numbered for ease of reference to them in the discussion that follows (see Example 3.1).
Having a part-time job also teaches time management skills and gives a student more room to breathe financially. This kind of responsibility developed over time makes for a happy hiring official. It is important to note however that there are other ways of impressing HR personnel, including internships, leaderships in university of organizations, and volunteer work. However, a part-time job builds both your resume and your income, and should be considered an essential part of a university education.

Example 3.1

In this excerpt from a text from the NES corpus, [Unit 2] offers an explanation, i.e., elaboration relation, for the idea presented in [Unit 1] and together they form [Span 1-2]. [Unit 3] then is connected to this span as an antithesis creating [Span 1-3]. [Unit 4] justifies the statement made in [Unit 5] and together they form [Span 4-5]. [Span 4-5] is essentially a restatement of [Unit 1]. This coherence relation between [Unit 1] and [Span 4-5] is problematic, however, as it appears [Span 4-5] intrudes on the relationship already formed in [Span 1-3]. This kind of intrusion is what Skoufaki (2009) referred to as “crossed dependencies.” In other words, units cross over into other spans and interrupt the coherent relationship formed between the units within that span, which violates the connectedness and adjacency constraints of RST.

Another type of abnormality is the misplacement of coherence relations. In this case, the relations form a coherent whole but their location within the text may be inappropriate. For example, a “background” relation generally occurs more towards the beginning of a text as it provides the background the reader needs to better comprehend the subsequent ideas. Therefore, it appears in preference before the nucleus. If it is placed after the nucleus the relation still holds but the coherence between the two units is awkward. According to Skoufaki (2009), this type of abnormality is indicative of “inductive content order” (p. 193).

It is clear then that RST can be applied to less well-structured texts thereby excavating the coherence breaks occurring within those texts which are making them appear to be illogical, unorganized, and ambiguous. One of the greatest advantages of RST is that, as Taboada and Mann (2006) argue, “ . . . [it] points to a tight relation between relations and coherence in text, thus constituting a way of explaining coherence” (p. 6). In other words, RST offers a way to articulate the unarticulated rules, to speak the unspoken values, and to reveal the hidden conventions with
which learners of English struggle. Consequently, RST is exceedingly suited for answering the question: Why is JEFL writing generally regarded as more illogical, unorganized, and less coherent than that of NES writing?

**Methodology**

The major steps carried out for the annotation of the corpora and comparison of the data were as follows:

1) Elementary Discourse Unit segmentation.
2) Rhetorical Structure Theory annotations.
3) Statistical analysis of the JEFL and NES texts.

**Step 1: Elementary Discourse Unit Segmentation**

Because segmentation is the first step before a RST analysis, quality discourse segments are necessary in order to build reliable RST annotations, just as is the case with any discourse parsing (Hoek, Zufferey, Evers-Vermeul, & Sanders, 2017; Soricut & Marcu, 2003). Therefore, to assure that the segmentation of the texts would be as precise as possible, I chose to implement Tofiloski, Brooke, and Taboada's (2009) automatic syntactic and lexically-based segmenter or SLSeg, which, when run, breaks each text into smaller units of analysis, i.e., EDUs. SLSeg identifies EDUs according to the following definition:

... complements of attributive verbs and cognitive verbs ... are not EDUs.
... adjunct, but not complement clauses are discourse units ... all discourse segments ... contain a verb. Whenever a discourse boundary is inserted, the two newly created segments must each contain a verb ... coordinated clauses (but not coordinated VPs), adjunct clauses with either finite or non-finite verbs, and non-restrictive relative clauses (marked by commas) [are segmented] ... the [segmentation] choice is motivated by whether a discourse relation could hold between the resulting segments. (p. 78)

Because the performance of SLSeg has been shown to have high precision with nearly identical recall (F-score of 0.79) in inserting discourse boundaries (Tofiloski, Brooke & Taboada, 2009), it was not necessary to test for inter-rater reliability.

