LEARNER CORPUS: PRONOUN USE IN WRITTEN
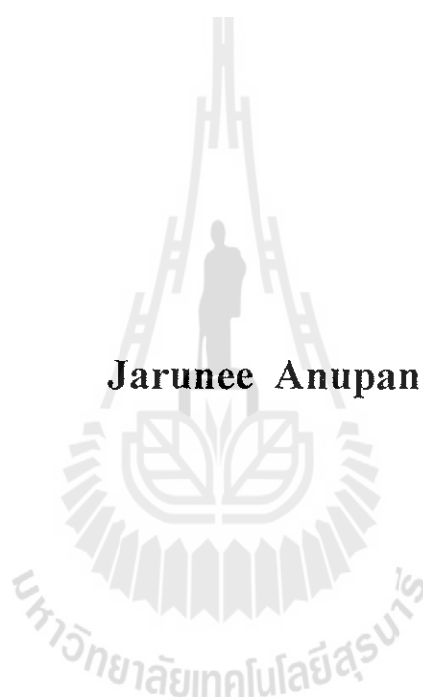
LANGUAGE OF THAI UNIVERSITY

ENGLISH MAJORS

Jarunee Anupan

A Thesis Submitted in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy in English Language Studies

Suranaree University of Technology

Academic Year 2012

ชุดข้อมูลผู้เรียน: การใช้คำสรรพนามในภาษาเขียนของนักศึกษา

วิชาเอกภาษาอังกฤษชาวไทย

นางจารุณี อนุพันธ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศิลปศาสตรดุษฎีบัณฑิต

สาขาวิชาภาษาอังกฤษศึกษา

มหาวิทยาลัยเทคโนโลยีสุรนารี

ปีการศึกษา 2555

LEARNER CORPUS: PRONOUN USE IN WRITTEN LANGUAGE

OF THAI UNIVERSITY ENGLISH MAJORS

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

_____

(Assoc. Prof. Dr. Anchalee Wannaruk)

Chairperson

_____

(Dr. Sanooch Segkhoonthod Na Thalang)

Member (Thesis Advisor)

_____

(Assoc. Prof. Dr. Jeremy Ward)

Member

_____

(Asst. Prof. Dr. Supath Kookiattikoon)

Member

_____

(Dr. Dhirawit Pinyonatthagarn)

Member

_____   _____

(Prof. Dr. Sukit Limpijumnong)   (Dr. Peerasak Siriyothin)

Vice Rector for Academic Affairs   Dean of Institute of Social Technology

and Innovation

จารุณี อนุพันธ์ : ชุดข้อมูลผู้เรียน: การใช้คำสรรพนามในภาษาเขียนของนักศึกษาวิชาเอก ภาษาอังกฤษชาวไทย (LEARNER CORPUS: PRONOUN USE IN WRITTEN LANGUAGE OF THAI UNIVERSITY ENGLISH MAJORS) อาจารย์ที่ปรึกษา : อาจารย์ ดร.สานุช เสกขุนทด ณ ถลาง, 275 หน้า

งานวิทยานิพนธ์ฉบับนี้เป็นรายงานการศึกษาวิจัยคลังคำของผู้เรียนเพื่อศึกษา 1) คำสรรพ นามภาษาอังกฤษ (ในรูปประธาน กรรม และแสดงความเป็นเจ้าของ) ที่นักศึกษาเอกภาษาอังกฤษ ชาวไทยใช้ในงานเขียนบรรยายเรื่อง 2) ความแตกต่างของคำสรรพนามที่นักศึกษาสี่กลุ่มใช้ในงาน เขียนบรรยายเรื่อง และ 3) รูปแบบการพัฒนาในการเรียนรู้ภาษาเรื่องการใช้คำสรรพนามที่พบเมื่อ เวลาผ่านไป งานวิจัยได้ออกแบบชุดแถบป้ายบอกข้อมูล (tagset) ทั้งสำหรับคำสรรพนาม และสรรพ นามที่ผิดโดยอิงจากการใช้คำสรรพนามภาษาอังกฤษในงานเขียน และได้นำวิธีใช้คอมพิวเตอร์ช่วย การวิเคราะห์ภาษาที่ผิด (Computer-aided error analysis approach: CEA) มาใช้ในงานวิจัยนี้ ก่อนที่ จะเก็บข้อมูลงานเขียนของกลุ่มเป้าหมาย โดยข้อมูลที่ได้จะนำมาวิเคราะห์ทั้งในเชิงปริมาณและเชิง คุณภาพ

กลุ่มเป้าหมายคือนักศึกษาเอกภาษาอังกฤษจำนวน 231 คน จากชั้นปีที่ 1 ถึงชั้นปีที่ 4 เขียน บรรยายเรื่องสามเรื่อง ได้จำนวนคำ 228,608 คำ นำมาวิเคราะห์จับหาคำสรรพนามและตรวจสอบ คำผิดโดยใช้ชุดแถบป้ายบอกข้อมูลที่ออกแบบไว้ ผลที่ได้แสดงให้เห็นว่านักศึกษาปีที่ 1 มีค่าเฉลี่ย การใช้คำสรรพนามในงานเขียนบรรยายเรื่องมากที่สุด แต่สรรพนามที่ถูกต้องมีค่าเฉลี่ยน้อยที่สุด (ค่านัยสำคัญทางสถิติที่ $\leq$ 0.05) นอกจากนั้น เมื่อพิจารณาถึงสรรพนามที่ผิดกลุ่มนี้มีค่าเฉลี่ยสูง ที่สุดอย่างมีนัยสำคัญเมื่อเทียบกับกลุ่มอื่นๆ ในการใช้สรรพนามผิดประเภท 01 -05 (omission, overuse, case error, gender error, and number error) และสรรพนามที่ผิดเหล่านี้มีแนวโน้มที่จะ ลดลงเมื่อนักศึกษาเรียนรู้ภาษานานมากขึ้น นักศึกษาปี 4 อาจได้รับอิทธิพลจากภาษาที่ 1 ในขณะที่ นักศึกษาชั้นปี 1 อาจจะได้รับอิทธิพลจากภาษาที่ 1 มากที่สุด อย่างไรก็ดี ในส่วนของค่าเฉลี่ยคำ สรรพนามผิดประเภท 06 (Mis-coreference) ของทั้งสี่กลุ่ม ไม่มีความแตกต่างอย่างมีนัยสำคัญ สรรพ นามผิดประเภท 06 นี้คงที่แม้เวลาจะผ่านไปก็ตาม รูปแบบการพัฒนาของการใช้สรรพนามเมื่อเวลา ผ่านไปนั้นพบว่านักศึกษาในระยะเริ่มต้นนักศึกษามีแนวโน้มที่จะใช้สรรพนามเพื่อแทนที่คำนาม มากที่สุด เมื่อเวลาผ่านไปนักศึกษามีแนวโน้มที่จะแทนที่คำนามด้วยคำแทนที่คำนามอื่นๆ แทนที่จะ ใช้แต่คำสรรพนาม อย่างไรก็พบว่านักศึกษายังใช้สรรพนามผิดในประเภทที่ 06 อาจเพราะมีการ พิจารณาถึงหลักการลำดับการเข้าถึงในการแทนที่คำนาม ไม่มากเท่าที่ควร ดังนั้นการจัดสื่อและ กระบวนการสอนการใช้คำสรรพนามเพื่อระบุการใช้สรรพนามผิดประเภท 06 ในบริบทจึงจำเป็น

