COHERENCE IN THAI STUDENTS' ESSAYS: AN ANALYSIS USING CENTERING THEORY[1]

Saranya Pathanasin[2]

บทคัดย่อ

ทฤษฎีเซ็นเทอริ่งถูกนำมาใช้อย่างกว้างขวางใน
งานวิจัยทางภาษาศาสตร์ แต่ยังไม่มีการนำมาใช้กับ
การศึกษาที่เกี่ยวกับทฤษฎีอิทธิพลของภาษาแม่
งานวิจัยนี้นำทฤษฎีเซ็นเทอริ่งมาใช้ในการศึกษา
อิทธิพลของภาษาแม่ (ไทย) ที่เกี่ยวกับการเกาะเกี่ยว
ความในการเขียนเรียงความภาษาอังกฤษของผู้เรียน
ชาวไทยในระดับมหาวิทยาลัย เรียงความจำนวน 50
ชิ้นของนักศึกษากลุ่มตัวอย่างถูกนำมาสร้างเป็น
คลังข้อมูล แล้ววิเคราะห์ด้วยทฤษฎีเซ็นเทอริ่ง ทั้งเชิง
ปริมาณและคุณภาพ ผลวิจัยทางสถิติชี้ว่า ทฤษฎีเซ็น
เทอริ่งมีประสิทธิภาพในการวัดระดับของการเกาะ
เกี่ยวความมากกว่าการวัดด้วยผู้อ่านสี่คน เนื่องจาก
ทฤษฎีเซ็นเทอริ่งปราศจากอคติในการวัดการเกาะ
เกี่ยวความ ผลการวิเคราะห์พบว่ารูปแทนที่กลุ่ม
ตัวอย่างใช้มากที่สุดคือคำซ้ำ (56%) และอันดับที่สอง
คือสรรพนาม (39.6%) ข้อค้นพบนี้ขัดแย้งกับทฤษฎี

เซ็นเทอริ่งที่ชี้ว่าสรรพนามเป็นรูปอ้างที่ใช้มากใน
ภาษาอังกฤษ ผลการวิเคราะห์เชิงคุณภาพชี้ว่า
ปรากฏการณ์นี้เป็นหลักฐานของอิทธิพลของภาษาแม่
ที่ปรากฏในการเขียนเรียงความภาษาอังกฤษของ
นักศึกษาชาวไทย เนื่องจากการใช้คำซ้ำเป็นลักษณะ
ของการเกาะเกี่ยวความในภาษาไทย อย่างไรก็ตาม
อิทธิพลของภาษาแม่นี้ ช่วยให้ผู้อ่านสามารถติดตาม
ความคิดในเรียงความได้ง่าย ผลการวิจัยยังชี้ว่า
นักศึกษาไม่ตระหนักถึงความแตกต่างในการใช้รูป
แทนต่างๆ อีกทั้งไม่มีความมั่นใจในการใช้สรรพนาม
ในการเขียนเรียงความภาษาอังกฤษ ผู้อ่านชาวต่างชาติ
ที่ใช้ภาษาอังกฤษเป็นภาษาแม่ไม่ทราบถึงอิทธิพลของ
ภาษาแม่นี้ในข้อมูล แต่ผู้อ่านชาวไทยสามารถบอกได้
ว่าลักษณะเช่นนี้คืออิทธิพลของภาษาไทยในการเกาะ
เกี่ยวความภาษาอังกฤษ

Abstract

The contributions of centering theory (CT) have been widely accepted in linguistics, but there has not yet been much published research applying the theory to first language (L1) interference. This study applies CT to investigate L1 interference in discourse coherence in essays written in English by Thai university students. A corpus compiled from 50 essays is analyzed quantitatively and qualitatively. Statistical results suggest that CT is more reliable than human raters in measuring coherence, since the data shows no influence of subjective measurement. Further analysis by CT shows that word repetition is the most common form of Cb (focus of attention) in the continuation states (56.6%), and

pronouns are the second most frequent form (39.6%). These findings conflict with the predictions of CT, which assert that pronouns are usually preferred over noun phrases. Results of qualitative analysis pointed out that this discrepancy is caused by L1 interference in writing, since word repetition is commonly used in Thai discourse. The Thai students' English-language writing abilities were at pre-intermediate levels of competence. However, this aspect is not considered negative L1 interference since the use of repetition in the students' compositions helped the raters follow the ideas in the writing. The results also show that the students were not aware of issues of cohesion and were not confident in using pronouns. The L1 interference in discourse coherence, specifically word repetition, was recognized by Thai raters but not by native English-speaking (NES) raters.

Introduction

Coherence cannot be overlooked in the study of discourse analysis because it is an important quality in any text. This project aims to find proper ways to guide students to improve their writing ability by focusing on discourse coherence analysis.

Since coherence involves the thinking process, measuring coherence can be problematic and is subjective. Levels of coherence are not objectively measurable on the surface of a text, unlike other errors such as grammatical errors, spelling errors, and errors in word choice. Therefore, measuring coherence in writing is always challenging, and particularly so in the EFL (English as a foreign language) field. Several methods

have been proposed for measuring text coherence. For example, Todd et al. (2004) proposed topic-based analysis, whereas Lee (2004) applied tree diagrams in measuring coherence in students' essays. Some electronic tools have also been designed specifically to measure coherence in a text, as described, for example, by Higgins et al. (2004), Miltsakaki and Kukich (2000), and Hasler (2008).