Automatic segmentation, however, is not without its share of weaknesses. As Tofiloski, Brooke, and Taboada explain, the SLSeg adheres “to the original RST proposals” of Mann and Thompson (1988) that “defined as ‘spans’ adjunct clauses, rather than complement (subject and object) clauses” (p. 78). However, there are
complications that may arise during segmentation based on this algorithm and, in turn, affect the overall analysis of the text. In particular, syntactically embedded relations, such as complement or (non)restrictive clauses may interfere with established segmentation principles (Hoek et al., 2017; Schilperoord & Verhagen, 1998; Verhagen, 2001).

Verhagen (2001) explains that while the application of “Mann & Thompson’s rule” to less complex cases of embedding may bring about the “desired segmentation,” the same cannot be claimed when applied to “more complicated cases” (p. 341). Accordingly, Schilperoord and Verhagen (1998, p. 150) (also see Verhagen, 2001) have proposed a condition on discourse segmentation summed up as follows: “If a constituent of a matrix-clause A is conceptually dependent on the contents of a subordinate clause B, then B is not a separate discourse segment.”

Based on this position of discourse segmentation, it is clear that SLSeg may indeed encounter problems when applied to intricate texts made up of varying clauses, as its algorithm may not always be able to accurately produce segments due to the implicitness of their dependent relationships. The texts used in this study, however, are all relatively short in length, consist mostly of simplex clauses, and contain virtually no complex or non-restrictive relative clauses. As a consequence, adjusting SLSeg would not have an impact on the segmentation in my corpora. This claim is substantiated by the fact that a manual check for the occurrence of multiple complements found zero cases in either corpus.

On a different topic, Hinkel (2005) has pointed out that tagging of syntactic and lexical features of L2 writers’ texts by means of automatic tagging can pose problems due to errors in the language of the texts. Such errors can make it difficult for a program to accurately tag a text. In the case of my corpora, I did indeed observe a few minor errors in segmentation, particularly in the JEFL texts where the JEFL writers had made mistakes in their English. For example, in one text the writer wrote:

# [But then some people may say that college students should not spend their time] # [to do part-time job.]

13 Recent research extends these problems to dependency on propositional content in text fragments and has proposed that segmentation be permitted within certain embedded clauses (see Hoek et al., 2017).
SLSeg marked a boundary (#) between time and to do part-time job. It appears in this instance the program identified to do part-time job as an infinitive clause, which would according to SLSeg principles rightly be marked as an EDU; however, this clearly cannot function as an infinitive clause but should in fact be a gerund clause: doing part-time job. With this grammatical correction, this sentence could not be broken into two separate EDUs. Grammatical errors such as this that affected the outcome of the segmentation were manually revised. Other errors, such as omitting particles (doing [a] part-time job), were not corrected, as these types of minor errors played no significant role in segmentation results.

In another text, the JEFL writer omitted a verb that was needed to make the sentence grammatically accurate:

# [However some people said that students have to work hard and late, and the environment in their workplace sometimes bad.]

SLSeg marked this sentence with a single boundary at the beginning, which would be in accordance to the system's programmed principles; however, without the finite verb is, this sentence is ungrammatical. Since correcting this text by inserting the needed verb is between workplace and sometimes affected the segmentation of this text, this grammatical error was corrected, which resulted in a second boundary:

# [However some people said that students have to work hard and late,][and the environment in their workplace is sometimes bad.]

In all similar cases where the grammatical errors affected the segmentation results, these errors were manually adjusted. There were also cases where segmentations did not appear to be consistent:

# [Also, by having a part-time job, college students can explore their interests and work on things that interest them.]

In this very same text, however, just one sentence before, the SLSeg identified by having a part-time job as a separate segment:

# [By having a part-time job,][college students can learn to be more responsible.]
A possible cause for this incorrectly marked boundary could be that the clause by having a part-time job is embedded in Also, college students can explore their interests and work on things that interest them. Since, however, the program had marked a preposition followed by a gerund clause as an EDU previously in this same text as well as in several others, e.g., without/IN working/VBG, before/IN going/VBG, I manually marked distinct EDUs in all instances where prepositions followed by gerund clauses were embedded in a segment, including the above example.