ต่อผู้เรียนที่จะทำให้งานเขียนมีความเชื่อมโยงเมื่อสื่อสารด้วยการเขียน ข้อเสนอแนะสำหรับงานวิจัย คือการพัฒนาการออกแบบชุดแถบป้ายข้อมูลการใช้ภาษาในส่วนอื่นๆ ของผู้เรียนภาษาอังกฤษเป็น ภาษาต่างประเทศชาวไทยต่อไป

สาขาวิชาภาษาต่างประเทศ   ลายมือชื่อนักศึกษา_____

ปีการศึกษา 2555   ลายมือชื่ออาจารย์ที่ปรึกษา_____

JARUNEE ANUPAN : LEARNER CORPUS: PRONOUN USE IN WRITTEN LANGUAGE OF THAI UNIVERSITY ENGLISH MAJORS. THESIS ADVISOR : SANOOCH SEGKHOONTHOD NA THALANG, Ph.D., 275 PP.

LEARNER CORPUS/COMPUTER-AIDED ERROR ANALYSIS/ ERROR TAGGING/PRONOUN/INTERLANGUAGE

This thesis is a learner corpus-based study of 1) how pronouns (subject, object, and possessive cases) are produced by Thai university English majors in narrative writings; 2) what the differences of pronoun errors are as they occur in the narrative writing produced by students in four different groups; and 3) what are the developmental patterns of pronoun acquisition that can be seen to occur over time. Pronoun and error tagsets were designed based on Thai EFL learners' writing and employed based on a computer-aided error analysis approach (CEA) before collecting the informants' writings. Both quantitative and qualitative data analysis are employed in the study.

The subjects were 231 university English majors, Year 1 to Year 4. Three story telling tasks were collected for 228,608 tokens and detected for pronouns and errors by using the tagsets. The results revealed Year 1 students produced the highest mean number of pronouns but they produced the lowest mean rates of pronoun accuracy (significant level at $p \leq 0.05$). Additionally, when considered with errors they produced significantly higher means of errors than the others for error types 01 to 05 (omission, overuse, case error, gender error, and number error); these errors tended
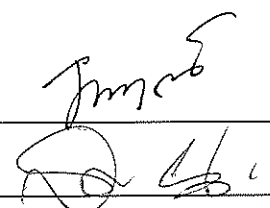
to decrease over time. Year 4 students might have been the least influenced by L1, while Year 1 might have been the most affected by L1. However, no significant differences in the means of error type 06 or mis-coreference across the four groups were found; a trend of the error was likely to be unchanging over time. Development patterns of pronoun use over time found that the students used pronouns most frequently used pronouns to replace a noun phrase at the beginning; as time went by, the students tended to replace the noun phrase with other pronominals. However, with little consideration of referent accessibility hierarchy, the students persisted in error type 06 or mis-coreference. More effective teaching materials or methods for identifying mis-coreference of pronoun in contexts, therefore, are needed so that the students are able to maintain better cohesion when communicating in writing. For future research, the tagsets that were designed for Thai EFL learner language can be extended to larger learner corpora in the future.

School of Foreign Languages   Student's Signature_____

Academic Year 2012   Advisor's Signature_____

ACKNOWLEDGEMENT

Thanks to my friends who have helped me with invaluable advice and discussions: to Jenny for both inspiring theoretical discussions and detailed comments, to Lin and Bass, for their constant willingness to encourage me to pass the difficult time, and to all staff at SUT, Ms. Saruta Jantaros in particular for her kind support and understanding.

My biggest thanks go to my family. My parents, Mr. Sawat & Mrs. Wannee Lagkham, have encouraged me, for as long as I can remember, to go through life asking questions. Although this thesis is a milestone, it would mean nothing were it not against the backdrop of enormous support and love of my parents, my brother Joe, my sister Jang. Finally, I am grateful that my life as a graduate student has been shared with my little family. Thanks to Santi, my husband, for making me realize that life is not always easy and for taking care of my two little birds: Jakkajee and Ja while I was busy with theses both master and doctoral degrees. Thanks for the kids' smiles and hearts cheering me when I face challenging conditions. The thesis is dedicated to my Supermom and kids; with these people, I am who I am today.

Jarunee Anupan

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This chapter includes background of the present study: 1.1) rationale, 1.2) statements of problems, 1.3) purposes, 1.4) research questions and hypotheses, 1.5) significance of the study, 1.6) scope and limitations, 1.7) definitions of key terms & abbreviations, 1.8) overview of the chapters and 1.9) summary of the chapter.