Higgins et al. (2004) developed a system called Criterion, which is an online essay evaluation service, to measure coherence based on discourse elements. In the work of Miltsakaki et al., centering theory (CT) was applied in an e-rater tool by focusing on rough-shift transition in the centering model. Such machines are, however, not available in regular classrooms. In addition, the methods proposed for the electronic tools, and in previous studies (i.e. Milsakaki et al. 2000, Higgins et al. 2004, Hasler 2008, Hu et al. 2001) were limited only to the measurement of specific types of coherence. In order to extend CT to the field of EFL, this study adopts CT not only to measure coherence but also to analyze the techniques which Thai students use in writing their essays coherently. The results of the CT analysis reveal particular aspects in students' writing which can pinpoint L1 interference with regard to discourse coherence.

Due to the fact that the organization of discourse is different from language to language, it is possible that EFL learners may construct discourse in their first language (L1) pattern. Like many other EFL learners around the world, Thai students have been documented as demonstrating mother tongue (L1) interference (Pongpairoj

2002, Bennui 2008, Watcharapunyawong and Usaha 2013). Therefore, the researcher hypothesized that the students who form the demographic base in this study tend not to write coherently in a natural English discourse structure, but rather tend to relate ideas and express focus of attention using a discourse structure which more closely matches conventional Thai writing. The present study aims to investigate the techniques that Thai university students are likely to employ in writing coherently, and what the Thai (L1) influence is on coherence in students' writing.

By employing CT in the analysis, the present study aims to reveal two points: firstly, to what extent CT can be applied to analyze coherence in students' essays; and secondly, how L1 interference influences coherence in students' writing. The results are explained from the centering theory perspective.

Review of Relevant Studies

This section attempts to synthesize published work in two main areas, namely centering theory (CT) and first-language interference (L1 interference).

Centering theory (CT) is the main framework of analysis used in the present study. The theory was chosen because it can measure discourse coherence by capturing the focus of attention in a discourse segment. Maintaining and changing the focus from one utterance to another utterance shows how ideas are connected in a discourse segment.

The concept of L1 interference is then presented, and previous research studies in this field reviewed, in order to explore current theoretical understanding in the published literature of L1 interference especially as it applies to Thai students learning English.

Centering Theory

Centering theory (CT) is a linguistic theory which focuses on the center of attention in an utterance and how the *center* or *focus* is maintained or shifted between utterances in a discourse segment. The theory has been used in many discourse analysis works, since it can explain coherence in discourse both in terms of levels of coherence and the use of referring expressions. From the original work of Barbara Grosz in 1997, the theory has been adopted and developed by several researchers in the fields of discourse analysis, computational linguistics, and psycholinguistics, for example Brennan et al. 1987, Gordon et al. 1993, Aroonmanakun 2000, Miltsakaki et al.2000, Hu et al. 2001 among many others.

Many experts in these fields have employed the centering model to explain the different degrees of discourse coherence, for example Gordon et al. (1993) for English, Miltsakaki et al.(2000) for Japanese, Hasler (2008) for computer-aided summarization and Pathanasin et al. (2014) for English and Thai. These works measured coherence by focusing on the use of referring expressions to point to the center of attention. One well-known example by Grosz, Joshi, and Weinstein (1995) is as follows:

Example 1:

a) Susan gave Betsy a pet hamster.
b) She reminded her that such hamsters were quite shy.
c) She asked Betsy whether she liked the gift.

In the above discourse segment, there are two discourse entities realized by *Susan* and *Betsy* which are possibly being referred to by the pronoun 'she' in utterance (b). According to Grosz et al., readers will prefer to interpret "she" to mean *Susan* rather than *Betsy* due to the fact that if *Susan* is the antecedent, the center of attention will be maintained between (a) and (b), resulting in less processing load on readers. By contrast, if 'she' in (b) refers to *Betsy*, the center of attention is changed, which would increase the processing load. According to CT, it would be easier to process if the pronoun refers to *Susan*. In conclusion, *Susan* is preferred over *Betsy* for the coherence of discourse.

CT provides a number of elements to be integrated in order to explain different relationships between utterances in a discourse segment. These elements, namely an utterance's members, constraints, ranking, and rules, are functional in formulating transition states in local discourse. There are four types of transition states. Each type reflects different relationships between utterances. Levels of discourse coherence can be measured by these transition states. Each element in the theory is explained in brief below.

Semantic entities in utterance

A discourse segment consists of a number of utterances, which can each be labeled as, for example $U_1 \ldots U_m$ (utterance 1 to the last utterance). To understand the relationship between such utterances, it is crucial to understand the primary role of each entity in an utterance. According to CT, an utterance consists of a set of semantic entities which are included in a list of forward-looking centers (Cf).

In a Cf set, there is only one entity, namely the backward-looking center (Cb), that links the current utterance (U_i) with its immediately previous utterance (U_{i-1}), and there is one entity called the preferred center (Cp) which is predicted to be the Cb of the next utterance (U_{i+1}). Note that the prediction might not be true, so that the Cb and Cp of a particular utterance might not be the same entity.

According to Brennan, Friedman, and Pollard (1987), the Cf members are ranked according to their grammatical roles which are: Subject > Object(s) > others. This ranking explains that the subject is higher priority than the object(s), which are again higher in priority than others. For example, we can determine the Cf members of utterance (a) in example 1 above, as follows:

Example 2

(U_1) Susan gave Betsy a pet hamster.
Cf: [Susan, Betsy, pet hamster]
Cp:[Susan]

Utterance (U_1) consists of three entities which are *Susan*, *Betsy*, and *pet hamster*. These three are therefore members of the Cf set and are ranked according to their grammatical roles. That is, *Susan* is the subject, so is ranked highest. *Betsy* is the first object, so is ranked second. The pet hamster is ranked third, respectively. All members are put in square brackets ([ ]) for analysis. The first member in this set is the preferred center (Cp) which is predicted to be the backward-looking center (Cb) of the next utterance. According to the theory, the Cb is the most focused entity in an utterance and is realized in the immediately previous utterance.