Once these above corrections were made, segmentation was complete. However, because SLSeg only works with the English language, the Japanese text, which acts as a point of reference, had to be manually segmented by two annotators, the author and a native-Japanese speaker. The second annotator was given Tofiloski, Brooke, and Taboada's (2009) definition of EDUs and asked to segment the Japanese text according to that definition. To test the reliability of the tagging of the EDUs, I used the kappa coefficient $\kappa$, commonly used in discourse analysis. The kappa coefficient is more robust than percentage agreement as it takes into account chance agreement: $\kappa = \frac{p_o - p_e}{1 - p_e}$ (where $p_o$ is the observed proportionate agreement and $p_e$ is the probability of chance agreement).

Marcu, Amorrottu, and Romera (1999) found that when inter-annotator agreement assumes that EDU boundaries can be inserted between each word the chance agreement is very high since many words are within units and not at boundaries. To avoid this pitfall, I assumed that EDU boundaries could only occur between clauses rather than between words. This resulted in a more conservative measure. Once the total number of EDU boundaries was calculated, $\kappa$ was computed with SPSS 23 to evaluate the agreement between the two annotators, the results of which were $\kappa = 0.709$, suggesting a satisfactory agreement rate (Arstein & Poesio, 2008; Cohen, 1988).

Step 2: RST Annotations
Following segmentation, the texts were analyzed according to the RST guidelines and the definitions of each of the 28 relations as originally identified by Mann and Thompson (1988) and augmented with four more recently defined relations (Means, Preparation, Unconditional, and Unless) made available by Taboada and Mann (2006). These relations are classified into two main types: nucleus-satellite

---

14 The author practiced applying RST by annotating previously analyzed texts, comparing the results to the published RST analyses (http://www.sfu.ca/rst/), and adapting the annotation practice accordingly. This was followed by a more formal training with an RST expert, before the RST annotations of the texts in the present corpus were produced.
and multinuclear. Relations between segments that did not fit into one of the 32 relations as identified by RST were tagged as unknown and marked “[???]”. This tag occurred only with dangling units that were incomprehensible due to the writer’s lack of language proficiency and therefore only appears in the JEFL corpus.

The aforementioned texts, i.e., 22 JEFL texts, 22 NES texts, and the Japanese text, were analyzed to create a structural analyses of each text using the UAM Corpus Tool (O’Donnell, 2009), a highly versatile annotating software that allows the user to tag data according to a specified annotation scheme, in the case of this study, the RST scheme, and create RST tree structures (see Figure 3.6).

To ensure the reliability of the RST analyses of the corpora, a second rater, who has a background in applied linguistics, was trained according to the same RST guidelines and principles. Training consisted of reading through the 32 RST relations, discussing the definitions, and practicing annotation of a text together. Following the practice text, all of the texts in the JEFL and NES corpora were shuffled together and 10 segmented texts were randomly selected, which resulted in three JEFL texts and seven NES texts that together totaled 179 segments (approximately 10% of the total number of segments of the JEFL and NES corpora). The second rater and the author then worked together and negotiated the spans and the nuclei in each of those spans until a consensus was reached on all these elements in each of the ten texts. This was done because macrostructure (i.e., spans) influences agreement when establishing rhetorical relations (Iruskieta, Ilarraza, & Lersundi, 2014). In other words, identifying the main idea of a text lends to identification of the structure of the text at the micro-level, which results in improved agreement between annotators. Though it may be argued that this would artificially raise the level of agreement between annotators, the fact that there is a correlation between these two kinds of agreement suggests that establishing agreement at the macro-level would ensure a more reliable annotation at the micro-level. Since reliability of annotations is one of the biggest challenges and criticisms of annotating the rhetorical structure of discourse, it is clearly best to take any and all steps that may lead to a more dependable annotation.