The present study investigates Computer Learner Corpus (CLC), a computerized collection of learner language (Leech, 1998), providing interesting insights into several aspects for people desiring to explore how the target language is learned as well as what has been and has not been acquired by the learners in order to support better learning processes (For overviews of the current state of learner corpus research see Granger (2002 and 2008) and Pravec (2002)). Computer learner corpora are used to test hypotheses in second language acquisition in two main ways. First, learner corpora can be used for the so-called Contrastive Interlanguage Analysis (CIA), i.e. the quantitative comparison of learner language and native language to find patterns of overuse or underuse. For CIA, a corpus does not have to be tagged. Second, learner corpora can also be used for the computer-aided error analysis (CEA), i.e. annotation or tagging of error in particular aspects. Error tagging is the main tool of this study, English pronoun for Thai EFL learners in particular. CIA is an additional approach to be applied to facilitate redesigning of error tagsets and to examine learner language development.

With Computer-aided Error Analysis (CEA) and Contrastive Interlanguage Analysis (CIA) approaches specifically adapted for Thai university English majors, tagsets of pronouns and pronoun errors will be designed for Thai EFL learners in particular according to the pronoun in their mother tongue; pronouns in written mode will be explored regarding their errors, accuracy, and developmental patterns over time.

1.1 Rationale

Although second language acquisition (SLA) studies have long been quite common, more innovative approaches like Computer Learner (CLC) have been conducted more in the present decade to reduce some limitations in previous SLA studies (Granger, 2003, Tono, 2000b). Many problems in written performance of English produced by EFL learners have been raised in many studies of second language acquisition (e.g. Blagoeva, 2002; Dagneaux et al, 1996; Gennari, 2004; and Wendan, 2004) showing EFL learners have problems in acquiring pronouns. In addition, some studies (e.g. Huang, 1994, 2000; Pu, 1995; Takami, 1987; Valin, 1990; and van Hoek, 1995) explored the language produced by Asian EFL learners in written texts and found that English pronouns are challenging for EFL learners since they require not only syntax but also other linguistic elements (pragmatic, discourse, and semantic) to understand and employ English pronouns correctly (Clark and Bangerter, 2004; Levinson, 1995; Unger, 2006; and Walker, 1998). The results conform to what has been found in some studies of Computer Learner Corpora, considered as a better approach fonr storing, generalizing, and applying larger and more particularized data as pedagogical materials (McEnery, Xiao, and Tono, 2006).

From second language acquisition (SLA) perspectives, Computer Learner Corpora (CLC) have been increasingly investigated in order to examine learner language in computerized textual forms (Leech, 1998); the studies have been specified to operate with larger sets of data and have resulted in more reliable conclusions (Granger, 2008) with corpus-based approach employed.

Corpus-based studies undoubtedly provide meaningful information for syllabus and material design; however, many researchers (e.g. Aston, 2000; Flowerdew, 1998; Gilquin, Granger, & Paquot, 2007) found that studies based on the analysis of expert or native corpora are inadequate to inform the design of teaching materials. To improve language teaching materials and later syllabi in reality, the studies need additional information of students' interlanguage. Aston (2000: 10) stated that information provided from corpora analysis of native speaker texts is insufficient to provide information about the practical difficulty and learnability of particular features. Such information might relate to interlanguage development possibly helping to identify features that should be underscored in teaching and to evaluate the learners' limitations. CLC studies have been increasingly conducted around the world throughout the last decade (Granger, 2008; Römer, 2004; Pravec, 2002).

Studies of learner corpora have been developed in many countries for more than twenty years so far (Granger, 2001, 2008; Guo, 2006; and McEnery, et al, 2006); major studies have been concluded in Europe where most of the people use English as a second language (ESL) and Asia where English is, for the most part, a foreign language (Granger, 2002; 2008, and Pravec, 2002). The most well known learner corpora studies have been conducted in ICLE, the International Corpus of Learner

English, (McEnery et al, 2006) primarily with both Contrastive Interlanguage Analysis (CIA) approach to find out how the ESL learners acquire English, and alternatively with Computer-aided Error Analysis (CEA). The latter will be emphasized more for investigating the issues of pronouns in written language in the present study.

As with the findings of SLA research mentioned earlier, that L2 learners have problems with pronouns when writing, findings from some recent studies of learner language in written mode conducted by ICLE project, the International Corpus of Learner English, in the University of Louvain-La-Neuve, Belgium, revealed seven grammatical errors in both intermediate and advanced learners of English. They found that pronoun errors ranked as the third most frequent (intermediate) and second (advanced) out of seven frequent errors produced by the EFL intermediate and advanced subjects (Dagneaux et al, 1998: pp.166). Remarkably, the EFL advanced learners tended to produce more pronoun errors than the intermediate did. However, most of the studies under ICLE project compiled data from subjects, the majority of whom studied English as a second language; the results possibly different from ones conducted in different contexts.

Despite the fact that the learner corpus project has been established in many Asian countries including Thailand, Computer Error Analysis has been specifically explored for particular Asian learners (e.g. Japanese, Chinese, Taiwanese, and Korean), but has not been examined for Thai learners with a particularized design under the CEA approach, based on information of learner corpora around the world (http://www.uclouvain.be/en-cecl-lcWorld.html). The studies conducted in these countries have examined several aspects of EFL learner language: collocation (Chi et

al, 1994; Chen, 1998), connectors (Milton and Tsang, 1993), irregular verbs (Tono and Aoki, 1998), grammatical morphemes (Tono 1998, 2000a), and verb semantics (Oshita, 1997). Thai learner corpora, hence, should be conducted so that we can know more about Thai EFL learner language with the more innovative CEA approach. The present study will use CEA in the exploration of learner language of English pronouns.

In brief, some second language acquisition studies have revealed L2 learners experience some consistent problems when writing, especially pronouns used as pronominal referents in contexts (Blagoeva, 2002; Dagneaux et al, 1998; Huang, 1994, 2000; Paddungwiang, 2009; Pu, 1995); the present study with the CEA approach aims to collect written data produced by Thai learners of English at a university level, to explore how English pronouns and pronoun errors in the written products are alike and to examine the learners' inter-language of English pronouns. The approach will store and analyze the computerized learner performance with a specific error tagset designed for learners whose mother tongue is Thai; real need analysis of Thai EFL learners will be noted as another pioneering step of the research in the innovative area of CLC, since the error tagset will be particularly designed for Thai EFL learners based on the learners' problems of pronouns.