According to the theory, the highest ranked member in Cf(U_1) realized in (U_2) will be the Cb(U_2). Once again, it is predicted that *Susan* will be Cb(U_2) since *Susan* is the Cp(U_1), as demonstrated in the following.

Example 3

(U_1) Susan gave Betsy a pet hamster.
    Cf: [Susan, Betsy, pet hamster]
    Cp: [**Susan**]

(U_2) She reminded her that such hamsters were quite shy.
    Cf: [Susan (she), Betsy (her),
        pet hamster]
    Cp: [Susan (she)]   **Cb [Susan** (she)**]**

        Grosz, Joshi, and Weinstein (1995)

Following CT, *Susan* is the preferred antecedent of the pronoun 'she' since the entity is the Cp(U_1) and *Susan* has the highest rank by means of the Cf ranking.

The above example shows that the focus of (U_1) is maintained in (U_2).

Centering Transitions

As mentioned above, the focus (Cb) of an utterance may or may not be maintained between utterances. This relationship of change (or non-change) of Cb from one utterance to the next is called a transition state. There are four types of transition state: *Continuation*, *retain*, *smooth shift*, and *rough shift*. Transition states express different relations between the present utterance (U_i) and the previous utterance (U_{i-1}) and they reflect levels of coherence in a discourse segment. Basically, the more Continuations there are, the more coherent the discourse is.

This study will adopt the centering transition states proposed by Brennan, Friedman, and Pollard (1987). The transition states are defined by the relationship between Cb and Cp, as shown in Table 1. The symbol "[?]" means that the current utterance does not have a Cb. Note also that there are cases where transition states cannot be calculated, such as in the first utterance of a discourse.

Table 1 Transition states

	Cb (U_i) = Cb (U_{i-1}) or Cb (U_{i-1}) = [?]	Cb (U_i) ≠ Cb (U_{i-1})
Cb (U_i) = Cp (U_i)	CONTINUATION	SMOOTH SHIFT
Cb (U_i) ≠ Cp (U_i)	RETAIN	ROUGH SHIFT

To demonstrate the computation of transition states, example (3) above is taken once again.

(U₁) Susan gave Betsy a pet hamster.
$\quad$ Cf: [Susan, Betsy, pet hamster]
$\quad$ Cp: [Susan]   Cb [?]
$\quad$ Transition: **None**

(U₂) She reminded her that such hamsters were quite shy.
$\quad$ Cf: [Susan (she), Betsy (her),
$\quad$ pet hamster]
$\quad$ Cp: [Susan (she)]   Cb [Susan (she)]
$\quad$ Transition: **Continuation**

It can be seen in the above that U₁ has no transition, which is expressed as **None**. Then the utterance U₂ is in the **continuation state** according to the transition state formula proposed by Brennan, Friedman, and Pollard (1987) which can be demonstrated as follows:

$$Cb (U_2) = Cp (U_2) \text{ and}$$
$$Cb (U_1) = [?]$$

Rules

The theory has two rules. The extent to which a particular discourse follows, or does not follow, these rules is related to the processing load in understanding that discourse. According to Grosz et al. (1995), CT rules are:

For each Uᵢ in a discourse segment D consisting of utterances U₁, … Uₘ:

1. If any element of Cf(U$_{i-1}$) is realized by a pronoun in Uᵢ, then the Cb (Uᵢ) must be realized by a pronoun also.

2. Sequences of continuation are preferred over sequences of retaining; and

sequences of retaining are preferred over sequences of shifting.

Rule #1 is generally called the 'Pronoun Rule'. It deals with the use of a pronoun that refers to the Cb of an utterance.

Rule #2 deals with the transition states. The preference of the transition state is for coherence of discourse. At this point, the discourse in example (3) will be taken again to demonstrate the importance of CT rule #1.

(U₁) Susan gave Betsy a pet hamster.

(U₂) **She** reminded her that such hamsters were quite shy.
$\quad$ Cf: [Susan (she), Betsy (her),
$\quad$ pet hamster]
$\quad$ Cp: [Susan (she)]   **Cb [Susan** (she)**]**

U₂ in the above example obeys CT rule #1. Since *Betsy*, which is not the Cb, is referred to by a pronoun 'her', *Susan,* which is the Cb, must be referred to by a pronoun as well. According to the theory, if the Cb were not referred to by the same pronoun as before, but rather spelled out in full, then this would increase the readers' processing load, an eventuality which competent writers/speakers prefer to avoid.

It can be seen at this point that the centering model is an explicit way to measure the level of coherence in a discourse segment. Therefore, the present study employed the theory in its analysis. The present study is not the first attempt to make use of the CT in measuring students' essays. In 2000, Miltsakaki and Kukich exploited the

centering model to improve performance of the E-rater, or electronic essay rater. Their experiment, which focused on CT rule #2, aimed to explain the role of rough-shift transitions in students' essays. They pointed out that much of the incoherence in students' essays was caused by undeveloped topics, and that coherence could therefore be measured by the percentage of rough shift transitions.

In contrast, a study by Gordon, Grosz, and Gilliom (1993) focused on centering theory Rule #1. This experiment examined the role of pronouns in discourse. In their work, the use of pronouns was explained by centering theory Rule #1. The experiment's so-called 'Repeated-Name Penalty (RNP)' revealed that if the Cb was not referred to by a pronoun, it would increase the processing load in reading.