Once the second annotator and the author agreed upon the spans and nuclei, the second annotator was asked to tag the relations within each text in accordance to the RST principles on his own. Reliability of the tagged relations between the two annotators was tested using Cohen’s (1988) $\kappa$ and computed on SPSS 23, the results of which were $\kappa = 0.806$. 
Figure 3.6. An example of the RST tree diagram created using the UAM Corpus Tool.

1) Having a part-time job is one of the best things college students do to help prepare themselves for their futures.

2) It is well known that students to graduate from college with no work experience have an extremely tough time finding employment.

3) and along with this, the employment which they can find is not as lucrative or as engaging as a work that can be found.

4) if they only had some work experience.

5) Almost everyone in the world will have to work at some point in their lives.

6) so why not get started as early as possible.

7) I do think that college students should focus most of their time on learning.

8) but it is very easy to find a place that will allow you to work with a flexible schedule.

9) so that you can stay focused on your studies.

10) Furthermore, if you’re like me and have been going to school for a long time or are pursuing an advanced degree,

11) you will inevitably have to work.

12) in order to make enough money to live and eat.

13) Except for a few rare cases of kids that have extremely wealthy parents,

14) most students will need to work.

15) in order to support either their living situation or their entertainment.
In order to further demonstrate the reliability of the RST analyses and show that the analyses of the NES texts were not designed to fit the raters’ initial ratings, these texts (i.e., the 22 NES texts) were classified into well-written and less well-written texts (as these texts had been classified as “good” writing, it is useful to change the terminology at this point in order to avoid confusion). Classification was based on the results of a second evaluation by native-English speakers using the same one-to-five scale originally used to select the 25 texts for the NES corpus. Well-written texts were identified on the basis of scores of two or more standard deviations above the mean (12 or better), whereas, less well-written texts had two or more standard deviations below the mean (eight or below). Among the 22 texts in the NES corpus, six texts were classified as well written and five texts were classified as less well written. These results were then qualitatively compared to the RST analyses to test whether or not there was a correlation with native-English speakers’ intuitive evaluations of the texts and their rhetorical structure. Specifically, the order of nuclei and satellites were carefully investigated, as well as the actual content of the texts to confirm whether the content held for the intended function, e.g., did an Evidence satellite in fact provide evidence for the nucleus, or did the content of the nucleus actually lend itself to the central role of a span? This will be elaborated on further in the following chapter.

Following the inter-rater reliability test and classification of NES texts, the features of the texts in each corpus for each rhetorical relation were tabulated. The frequency rate of relations in the JEFL corpus was compared to that of the NES corpus (Step 1) and statistical significance was tested between the two corpora’s relation frequencies (Step 2) at the per word level, the per EDU level, and the per text level. A qualitative comparison of the rhetorical features identified in each corpus was then conducted (the results of which will be discussed in the following chapter). This was followed by qualitatively comparing a text representative of “good” Japanese writing to the JEFL texts (Step 3) in an effort to determine any plausible influence and preferential conventions in the Japanese writing that seemed to be utilized in their English writing as well and that may account for differences in the features that could not be identified through the quantitative results (see Figure 3.7); the results of which will also be discussed in the next chapter (the Japanese text will be introduced in Chapter 4).
Step 3: Statistical Analysis of JEFL & NES Texts
The purpose of the statistical analysis is to compare the frequency rate of the 32 RST relations (see Table 3.1) in the JEFL corpus and the NES corpus and determine if there is a statistical significance between the two corpora’s relation frequencies at the per word level, the per EDU level, and the per text level. As mentioned above, the total number of words was 5763 and the total number of segments was 1173 in the JEFL corpus. The total number of words was 4544 and the total number of segments was 673 in the NES corpus.
Normalization of data
As with all corpus linguistic studies, this study attempts to present a quantifiable piece of information regarding the structure of the JEFL and NES texts, which necessitates a solid statistical basis on which significant variables can be identified. And, since the data in each of the corpora differ in size, a direct comparison between the two corpora certainly would not have produced dependable information on the corpora. Therefore, as suggested by numerous scholars (see Biber, Conrad, & Reppen, 1998; Kilgarriff, 2012; Roland, Dick, & Elman, 2007) the raw frequency counts of relations were normalized. Normalization refers to the process of establishing comparability among observations and applies a simple formula to achieve this: the frequency of the observation is divided by the word count and multiplied by the unit in which the feature is analyzed.