1.2 Statements of Problems

In accordance with results of some studies of pronouns produced by L2 learners in written tasks (Chen, 2002; Huang, 1994, 2000; Pu, 1995; Takami, 1987), some Asian EFL learners experience problems with pronouns. The primary problems with pronouns in context consist of several error types (Ann, 2005): overuse, underuse

or omission, malformation (case, gender, and number errors) and mis-coreference. All error types of pronouns in written language will be determined in the study.

Some pronoun errors have been generally found in contexts produced by the EFL learners (Huang, 1994, 1995; Hyland, 2002; Ivanič & Camps, 2001; and Martinez, 2005). One compelling example of research into pronoun error, presented at the Thailand TESOL 2007 conference (Chimongkol 2007) found that UBU second year English majors show some errors of clause heading including pronoun errors such as omission and overuse. These errors, she concluded, draw from the students' mother language, Thai. For examples:

    (a) *Love you very much. (Subject omission)

    (b) *My uncle he is a teacher. (Pronoun overuse)

(Chimonkol, 2007)

In addition, other errors of malformation (gender, number, and case errors) are considered pronoun problems even though they are not very serious ones in communicative English. However, the errors are still considered ungrammatical, for examples:

    (c) *I like my new friend, Lisa, because *he* is helpful and funny.

        (a gender error)

    (d) *Everyone wants to take a break during New Year, for *they* work

        hard all year long. (a number error)

    (e) *It is harder for him than for *she*. (a case error)

        (adapted from Broukal, 2002: pp.11-22)

Lastly, other pronoun errors that occurred in Hoek (1995) and Valin (1990) are exemplified below:

(f) *He_i went home, and then **Peter**$_i$ took a nap.

(g) *In **John**$_i$'s apartment, he$_i$ holds wild parties.

(h) *Near **John**, he, saw a snake.

The ungrammatical sentences showed in (f) to (h) reveal a problem of pronoun mis-coreference within the sentences, according to the Principle B (Chomsky, 1981) that a pronoun must be free within the minimal category. Mis-coreference errors possibly convey more misunderstanding in both communicative and academic English since readers never have any chance to negotiate the meanings in the written task.

The other examples of pronoun errors are mis-coreference across sentences. The errors may bring higher misunderstanding since they may collapse cohesion between sentences (Huang, 2000). Moreover, in reality negotiation of meaning between readers and a writer for understanding pronoun referents in written contexts seems unfeasible especially if compared to one in spoken situations. Some studies (e.g. Fox, 1993 and Lappin and Leass, 1994) have been conducted to examine pronoun problems between sentences as a case of pronoun resolution of EFL learners (further discussed in chapter 2).

An additional example from an excerpt taken from Hatch's 1992 work contributed as a completion task presented at a Thailand Post TESOL in Ubon Ratchathani University 2006 by Boonmee is shown below. An additional example, drawn from Hatch 1992, was presented by Boonmee (2006). It is a completion task which requires either corresponding pronouns or the names to appropriately refer to the names provided in the parentheses (P = Patricia Hayward, G = Gregory Hayward).

The River of No Return Wilderness is blanketed by several feet of snow for up to eight months a year, and **Patricia and Gregory Hayward** have seen much of it on cross-country skis. Alumni of the University of Idaho (P) (1)*she (#Patricia) has a bachelor's degree in wildlife ecology and (G) (2)*he (#Gregory) has just finished his doctorate. (P&G)(3) They met on an owl study in Rockies and from 1984 to 1988 (P&G)(4) they persevered at the demanding task of tracking these birds. The job, says (G)(5) Gregory, was "far from work. My fondest memories are skiing through miles of unbroken powder knowing that (P)(6) *she (#Patricia) was the only other human for more than 20 miles or listening at night to the soft call of a boreal owl." (G) (7)*He (#Gregory) intends to continue to monitor the boreal owl population and to study old-growth forest. Until recently, (P) (8) Patricia, and avid horsewoman, spent almost all of (P's) (9) her leisure time participating in dressage. But most of (P&G's) (10)*their (#Patricia and Gregory's) time now goes to their son, born in June. Distance running keeps (P) (11)*she (#Patricia) in shape for chasing owls, horses, and little boys. (P&G) (12)*They (Patricia and Gregory) recommend two books for readers who wish to learn about boreals and other owls: Johngard's *North American Owls* and Heimo Mikkola's *Owls of Europe*.

The errors occurring in the passage were examined across sentences for mis-coreference errors. The pronouns must match the number and gender of the noun phrases they stand for and be in a case that matches its function. For example, in item (6) the pronoun '*she*' is not correct since there is no antecedent with the same gender and number in the previous clause or sentence. The correct pronoun use in context will help readers to carry on the linkage between sentences, thus enabling cohesion, the grammatical and lexical links within a text or sentence that holds a text together and gives it meaning. To continue cohesion in their writings, Thai EFL students in the study still selected incorrect or inappropriate pronouns to replace the antecedents from the previous sentences; accordingly, they might produce incorrect or inappropriate pronouns when writing in context. The error of mis-coreference, both within and

across sentences, will be considered in the present study since it has been found to be common among Thai EFL learners.

In studying how Thai EFL learners produce pronouns in written tasks, pronoun accuracy and errors will be detected and analyzed in the present study. After detecting the target features, each pronoun error will be analyzed in terms of overuse, omission, malformation (nominative, accusative, numeral, and genitive cases), and mis-coreference. The analysis of pronoun errors in writing can be accomplished with CEA, one of the Computer Learner Corpora approaches, which is particularly designed for Thai learners as addressed in chapter 3. The examination of written interlanguage for pronouns will contribute to an understanding of pronoun development over time.

1.3 Purposes of the study

The study aims to:

1) explain how pronouns are produced in written mode by Thai university English majors;

2) categorize and compare actual types of pronoun errors occurring in Thai university English majors' narrative writing within and across groups based on years of study; and

3) explore development of pronoun acquisition in the language learning process of Thai university English majors over time.

1.4 Research Questions and Hypotheses

1.4.1 Research questions

1) How are pronouns produced by Thai university English majors in narrative writings?

2) What are the differences of pronoun errors occurring in the narrative writing produced among the students in different four groups?

3) What developmental patterns of pronoun acquisition can be seen to occur over time?