Next, a brief review of the notion of first language interference, together with brief summaries of some previous studies on L1 interference involving Thai students, are presented below by way of background knowledge underlying the present study.

First Language Interference in Thai Students

First language (L1) interference is widely accepted as an important barrier in foreign language learning. Ahmed (2010) pointed out that among many other factors such as teaching method and socio-cultural issues, the difference in rhetorical conventions between English and Arabic was an important challenge for Egyptian students in choosing cohesive devices and developing coherence in English essays. Due to the fact that coherence is language specific, a coherent method might be suitable in one language but not another.

A wide body of research has shown that L1 interference is an important problem for Thai learners. In an attempt to help Thai students overcome this barrier, many studies e.g. Pongpairoj (2002), Bennui (2008), Watcharapunyawong and Usaha (2013) have been conducted in order to identify specific L1 interference in Thai students. Some common errors have been identified, and suggestions made on how to solve these problems.

Pongpairoj (2002) analyzed errors in English paragraphs written by 100 Thai university students. Contrastive analysis was the method employed in her study to explain syntactic errors, morphological errors, and word usage errors caused by first language interference. Statistical results showed that students produced the highest number of errors in word usage. Morphological errors and syntactic errors were respectively the second and the lowest number of errors found in her study. Causes for errors were explained through contrastive analysis. Her paper included some recommendations for the design of better teaching materials.

Another study in the same area was done in 2008 by Pairote Bennui. The researcher combined contrastive analysis, error analysis and inter-language analysis with contrastive rhetoric in analyzing 28 English paragraphs written by Thai students. L1 interference was found on word, syntactic, and discourse levels. It was found that students directly translated Thai words into English words due to their misunderstanding of different

semantic systems between the two languages. The wrong word order was often found, along with errors related to some English grammar points such as tense, articles, etc. The researcher pointed out that Thai students employed Thai writing style especially on the discourse level, with incomplete rhetorical patterns and a high level of redundancy.

Following Bennui (2008), error analysis was employed in the work of Watcharapunyawong and Usaha (2013) which focused on L1 interference in Thai EFL students. In their studies, 40 students were assigned to write paragraphs on different topics reflecting different genres, and errors were found in all genre types in 16 categories i.e. verb tense, word choice, article and pronoun. However, the results pointed out that L1 interference affected different genres differently. For example, article use was the highest type of error caused by L1 interference in descriptive writing while verb tense was the highest L1 interference errors in narrative writing. The researchers suggested that genre should be taken into consideration in English essay writing lessons.

Focusing on producing coherence in Thai discourse, Ratitamkun (2010)'s study showed that Thai speakers at 4 years of age used lexical expressions more frequently than reference expressions (i.e. pronoun) by significant percentages (88% vs 12%). She claimed that children will be capable of using referential forms when they get older. This could suggest that the use of lexical expressions to form coherence is common in Thai discourse.

This field of study has proven the problem of L1 interference in different aspects including coherence. Results from previous studies (i.e. Pongpairoj 2002, Bennui 2008, Watcharapunyawong et al. 2013) directly benefit the area of EFL in Thai contexts. These studies have helped curriculum developers to design better courses by increasing their awareness of these problems. Teachers of English should choose materials and prepare lessons to help students reduce L1 interference. However, previous studies on L1 interference (i.e. Pongpairoj 2002, Bennui 2008, Watcharapunyawong et al. 2013) have been conducted only on the surface level of texts, focusing on issues such as grammatical errors, word choice, and word order. There is an absence of research conducted to analyze specifically the way that L1 interference affects discourse coherence. The present study aims to fill this gap by exploring how L1 interference influences the way that Thai students connect their ideas between sentences in writing.

Methodology

Data

The data analyzed in the present study was randomly selected from students' essays at Prince of Songkla University, Phuket, Thailand. The demographic groups studied comprise first and second year undergraduate students who are studying different majors within the Faculty of International Studies. The corpus was compiled from 50 essays, each of which contains five paragraphs comparing and

contrasting two phenomena, and features a range of informative patterns.

These essays were part of the students' final examination; therefore, students wrote essays without using any books or electronic tools. In this task, students chose to write one essay from three different topics given: 1) The differences between classroom-based learning and online-based learning, 2) Advantages of using technology for communicating, and 3) Things foreigners might not understand about Thai culture.

Human Raters

To measure the level of coherence in students' writing, the essays were marked by two native English teachers (R1 and R2) and then again by two Thai teachers (R3 and R4). R1 and R2 are native speakers of English with 3-5 years of experience teaching English to Thai university students and both hold a Bachelor's degree. Living and working in Thailand, both have some communication skill in Thai. R3 and R4 are Thai university lecturers holding a Ph.D. in relevant fields and have 10-plus years of experience teaching English to Thai students on the university level.

All raters read all the essays and gave points specifically on coherence for each essay. They were asked to give between 1-5 points where 1 referred to the lowest degree of coherence and 5 referred to the highest degree. Each essay received different points and the points were calculated into percentages. For example, an essay that received 3 out of 5 points meant that the essay's level of coherence was 60%.

The human raters understood that coherence analysis was the main aim in the present study; therefore, they overlooked other errors such as spelling, grammar, capitalization, etc. After marking, all raters were interviewed specifically on coherence aspects. They gave opinions on students' techniques in giving coherence to their essays and in the use of referring expressions. The interview questions covered whether they could recognize any first language interference in the essays.