Though it is usual to provide measure of word count in corpus studies, word count really has little importance for this study since my focus is at the rhetorical level and the relations of segments to one another within the hierarchy of the text occur at that level rather than at the lexical one. Accordingly, individual words have little effect on how relations are marked since linguistic devices do not function as indicators of relations within RST, as was discussed previously. For this reason, normalization occurred at two levels: 1) by number of words (PNW) and 2) by number of segments, i.e., EDUs (PNS). The normalization was done by dividing each frequency by the number of words (or EDUs) in that corpus, and multiplying by 1000 to give frequencies per thousand.

Statistical tests
Line charts were used to demonstrate and offer a graphical representation of the two corpora’s relation frequencies at the word level, the EDU level, and the text level. Descriptive statistics (mean, standard deviation, and median) were used to summarize the frequencies. Though a chi-square test is commonly used in corpus studies such as this one, due to the low frequencies of many of the categories, which is a well-known disadvantage that generally prevents the use of chi-square (McEnery & Wilson, 2001), chi-square was not adopted for the purposes of this study. A Shapiro-Wilk normality test to determine if the data (raw and normalized frequencies for NES and JEFL) were normally distributed was conducted and revealed that the data were not normally distributed (p = 0.000 for Shapiro-Wilk normality test). Since normality of data distribution could not be assumed, a t-test, which is a parametric test, could not be legitimately applied. When this occurs, an appropriate non-parametric alternative is the Wilcoxon signed-rank test (Loewen & Plonsky, 2015).
Wilcoxon signed-rank test

Wilcoxon signed-rank tests were applied to determine if there was a statistical significance between the two corpora’s relation frequencies at the per word level, the per EDU level, and the per text level. A p-value less than 0.05 indicated significance. The effect size of the Wilcoxon signed-rank tests was computed as

\[ r = \frac{|z|}{\sqrt{n_{\text{NES}} + n_{\text{JEFL}}}} \]

where \( z \) is the standardized test statistic for the Wilcoxon signed rank test, \( n_{\text{NES}} \) is number of observation of relations in NES texts, and \( n_{\text{JEFL}} \) is number of observation of relations in JEFL texts (Field, 2013; Pallant, 2007). According to Cohen (1988), the criteria of the effect size \( r \) is: 0.1 = small effect; 0.3 = medium effect; 0.5 = large effect. The analyses were conducted for 1) all 32 relations ("unknown" was not included), and 2) by category (Presentational, Subject Matter, and Multinuclear). All analyses were conducted using SPSS 23.

Yule’s difference coefficient

In addition to the above tests, Yule’s (1944) difference coefficient was applied to assess the difference in the relative frequency of a relation in the two corpora:

\[ \frac{\text{Freq}_{\text{NES}} - \text{Freq}_{\text{JEFL}}}{\text{Freq}_{\text{NES}} + \text{Freq}_{\text{JEFL}}} \] (Baron, Rayson, & Archer, 2009). It should be noted, however, that the results of this test do not suggest whether the relations are used accurately or not; it simply provides a rate of use in comparison to each corpus. The value of the coefficient varies between +1 and –1. A positive value indicated relative overuse in the NES corpus, and a negative value relative overuse in the JEFL corpus.