1.4.2 Research Hypothesis

To all research questions, Thai university English majors might differently produce pronouns in narrative writings; pronouns produced by the language learners with more exposure to the target language are assumed better. The pronoun use in written language and errors made by Thai EFL learners is clearly influenced by L1, especially at the beginning. It is likely, therefore, that the learners in Year 1 will produce the highest rate of pronoun errors and the lowest level of pronoun accuracy

According to Corder (1981) suggesting that L2 learners with more exposure to the target language tend to be decreasingly influenced by their first languages, the students in year 4 are possibly the least affected by L1 in accordance with the most years of English exposure; the pronouns produced by this group might be the closest to native speakers'. The others with less English exposure might be influenced by L1 and produce more pronoun errors. However, first language might not be the main culprit influencing the errors; other possible causes of pronoun errors have to be counted.

Other than first language interference, some potential causes of pronoun errors produced by EFL learners mentioned by Corder (1981) are probably over-generalization (learning new language data in learners' minds and then generating flawed rules for their productions based on the evidence) and ineffective teaching materials and methods.

In brief, first language transfer, over-generalization, and ineffective teaching materials & methods might be main sources of pronoun errors. However, some additional causes may include: ignorance of rule restriction; incomplete application of rules; false concept hypothesizing (suggested by Richard 1974); and carelessness and translation (Norris 1983). These will be again discussed in chapter 2 including some discussions on avoidance, when usage structures are simply avoided by L2 students when the differences between L1 and L2 are significant (Schacher, 1974; Schacher and Celce-Murcia, 1977).

1.5 Significance of the Study

Accuracy and error rates of pronouns in writings produced by the subjects (Thai EFL learners) will be explored, categorized, and analyzed with CEA approach to observe development patterns of pronoun usage in narrative writing over time. The errors will be marked as negative evidence, information of ungrammatical form of the target language; while accuracy will be marked as positive evidence, information of grammatical forms of the target language (Marcus, 1993: 53).

Predominantly, the results in terms of learner interlanguage of English pronouns can be informed to Thai EFL teachers so that the teachers will be more aware of the errors and plan to maintain how the learners use English pronouns more

successfully in narrative writing. The pedagogical significance of second language acquisition studies has been underscored since the studies that reported routes and phases that learners go through in acquiring a second language and subsequent languages were first published (Nunan, 1996). In addition, some learner corpora studies (e.g. Aston, 2000; Gulquin, Granger, & Paquot, 2007) indicated that only information of native corpora analysis is not enough for syllabus and material design, but also information of learner corpora analysis is needed since the information can support the identification of features which teaching should emphasize, and to evaluate the learners' difficulty and learnability of specific linguistic features (Aston, 2000).

As there is no standard of error tagset for all second language learners (Díaz-Negrillo and Fernández-Domínguez, 2006), the study might contribute to CEA in the area of computer learner corpora (CLC) through its use of tagsets designed for Thai EFL learners and in employing computer error analysis approach. Each tagset might not be shared by the research in the same area; and when the tagset if for commercial purposes, such as the Louvain error tagging system, it does not come out to be applied as generally as it would be anticipated. With the unique design, both tagsets can be extended to other parts of speech (POS) and errors produced by Thai EFL learners. After POS assembling, consequently, the learner's written language tagged will let the teachers analyze models of typical students' linguistic knowledge, instead of deductive treat to individual student papers from reading, correcting, and responding.

1.6 Scope and Limitations of the Study

The main focus of the present study is the use of English pronouns as produced in written form by Thai EFL learners. For one thing, it is unworkable to concentrate on every part of speech (POS). However, one vital reason for having selected pronouns rather than other parts of speech is that "nouns are more topic-related than other parts of speech" (Leech 2001: 332); a pronoun functions as a replacer of a noun or noun phrase to avoid repetition in contexts. Plus, personal and possessive pronouns must convey the same gender and number to the antecedents they refer to in the previous clause or sentence (Huang, 2000). The scope of the study is, therefore, about pronouns in narrative writing by manual tagging for Thai EFL learners at university level.

Error tagging in the CEA approach will be manually processed because there is no learner corpus automatic tagging software applicable for speakers of Thai according to Prof. Granger S., the project director of ICLE, (personal communication, February 14, 2009). Although learner corpus linguistics has the use of other POS-taggers like Constituent Likelihood Automatic Word-tagging System (CLAWS) developed by the University Center for Computer Corpus Research on Language in Lancaster University, errors are not included for analysis by automatic tagging software, which requires both highly trained labor and costly financial support.

An additional limitation is that subjects are all English majors in Thailand. The study does not include EFL learners from other majors. This limitation may affect the validity of tagsets, which are designed for English majors, not for other majors and other EFL learners in lower or higher levels.

In short, the present study investigates Thai EFL learner usage of pronouns as employed in narrative writing by designing tag sets of English pronouns and pronoun errors for Thai university English majors in particular.

1.7 Key Terms & Abbreviations

1.7.1 Key terms

In this study the terms:

Interlanguage means pronoun production, not comprehension. It involves English pronoun used in written language as narrative writing, a story telling essay, acquired by the subjects; the interlanguage of a pronoun or of pronouns is/are theoretically investigated according to the learners' performance of pronouns in contexts.

Learner corpus means a collection of pronouns in narrative writings performed by Thai university English majors, in a computerized textual form.

Writing refers to narrative writing, picture story telling task (see chapters 2 and 3).

Pronoun means English personal and possessive pronouns. It includes personal pronoun in subject and object cases, and possessive pronouns: first person (I, we, me, us: mine, ours), second person (you, you: yours), and third person (they, he, she, it, them, him, her, it: theirs, his, hers, its); since these pronouns are governed by almost the same rules.

Error tagging represents an approach marking different types of errors with special tags (Granger 2003). The approach is designed for SLA research to compare types of error and error frequency across different groups, in this case Thai EFL

learners. The error tagset is predetermined as a list of error types based on contrastive analysis of English and Thai pronouns (see more information in chapters 2 and 3)

1.7.2 Abbreviations

List of abbreviations commonly used in the study

CA	=	Contrastive Analysis
CEA	=	Computer-aided Error Analysis
CIA	=	Contrastive Interlanguage Analysis
CLC	=	Computer Learner Corpora
EA	=	Error Analysis
EFL	=	English as a Foreign Language
ESL	=	English as a Second Language
IL	=	Interlanguage
L1	=	First Language
L2	=	Second Language
LT	=	Language Transfer
POS	=	Part of Speech
PRO	=	Pronoun
SLA	=	Second Language Acquisition
TL	=	Target Language

1.8 Overview of the Chapters

This thesis includes six main parts: introduction, literature review, research methodology, results, discussions, implications and conclusions. Contents of each chapter are briefly described below.