In the meantime, the researcher constructed an informal group interview with students in the form of a classroom discussion. The researcher asked students whether they realized their own choice of cohesive devices and asked them to give reasons for their choices of reference expressions in essay writing.

Compiling Corpus

In this step, a corpus was compiled for CT analysis. All essays were separated on the sentence level. Then, all utterances were input in Excel: one worksheet for one essay. In each worksheet, utterances were arranged in the same column as presented in the following:

	Advantages of using technology for communicating
1.	In the world everything are depend on technology.
2.	Technology is beyond developed and make a lot of good thing in today.
3.	Technology have a lot of advantages such as, communicating, business and internet something like that.
4.	There are 3 main advantages of technology for communicating in daily life.

Centering Analysis

To measure coherence in the data Cf (forward-looking centers), Cb (backward-looking center), and Cp (preferred center) of each utterance were identified and input in the next columns respectively. Then transition states between utterances were calculated and recorded in the last column as below:

	Technology in dairy life	CF	CB	CP	TRANS
1.	In the world everything are depend on technology.	the world, technology	?	the world	None
2.	Technology is beyond developed and make a lot of good thing in today.	technology, good thing	**technology**	Technology	Continuation
3.	Technology has a lot of advantages such as, communicating, business and internet something like that.	technology, advantages, communicating, business, internet	**technology**	Technology	Continuation
4.	There are 3 main advantages of technology for communicating in daily life.	3 main advantages of technology, communicating, daily life	**technology**	3 main advantages of technology	Continuation

Coherence analysis

Next, results of the three assessments - by native English speaking raters, by Thai raters, and by CT analysis - were compared through descriptive statistical methods (Box plot) to describe the distribution of coherent scores by each rater and by CT analysis. This stage of analysis aimed to measure the levels of coherence in discourse and prove the usefulness of CT in measuring discourse coherence.

The last step was a deeper analysis of the CT in order to determine if the use of referring expressions conformed to the proposed rules, particularly rule #1 of the CT, which states that '*If any element of Cf(U_{i-1}) is realized by a pronoun in U_i, then the Cb (U_i) must be realized by a pronoun also.*' The consequences of the conformation were investigated, and the results of this analysis are presented below.

Results and Discussion

The results from the quantitative descriptive analysis show the differences in evaluation between CT analysis and the four raters, as presented in figure 1.
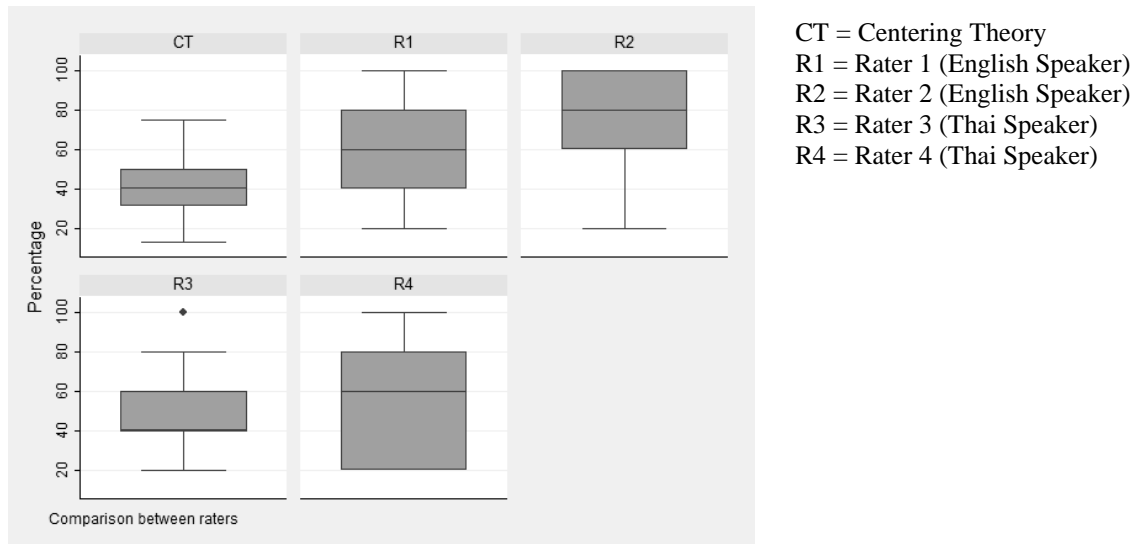
Figure 1 Box Plot of distribution percentages among 50 students evaluated
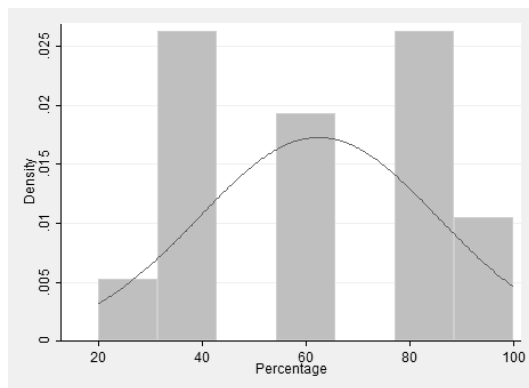by CT and four different raters.

Figure 2: Distribution of percentages among
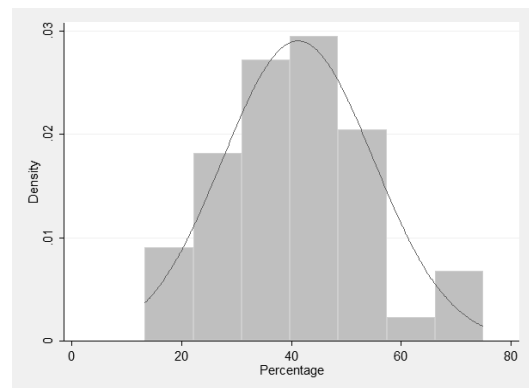50 students rated by R1.