Results of Statistical Analyses

Raw Frequency

Table 3.3 in Appendix A presents the raw and normalized RST relation frequencies of NES and JEFL. Figure 3.8 shows the line charts of raw relation frequencies of NES and JEFL. It appears that most relation frequencies were higher for the JEFL corpus than for the NES corpus when looking at the raw frequencies (see Table 3.3 in Appendix A); however, as discussed previously, the different sizes of the data in each corpus prevents a reliable statistical analysis from being conducted. Therefore, it was necessary to normalize the data of each corpus to make it of comparable length so that a solid statistical basis could be established before identifying significant variables.

Normalized Frequency

Table 3.4 in Appendix A shows the frequency rates of relations normalized by per 1000 words and per 1000 segments. Figures 3.9 and 3.10 show the line charts of relation frequencies of NES and JEFL normalized at the per word level and per segment level respectively. When normalized by number of words (PNW), the frequency rates appeared to remain higher among the JEFL texts; however, since
rhetorical relations do not occur at the lexical level but at the EDU level, relation frequencies were also normalized by segments (PNS). This on average raised the frequency rates in the NES corpus and lowered them in the JEFL corpus. For example, “evidence” relation occurred more frequently in the JEFL corpus at both the raw and normalized PNW levels but when normalized PNS, the “evidence” relation occurred more frequently in the NES corpus. Most relation frequencies were higher for the NES corpus than for the JEFL corpus when looking at the frequencies normalized by PNS (see Figure 3.10).

**Difference Coefficient**

Yule’s difference coefficient was used to reveal the diversity of the rhetorical relations used in each corpus (see Table 3.5 in Appendix A). The results show that the frequency rate of the 32 relations was fairly equal between the two corpora. The median value of the difference coefficient was 0.07, suggesting that essentially the 32 relations were represented at nearly an equal rate (with 0 representing an equal rate) between the JEFL corpus and the NES corpus.

**Descriptive Statistics**

Tables 3.6-3.8 in Appendix A show the descriptive statistics (mean, standard deviation, and median) for the raw relation frequencies and normalized relation frequencies (normalized by number of words and by number of segments).

**Results at the text level (raw)**

The analysis results of the Wilcoxon signed-rank tests indicated that at the text level (raw relation frequencies) there was a statistically significant difference in relation frequencies between NES and JEFL (p = 0.000), when comparing all 32 relations (see Table 3.6 in Appendix A). The relation frequency for JEFL (M = 15.81, SD = 20.00) was significantly higher than NES (M = 9.59, SD = 12.17).
Figure 3.8. Line chart of raw relation frequency for NES and JEFL corpora.
Figure 3.9. Line chart of relation frequency (normalized by number of words) for NES and JFL corpora.
Figure 3.10. Line chart of relation frequency (normalized by number of segments) for NES and JFL corpora.
There was a statistically significant difference in relation frequencies between NES and JEFL (p = 0.024) when comparing only the Presentational relations. The relation frequency for JEFL (M = 19.50, SD = 26.13) was significantly higher than NES (M = 11.30, SD = 15.33).

There was a statistically significant difference in relation frequencies between NES and JEFL (p = 0.045) when comparing only the Subject Matter relations. The relation frequency for JEFL (M = 15.13, SD = 19.38) was statistically significantly higher than NES (M = 10.20, SD = 12.23).

There was no statistically significant difference in relation frequencies between NES and JEFL (p = 0.063) when comparing only the Multinuclear relations.

Results at the word level
The analysis results of the Wilcoxon signed-rank tests indicated that at the word level (see Table 3.7 in Appendix A), there was a statistically significant difference in relation frequencies between NES and JEFL (p = 0.019) when comparing all 32 relations. The relation frequency for JEFL (M = 2.74, SD = 3.47) was significantly higher than NES (M = 2.11, SD = 2.68).

There was no statistically significant difference in relation frequencies between NES and JEFL (p = 0.066) when comparing only the Presentational relations.

There was no statistically significant difference in relation frequencies between NES and JEFL (p = 0.495) when comparing only the Subject Matter relations.

There was no statistically significant difference in relation frequencies between NES and JEFL (p = 0.063) when comparing only the Multinuclear relations.