Chapter 1: *Introduction* provides research rationale, statements of problems, objectives & research questions, and hypothesis. The introduction provides a common overview of the study informing why and how the research will be carried out. In addition, some key terms are defined since the definitions might be somehow different from the same terms used in other studies; the readers can better understand the terms based on the research framework.

Chapter 2: *Literature Review* includes four main parts according to the research theoretical framework: 1) various SLA approaches (contrastive analysis, error analysis, and interlanguage) related to the present study with previous studies including some corpus-based studies, 2) current situations of studies of Computer Learner Corpora (CLC), 3) Computer-aided Error Analysis (CEA) & related studies, and 4) English & Thai pronouns including related studies.

Chapter 3: *Research Methodology* is devoted to the elaboration of research methods and procedures. The chapter starts with a pilot study conducted before adjustment. The research participants' background and group characteristics are defined next. The research design is then discussed in accordance with the research questions. Then, the chapter focuses on how the main research instrument has been designed based on CEA for specified tagsets of pronouns and pronoun errors for Thai EFL learners. With corpus-based processes, annotated key words will be computerized and analyzed by a concordancer, in this case WordSmith Tool version 5.0. After counting the occurrences of pronouns and their errors in the writings using the above-mentioned software tools, statistical and text analysis techniques are presented in accordance with the research questions.

Chapter 4: *Results* are presented after statistically calculating and analyzing both quantitatively and quantitatively according to the research questions. The first part is about general information of how the students of four groups write in terms of tokens, pronouns, pronoun accuracy, and pronoun errors; the section also provides some examples of pronouns in contexts. The second part emphasizes errors produced by the four groups including errors in contexts and how patterns of errors occur among the four groups. Lastly, developmental trends of pronouns over time, as seen in contrasts between the groups of subjects, are reported.

Chapter 5: *Discussions* examine what the results revealed in terms of the theoretical framework reviewed previously. The explanations of how pronouns are produced, the nature of differences of pronoun errors produced by the students in four different groups, and how pronoun use develops over time will be provided in this chapter.

Chapter 6: *Pedagogical Implications and Conclusions* consists of suggestions for pronoun teaching including examples of teaching materials and further studies. Finally, conclusions of the study are stated.

Appendices provide materials separated into two appendices. Appendix A includes information about the English major curriculum at Ubon Ratchathani University. Appendix B contains the materials to be employed in the study, the picture story telling tasks: Little Red Riding Hood, Snow White & the Seven Dwarfs and Cinderella.

1.9 Summary of the Chapter

Computer Learner Corpora studies have become more common among second language acquisition researchers due to their advantages of better storage, generalizing, and application for pedagogical material design. The innovative approach will be employed in the study for examination of learner language, pronouns and pronoun errors in written mode, associated with original SLA approaches (contrastive analysis, error analysis, and interlanguage). The new approach is called 'Computer-aided Error Analysis (CEA)'. As a result, pronoun and pronoun error tagsets are designed based on Thai EFL learners' background. In addition, results from different groups will be compared and analyzed in terms of interlanguage of pronouns in written mode.

As is discussed in detail in the review of literature in the next chapter, Thai computer learner corpus has been conducted from the perspectives of computational linguistics.

CHAPTER 2

LITERATURE REVIEW

This chapter is organized into four main sections: 2.1) Second language acquisition approaches: contrastive analysis, error analysis, and interlanguage; 2.2) Computer Learner Corpora (CLC) and the current situation of learner corpora around the world; 2.3) Computer-aided Error Analysis (CEA) and error annotation; and 2.4) English & Thai pronouns. Also related studies will be provided within the sections.

Research into the area of computer learner corpus has been a very recent branch of learner language (Granger, 1998, Leech, 1998 & 2001, Nesselhauf, 2004, Tono, 2003, and many others). Granger (1998: xxi) asserted that "with roots both in corpus linguistics and second language acquisition (SLA) studies, it uses the methods and tools of corpus linguistics to gain better insights into authentic learner language." To begin with, the chapter reviews in part the territory of SLA including Contrastive Analysis (CA), Error Analysis (EA), and Interlanguage (IL). The first two approaches are regarded as deriving from an earlier period of SLA, while the latter one is considered the most recent development of the same area. After revisiting for the initiation of computer learner corpora, a few prominent learner corpora as well as learner corpora in written languages around the world are introduced. Later, Computer-aided Error Analysis (CEA) and error annotation including some CEA studies are addressed for their insights; the error annotations to be employed in this study are planned based on the learners' first language background as compared to the

target language, similarly as is done in Contrastive Analysis. English and Thai pronouns, afterwards, are compared in linguistic detail.

2.1 Second Language Acquisition Approaches

The birth of Computer Learner Corpora has been from within the areas of SLA, Contrastive Analysis and Error Analysis (Granger, 1998, 2003, and 2008; and Tono, 2003). These are earlier approaches to learner language studies focusing on the product rather than the process of learner language (Corder, 1986; and Ellis, 1994, 1997). Besides the SLA approaches in the opening period of SLA, Interlanguage was proposed by Selinker (1972) for explaining language in between first language and second language as produced by ESL or EFL learners (Gass and Selinker, 2001; and Selinker, 1992); Interlanguage subsequently involves computer learner corpora especially in terms of language development.

The present study aims to find out pronoun patterns over time or the written Interlanguage of pronouns produced by Thai EFL learners by employing a Computer Learner Corpora approach, in particular Computer-aided Error Analysis. However, even the most recent developments in SLA theory and data analysis cannot ignore the roots of the discipline entirely, and since this study uses several of the basic ideas and data analysis techniques of earlier paradigms, a range of relevant areas will be reviewed. This section therefore traces the history of studies of IL, beginning with Contrastive Analysis (CA), and moving on to Error Analysis (EA) studies.