Figure 3: Distribution of percentages among
50 students using CT analysis.

Figure 1 shows that all five coherence evaluations, by CT analysis method, R1, R2, R3, and R4, are significantly different from each other. This can clearly be seen from the data distribution using the box plot visual data display method. The CT analysis method and R1 show symmetric data distribution, i.e., the data is equally divided into quartiles. The data distribution depicted in Figure 1 shows that R2, R3, and R4 did not have the same equal distribution, and so it is not possible to divide the data into four parts (i.e. quartiles) by this method.

Further analysis to differentiate CT analysis and R1 found that R1 was not normally distributed and not continuous, as shown in Figure 2. In contrast, an approximately normal distribution was produced from the results of the CT analysis (Figure 3).

The results could suggest that the coherence analysis is subjective because different raters evaluated the same set of data in different ways. However, the results show that CT analysis is the most reliable measurement because this method was the only method to produce a normal distribution of the data under study. The normal distribution showed no influence of subjectivity in the measurement. This is because CT analysis is an explicit way to analyze coherence in a discourse segment by capturing the focus of attention of an utterance and explaining the relationship between utterances. This confirms the usefulness of the CT model in coherence analysis.

The quantitative results led to a deeper investigation into how students make their writing coherent. The results showed that in all utterances with continuation states, which contribute to maximal coherence, the focus of attention, Cb, is carried on in different forms. The frequency of each type of referring expressions is presented in Table 2.

Table 2 Instances of referring expressions found in utterances with each continuation state

Reference expressions	Instances	%
Word Repetitions	226	56.6
Pronouns	158	39.6
Definite Noun Phrases	9	2.3
Zero Pronouns	6	1.5

As can be seen in Table 2, the types of referring expressions found in utterances with continuation states are word repetitions, pronouns, definite noun phrases, and zero pronouns respectively. Further analysis was done to investigate the reasons why the writers chose each type of reference expression. The results are presented below. Note that the students' spelling and typographical errors not relevant to the area under analysis have been corrected to enhance readability and protect the students' privacy.

Word Repetitions

Results showed that repetition is the most common form of referring expression found in the data of the present study, with a frequency of 56.6%. It was found that students kept focus of attention by using the

same words throughout the discourse segment. The following example is an excerpt from the beginning of an essay entitled 'Things foreigners might not understand about Thai culture'.

Example 4

U1: **Thailand** has a long history and an interesting culture.
<Cb = None, Cp = Thailand>

U2: Long time ago, **Thailand** was a developing country.
<Cb=**Thailand**, Cp=Thailand>

U3: Today, **Thailand** is one of the most popular countries in the world.
<Cb=**Thailand**, Cp=Thailand>

U4: Every year, **Thailand** has many foreigners come to visit.
<Cb=**Thailand**, Cp=Thailand>

U5: Many foreigners come to **Thailand** and want to learn about Thai language.
<Cb=**Thailand**, Cp=foreigners>

U6: Some **foreigners** want to learn and understand about Thai culture.
<Cb= **foreigners**, Cp= foreigners>

U7: Many **foreigners** might not understand Thai culture.
<Cb= **foreigners**, Cp= foreigners>

The above example showed that the student connected her ideas using repetitions, in this case, a proper noun: **Thailand**, and a common noun: **foreigners**. If we look at the

amount or percentage of continuation states in students' essays, this shows that the essays are relatively coherent. However, from the CT perspective, pronouns are preferred over noun phrases (Gordon et al. 1993). The high frequency of word repetition in the data, therefore, showed L1 interference in writing.

For the case of 'Thailand' the researcher would like to elaborate that in the Thai language a country name is neutral and is not referred to by a pronoun 'he' or 'she'. Therefore, Thai students with low English proficiency would not be keen to replace 'Thailand' with a pronoun, whether 'it' or 'she', as 'it' sounds inappropriate for a respected institution while 'she' is feminine although grammatically correct. Nevertheless, 'Thailand' can be referred to by a proper definite noun phrase such as 'the country'. Instead of trying to find a proper referring expression or a synonym, the writer picked the word repetition 'Thailand' to give coherence to her essay.

To support this claim, the researcher would like to refer to Chanawangsa (1986), who found in her research that word repetition is often used in Thai for four purposes: to avoid confusion and make a text easy to comprehend, to reaffirm one's viewpoint, to express interest and to cooperate in conversation. It could be pointed out here that the students employed Thai discourse structure in English writing.

From interviews with students, it was found that some students were not aware that a word should not be repeated many times in writing. They said, for example, "Nobody has ever told me that I cannot do this," and

"It is easier for me to do this because it is like when I think in Thai". The other reason that students gave was "I wanted to keep focusing on the topic that I was writing about."

Unsurprisingly, even though the use of repetition was noticeable by native English-speaking (NES) raters (R1 and R2), they did not recognize it as an example of L1 interference. The NES readers viewed this weakness as poor grammatical ability, or evidence of students' low proficiency level in English writing. One NES rater stated that he would expect lower frequency of word repetition in more advanced Thai students' writing. In addition, he even pointed out that the use of word repetition in the target group positively helped the essays to be more coherent. The use of synonyms, paraphrases or pronouns might even be more confusing and problematic if a wrong pronoun or synonym was used. The other NES rater said that word repetition maintained the central idea in the essay. In sum, the word repetition found in the corpus helped the raters to follow the ideas in the essay.