Thus, for all three subcategories (Presentational relations, Subject Matter relations, and Multinuclear relations), the statistical results support the null hypothesis.

Results at the segment level
The analysis results of the Wilcoxon signed-rank tests indicated that at the segment level (relation frequencies normalized by number of EDUs; see Table 3.8 in Appendix A) there was no statistically significant difference in relation frequencies between NES and JEFL (p = 0.100) when comparing all 32 relations.
There was no statistically significant difference in relation frequencies between NES and JEFL (p = 0.374) when comparing only the Presentational relations.

There was a statistically significant difference in relation frequencies between NES and JEFL (p = 0.047) when comparing only the Subject Matter relations. The relation frequency for JEFL (M = 12.90, SD = 16.52) was significantly lower than NES (M = 15.16, SD = 18.17).

There was no statistically significant difference in relation frequencies between NES and JEFL (p = 0.612) when comparing only the Multinuclear relations.

Thus, results of the Presentational relations and Multinuclear relations support the null hypothesis, but the results of the Subject Matter relations reject the null hypothesis.

**Summary of Results**
The statistical analyses showed that most relation frequencies were higher for the JEFL corpus than for the NES corpus at the raw frequency rate. These results changed slightly when normalized by words with more frequency rates more equally distributed between the two corpora; however, normalization per segments caused the frequency rates between the two corpora to swap, with frequencies higher for the NES corpus than for the JEFL corpus. As rhetorical relations occur at this level, it is this frequency rate, i.e., PNS, to which attention should be paid as opposed to the raw and PNW rates. Though the NES corpus did contain a higher rate of relations on average, there were certain relations that occurred more often in the JEFL corpus. The Antithesis relation, for example, appeared over 30% more in the JEFL corpus (N = 5.97) than in the NES corpus (N = 4.46). The Justify relation occurred nearly 80% more frequently in the JEFL corpus (N = 52.86) than in the NES corpus (N = 29.72). And while the Non-volitional Result occurs nearly equally in both corpora (NES = 14.86; JEFL = 14.49), the Non-Volitional Cause occurred almost twice as often in the JEFL corpus (N = 14.49) than in the NES corpus (N = 7.43). Furthermore, the Joint relation, which indicates that two units do not hold any rhetorical relation within the hierarchy of the text, occurred nearly four times more in the JEFL corpus. Yule’s difference coefficient normalized at the PNS level further showed that indeed these relations were overused more by the JEFL writers when compared to their use by the NES writers.
The Wilcoxon signed-rank tests showed that, when comparing the frequency rates of different categories of relations (Presentational, Subject Matter, and Multinuclear) between the JEFL corpus and NES corpora, there was no statistically significant difference at the PNS level for either the Presentational or the Multinuclear relations (though significant differences were found at the raw and PNW levels), i.e., the results support the null hypothesis. However, the relation frequency for Subject Matter relations was significantly lower for JEFL (M = 12.90, SD = 16.52) than for NES (M = 15.16, SD = 18.17), which indicates that the JEFL writers used the 15 different relations from the Subject Matter relations category less frequently than the NES writers did, thus requiring we reject the null hypothesis. Furthermore, the effect size (r = 0.36) also indicates that the difference between the frequency rate of Subject Matter relations in the JEFL and NES corpora is above medium size, suggesting that the difference is meaningful.

Conclusion

Chapter 3 has presented a detailed explanation of the participants, data, and collection method used to gather the data. This was followed by a description of the analytical framework, RST, on which this study is based. The suitability of this particular framework for contrastive purposes in light of the existing literature presented in Chapter 2 was also presented. I then demonstrated how the design of this study addresses weaknesses in research designs of past studies, specifically comparability of data and inter-rater reliability. A description of the statistical tests used to compare the data was also given, followed by the quantitative results of the analyses at the text (raw), word, and segment (EDU) levels. The following chapter discusses the quantitative results and their theoretical and pedagogical implications. Qualitative observations and their significance will also be explored.