Selinker (1992:4) states that IL has its roots in CA and EA and that the three areas of CA/EA/IL are inescapably linked. This view is also expressed by Corder (1981) in his discussion of EA as a basic learner strategy and a starting point for the

study of transfer in ILs. Each of these approaches developed as a result of changing views of the status of the language produced by learners, changing perceptions of the forces driving development and the need for methodological innovation in the study of this form of language. In the absence of a theoretical framework which adequately describes and explains all second-language acquisition (SLA) data, it is also probably best to take an eclectic approach (Tarone 1994). Furthermore, any study also needs to consider a variety of SLA phenomena including variability, systematicity, transfer, staged development and incompleteness (Towell & Hawkins 1994). Each of the three approaches (CA, EA, IL) discussed below accounts for each of these areas with varying degrees of success.

2.1.1 Contrastive Analysis (CA)

This section provides information about the Contrastive Analysis approach in terms of its major views and problems. The first section will provide a short background and concepts of the approach; and the latter sections will emphasize weak points and criticisms.

2.1.1.1 Major Views of Contrastive Analysis

The study of SLA as we know it today is rooted in early contrastive analysis (CA), which became the dominant approach during the 1950s and 1960s. According to Lado (1957), the purpose of CA is to carefully describe the source language (SL) and the target language (TL) in order to develop effective pedagogical materials. The basic assumption of CA is that learning a second language (L2) reveals the transferring of L1 linguistic features and meanings to L2 by learning different rules. Contrastive analysts predict that some languages might be easier to learn than

others, because where languages differed greatly in structure the learner would be required to automatically employ a more complex rule.

Three main theoretical claims result from CA:

- the learner expects to equally accomplish both the first language and the target language;

- the learner is required to learn the *differences* from the first language when learning the target language; and

- the learner encounter *difficulty* to learn the new language when structures and patterns in the two languages are not the same (Selinker, 1992).

Studies taking a CA perspective, therefore, focused primarily on transfer phenomena and especially negative transfer, which occurs due to structural differences between both languages. As a result of these structural differences, learning a language was not a simple matter of transferring a form directly from L1 to L2. The primary focus of CA studies was therefore on difference and types of difference (Long & Sato, 1983).

Detailed analyses of similarities and differences were carried out by comparing languages in terms of mainly phonology and syntax and, to a lesser extent, semantics (James, 1998; Lado 1957; Weinreich cited in Selinker, 1989). Many later studies of transfer take a contrastive approach, although a detailed CA is not always carried out. The present study will also compare English and Thai pronouns in terms of form and usage with contrastive approach to draw tentative predictions of difficulties and errors of pronoun usage as performed by Thai EFL learners in written mode, but not to draw any further applications due to the limitations of CA.

2.1.1.2 Problems with Contrastive Analysis

Unfortunately, Contrastive Analysis in its original formulation proved to be seriously flawed when applied to data from learners across the world in different language-learning situations. Most seriously, the strong predictions of difficulty and ease of learning which are naturally appealing were not always borne out by studies of learner language. In particular, some researchers found that when there was a great degree of difference between languages, learners seemed to be able to produce the form correctly (i.e. there was no negative transfer), whereas if there was a small degree of difference learners seemed to find it more difficult to produce the correct form (Odlin 1989; Towell & Hawkins 1994). As a result, some areas of error were not predicted by CA (Hyltenstam, 1977). Furthermore, students tended to avoid intricate areas to reduce the possibility of making errors, and the full range of possible errors, therefore, was not available for study in this approach (Schachter, 1974).

From an explanatory point of view, another limitation of CA studies lies in the extreme role of transfer posited by early theorists like Lado (1957). Early theorists believed that language transfer was the main process in SLA. However, later studies show that many errors are not simply traceable to the L1 (Dulay & Burt, 1973; Felix, 1980; Richards, 1974). In addition, some errors may be a result of performance problems (Zobl, 1984) and errors are subject to variability (Zobl, 1982 and 1984). Several theorists concluded that although there is some role for transfer, learners choose in an active and principled way whether or not to transfer and what to transfer (Gass, 1984; Selinker, 1992). Contrastive Analysis does not account for this active role of the learner, because it is primarily interested in the languages as linguistic systems and products rather than in learners using psycholinguistic processes (Long &

Sato 1984; van Els et al. 1984). As a result of the failure of the "strong" version of CA, Wardhaugh (1970) suggested a "weak" version of CA which proposed that the findings of CA could be used to explain transfer after the fact. This version had limited explanatory value, although it was later incorporated as part of EA (James, 1998). From a theoretical perspective, there are profound problems with surface comparisons between languages if one views them as independent systems. For example, it is highly controversial whether one can meaningfully compare English and Thai, because one is comparing the elements of one system with the elements of another system and trying to find similarities and differences across independent systems. The value of the elements may not be equal in each system and these elements may therefore be incomparable. A further problem encountered when comparing learner language cross-linguistically is that a learner and a language may use a variety of strategies and structures to express a particular concept or function (Comrie, 1984), making direct comparison difficult.

2.1.1.3 Conclusion

Contrastive Analysis identifies transfer as one of the key processes for understanding errors in language learning. In short, however, transfer may be one aspect of SLA but it is not able to entirely explain SLA. A thorough explanation of the process of development in SLA is therefore not provided by CA (Towell & Hawkins 1994). Zobl (1984:79) says CA "was not an acquisition theory; or, alternatively, it lacked one". Contrastive analysis was largely discarded during the 1970s, but it has been perpetuated in a modified form in Transfer Analysis (James, 1998). Transfer Analysis is concerned mainly with processes such as cross-linguistic influence (Giacobbe 1992; Kellerman & Sharwood Smith 1986) and language transfer

(Gass & Selinker 1992; Odlin 1989). This newer version of CA is not the same as the original because it is a comparison between IL and mother tongue, and not the mother tongue and a target language (James 1998); the present study will borrow a transfer analysis perspective occasionally though English and Thai are structurally very different.

2.1.2 Error Analysis (EA)

This discussion of the Error Analysis approach is arranged into three main sections: characteristics of error analysis studies, limitations of the approach and distinctions between errors and mistakes.