On the contrary, Thai readers (R3 and R4) recognized repetition in students' essays as showing a similar pattern to typical Thai discourse. They agreed that students were not confident in using pronouns and synonyms in English, probably due to a lack of practical experience in reading and writing complex discourses in English.

Pronouns

The second most frequent cohesive devices found in the data are pronouns. According to Centering Theory, Cb should be referred to (if not *must* be referred to) by a pronoun (Gordon et al. 1993). Interestingly, the percentage of pronouns found in the present study is relatively low (39.6%) when compared to word repetition (56.6%). In addition, analysis shows that the use of pronouns in the data is in an unclear pattern. That is to say, students unpredictably chose between pronouns and word repetition.

Example 5

U_{12}: Although technology for communication has advantages and disadvantages, it can help people.
<Cb=**technology**, Cp=technology>

U_{13}: In my opinion, using **technology** for communication has advantages more than disadvantages.
<Cb=**technology**, Cp=technology>

U_{14}: **It** can help me communicate with family, friends easily.
<Cb=**technology** (it), Cp=technology>

U_{15}: Using **it**, I save time, money, and if I have more free time, I can meet new friends from this technology.
<Cb=**technology** (it), Cp=technology>

U_{16}: For me, **this technology** is very good.
<Cb=**technology**, Cp=technology>

The above example is an excerpt from an essay entitled 'Advantages or disadvantages of using technology for communicating.' It shows the unpredictably mixed non-definite NP (U$_{12}$ and U$_{13}$), pronoun (U$_{14}$ and U$_{15}$) and definite NP (U$_{16}$) in referring to Cb, in this case, **technology**. This confirms the above claim that students were not confident in using pronouns and/or were not aware of the effect of using different cohesive devices in their writing. According to the CT rule proposed by Grosz et al. (1995), which the present study chose to follow, the use of different referring expressions in the data did not violate CT rule #1 regarding the use of pronouns in referring to Cb. However, from Gordon et al. (1993)'s strong view, the result could be considered a violation of the rule which states that *the Cb should be pronominalized;* otherwise it would increase processing load in readers.

In addition, it is found in the data that students use incorrect pronouns, such as using a plural pronoun to refer to a singular noun or vice versa. Note that these pronoun errors did not affect the analysis, as the researcher could identify the antecedents of the pronouns. For example:

Example 6

U$_{15}$: Last week, I went to Phang-nga.
<Cb=None, Cp=I>

U$_{16}$: **Two Spanish** asked me "What is Krab in Sawas-dee-krab?" and a lot of questions from them.
<Cb=I (me), Cp=two Spanish>

U$_{17}$: I know my English is not good but I use my body language and skills to speak with **him**.
<Cb=I, Cp=I>

The above excerpt from an essay titled 'Things foreigners might not understand about Thai culture' showed a pronoun error in U$_{17}$ where the student used the pronoun 'him' instead of 'them'. As mentioned above, the researcher could identify the antecedent of the pronoun. Therefore, the pronoun error did not affect the analysis.

Interestingly, discussion with students revealed that students did not believe they had a problem choosing the right pronoun, even though pronoun errors were noticeable. A deeper analysis showed that the students in the present study were not trained to use pronouns sufficiently in writing. It could be seen clearly in the results that pronouns, which are not commonly used in Thai writing, could be a problem for Thai learners when performing English writing tasks.

Definite Noun Phrases

Definite Noun Phrases are considered a frequently-used cohesive device in English achieved by the selection of vocabulary, called Lexical Cohesion in Halliday and Hasan (1976). However, they are found in only 2.3% of the utterances in this data set. The discourse segment in example 7 was selected from an essay on the topic 'Things foreigners might not understand about Thai culture'.

Example 7

U_{20}: They try to learn about Thai culture, and travel in Thailand.
    <Cb=foreigners (they), Cp=foreigners>

U_{21}: **The foreigners** don't understand Thai culture, but they like to learn Thai culture.
    <Cb=**foreigners**, Cp=foreigners>

From the analysis, the low number of definite NPs can be attributed to L1 interference in students due to the fact that Thai language does not have definite articles (*the*). Therefore, it is rare to see definite NPs used as reference expressions in the data. Thai language often uses word repetition without definiteness, and so it is not surprising that this type of reference expression is the most common reference expression found in the data, as discussed above.

Zero Pronoun

The least common type of referring expression found in the data is zero pronouns at only 1.5%. Example 8 shows an excerpt from an essay on the topic 'Advantages or disadvantages of using technology for communicating'.

Example 8

U_1: I think use technology to contact foreigners is more comfortable and very easy to use.
        <Cb= [?], Cp=technology>

U_2: Ø Can make my life very easy and very comfortable.
 <Cb=**technology** (Ø), Cp=technology (Ø)>

The above example presents the use of the zero pronoun (Ø) when the entity in U_2 is Cb: **technology**. This aspect violates the CT rule because the use of the zero pronoun does not exist in this position in English. However, the zero pronoun in Thai was the focal point in Aroonmanakun (2000) who proposed CT rule #1 for Thai discourse that *If some element of Cf(Ui-1) is realized as a zero pronoun in Ui, then so is Cb(Ui).* Also, Pathanasin and Aroonmanakun (2014) pointed out in their research that the zero pronoun is used significantly often to refer to the Cb entity in Thai discourse.

However, results showed that the zero pronoun, which is the least common cohesive device in the data of the present study with a frequency of only 1.5%, does not occur frequently enough to demonstrate L1 interference in the target group. It could be claimed that the target group generally understand the basic sentence structure S+V+O in English, and did not generally apply subject omission in Thai when writing English, at least in 98.5% of utterances. These results do not suggest a need for extra instruction on this point. The focus should be on choosing appropriate referring expressions such as pronouns or definite NPs in the subject position.

Conclusion

From applying centering theory in analyzing discourse coherence in students' essay writing, two significant points emerge from

the data. Firstly, the statistical results suggest that centering analysis is a more reliable method than four human raters for the reason that the CT analysis shows no influence of subjective measurement as depicted by Box Plot distribution and Distribution of percentages among 50 students by CT analysis in Figure 1 and Figure 3 respectively. Secondly, CT analysis shows that Thai university students employed word repetition as the most common form of referring expression at 56.6%, whereas pronouns are the second most important cohesive devices found, at 39.6%, despite the fact that pronouns are preferred over noun phrases in forming coherent discourse segments according to CT. The other two types of referring expressions found were definite noun phrases and zero pronouns, which were found in small numbers at 2.3% and 1.5% respectively.

A deeper analysis revealed L1 interference in the choice of referring expressions in the target group. That is to say, students employed Thai written discourse structure, in which word repetition is commonly used. In addition, students were not keen on using pronouns, due to their lack of English writing practice, even though they did not have problems identifying suitable pronouns. The interview revealed that students who are at a pre-intermediate level of ability in English were not aware of the effects of different cohesive devices on their writing.

Nevertheless, the use of word repetition over pronoun reference was not always a drawback. NES raters said that word repetition was useful in helping them follow students' ideas in essays. Interestingly, native raters saw students' choices of cohesive devices as evidence of weak grammatical ability in English, whereas Thai raters identified L1 interference in discourse construction.

Implementation

For further studies, it is recommended that investigations should be conducted with Thai students with higher proficiency in English to see the difference in coherency-forming techniques. In terms of curriculum development, the researcher recommends that lessons on the use of cohesive devices should be offered as part of English-language academic writing courses in Thailand.

Acknowledgements

References

Ahmed, Abdel Hamid. 2010. Students' Problems with Cohesion and Coherence in EFL Essay Writing in Egypt: Different Perspectives. *Literacy Information and Computer Education Journal* 1.4: 211-221.

Aroonmanakun, Wirote. 2000. Zero Pronoun Resolution in Thai: A Centering Approach. Paper presented at *the International Conference on Human and*

Machine Processing of Language and Speech, National Electronic and Computer Technology Center, Bangkok.

Brennan, Susan, Marilyn W. Friedman and Carl J. Pollard. 1987. A Centering Approach to Pronouns. Paper presented at *the 25th Annual Meeting of the Association of Computational Linguistics*, 6-9 July, 1987, Stanford.

Bennui, Pairote. 2008. A study of L1 interference in the writing of Thai EFL students. *Malaysian Journal of ELT Research* 4: 72-102.

Chanawangsa, Wipah. 1986. *Cohesion in Thai*. Doctoral dissertation, Georgetown University.

Gordon, P.C., Grosz, B.J., and Gilliom, L.A. 1993. Pronouns, Names, and the Centering of Attention in Discourse. *Cognitive Science* 17: 311-347.

Grosz, Barbara J. 1977. *The representation and use of focus in dialogue understanding (technical report no. 151)*. Menlo Park CA: SRI International.

Grosz, Joshi and Scott Weigstein. 1995. Centering: A Framework for Modelling the Local Coherence of Discourse. *Computational Linguistics* 21: 203-255.

Halliday, Michael and Ruqaiya Hasan. 1976. *Cohesion in English*. New York: Longman Inc.

Hasler, Laura. 2008. Centering Theory for Evaluation of Coherence in Computer-Aided Summaries. Paper presented at *the 6th Language Resources and Evaluation Conference (LREC 2008)*, 28-30 May, Marrakech, Morocco.

Higgins et al. 2004. Evaluating Multiple Aspects of Coherence in Student Essays. Paper presented at *HLT/NAACL, 2004*, Boston, MA.

Hu, Jianhua and Haihua Pan. 2001. Processing Local Coherence of Discourse in Centering Theory. Paper presented at *the 15th Pacific Asia Conference on Language, Information and Computation*, 1-3 Feb, Hong Kong

Lee, C. 2004. Seeing is Understanding: Improving Coherence in Students' Writing. *The Internet TESL Journal*. X.7. July, 2004 <http://iteslj.org/Techniques/Lee-Writing>.

Miltsakaki. Eleni. and Karen Kukich. 2000. Automated evaluation of coherence in student essays. Paper presented at *LREC 2000,* Athens, Greece

Pathanasin, Saranya and Wirote Aroonmanakun 2014. A Centering Theory Analysis of Discrepancies on Subject Zero Anaphor in English to Thai Translation. *MANUSYA, Journal of Humanities* 17.1: 45-63.

Pongpairoj, Nattama. 2002. Thai University
    Undergraduates' Errors in English
    Writing. *Journal of Language and
    Linguistics* 20: 66-98

Ratitamkun, Theeraporn. 2010. Referential
    Choices in Narratives of 4-year-old Thai-
    Speaking Children. *MANUSYA: Journal
    of Humanities* 13.1: 44-59

Todd et.al. 2004. Measuring the Coherence
    of Writing Using Topic-Based analysis.
    Assessing Writing 9.2: 85-104

Watcharapunyawong, Somchai and Siriluck
    Usaha. 2013. Thai EFL Students' Writing
    Errors in Different Text Types: The
    Interference of the First Language.
    English Language Teaching 6.1: 67-78.