2.1.2.1 Characteristics of Error Analysis Studies

During the 1970s, Error Analysis (EA) emerged as the next major development in the study of SLA. Although the studies sometimes attempted to explain how second languages are learnt, EA remained primarily a methodological approach rather than a theory of SLA. The central focus in EA is on the L2 learner rather than on the system, as in CA (van Els et al. 1984), although the "systematicity in development and the common processes posited to explain development" (Long & Sato 1984:256) became central features of EA. As the pendulum swung away from CA, theorists claimed that errors could be explained in terms of the target language (TL) only, with no reference whatsoever to the source language (SL), i.e. there was no transfer. Actually, this proved to be too extreme a view once data were cautiously analyzed.

The EA studies focused mainly on the performance of learners of a few languages who had learned in formal contexts and were studied in experimental conditions. These studies had two primary purposes. Firstly, the studies had been

conducted for making explanations available for errors, which were characterized as *learner-internal cognitive processes* and *learner-external causes*. Learner-internal errors include: *overgeneralizations* (Schumann, 1978); *developmental patterns* which remain the same regardless of the L1 of the learner (Dulay& Burt, 1974; Huang, 2000; Huang & Hatch, 1978); *ignorance of rule restrictions* (Richards 1985); *incomplete application of rules* (Richards 1985); and *learning strategies* (Selinker 1972).

Secondly, learners' external errors related to problems with the *input* received by the learner, mainly in the case of formal instruction (Faerch et al. 1984) and included errors such as transfer of training (Felix 1981; Selinker 1972). As a result of finding learner-external causes of errors, the second, later purpose of EA was to relate the social context of learning to the errors produced (Faerch et al. 1984; Selinker, 1983). These studies directly observed the input available to the learners and studied the errors related to the input received.

In the present study explanations of pronoun errors will be made; also existing input of pronoun will be explored. However, the approach has some limitations to be aware of when discussed.

2.1.2.2 Limitations of Error Analysis

EA had flourished during 1970s when experts of foreign language learning viewed errors as tickets to explore learners' interlanguage. Unfortunately, EA underwent criticism due to the emergence of a number of major weak points (Dagneaux et al, 1988; Granger, 2003; Spillner, 1991; Tono, 2003):

1) Unclear descriptions of various learner data in EA studies;

2) Unclear error categories in EA;

3) Unable to provide explanations for such phenomena as avoidance;

4) Restricted only to weakness of the learner; and

5) Fixed explanations for second language learning.

The first two limitations are methodological. With respect to the data, Ellis (1994: 49) highlights "the importance of collecting well-defined samples of learner language so that clear statements can be made regarding what kinds of errors the learners produce and under what conditions" and regrets that "many EA studies have not paid enough attention to these factors, with the result that they are difficult to interpret and almost impossible to replicate". The problem is combined with the issue of error categories used, which also suffers from various weak points: they are often ill-defined, rest on mixed criteria and involve a high degree of subjectivity. Terms such as "grammatical errors" or "lexical errors", for instance, are rarely defined, making results difficult to interpret and analyze, as several error types--prepositional errors for instance--fall somewhere in between and it is usually impossible to know in which of the two categories they have been counted. In addition, the error typologies often mix two levels of analysis: description and explanation. Scholfield (1995) illustrates this with a typology made up of the following four categories: spelling errors, grammatical errors, vocabulary errors and L1 induced errors. As spelling, grammar and vocabulary errors may also be influenced by L1, there is an overlap between the categories, which is "a sure sign of a faulty scale" (Scholfield, 1995, p. 190).

The other limitations deal with the scope of EA. EA's limited focus on overt errors means that both non-errors, i.e. instances of correct use, and non-use or

underuse of words and structures are disregarded. It is this problem in fact that Harley (1980: 4) is referring to when he writes:

> '[It] is equally important to determine whether the learner's use of 'correct' forms approximates that of the native speaker. Does the learner's speech evidence the same contrasts between the observed unit and other units that are related in the target system? Are there some units that he uses less frequently than the native speaker, some that he does not use at all?"

In addition, the picture of interlanguage described by EA studies is very fixed: "EA has too often remained a static, product-oriented type of research, whereas L2 learning processes require a dynamic approach focusing on the actual course of the process" (van Els et al., 1984, p. 66).

Although these weaknesses considerably reduce the usefulness of past EA studies, they do not call into question the validity of the EA enterprise as a whole but highlight the need for a new direction in EA studies. One possible direction, grounded in the fast growing field of computer learner corpus research, is sketched in section 2.3 about Computer-aided Error Analysis and its potential to reduce such limitations.

2.1.2.3 Errors vs. Mistakes

At the level of analysis, deciding whether a deviation is an error or a mistake is another problem. Corder (1967, 1981, and 1983) argues that mistakes should not be included in the quantification or analysis of errors and this is the approach taken by most analysts. Johnson (1988) believes that mistakes can be corrected by the learner, but in practice determining whether a learner cannot correct his/her own deviant utterances is very problematic (James, 1998). Errors can be found when the learners have not known the rules and required teaching with explicit error examples in particular situations (Shaughnessy, 1977). In a different view, Edge

(1989) rejects this error-mistake classification and calls all deviations from the norm mistakes. These mistakes include:

- *slips*, which are a result of "processing problems or carelessness" (Edge 1989:11);

- *errors*, which are comprehensible but which the learner cannot correct, although the form has been taught; and

- *attempts*, which are moderately incomprehensible and uncorrectable by the learner.

Snow (cited in James 1998) argues for two steps in error development. The first step is the presence of errors which the learner does not recognize as errors, and the second step is the presence of errors that the learner recognizes as errors but which he/she cannot correct. The mistake, where the learner is able to correct a wrong form, may be a third step. In other words, mistakes are a performance problem rather than a competence problem (Corder, 1967), rather like the lapses made by L1 speakers (Johnson, 1988). This performance-competence distinction is maintained by most theorists in distinguishing errors from mistakes.

Another way of determining whether a deviant form should be classified as an error or a mistake is to decide on the gravity of the error. In order to do this, James (1994:191) believes that criteria for error gravity need to be established (e.g. "are lexical errors more serious than grammatical?"), as well as who will judge the gravity (e.g. L1 teachers/L2 teachers/non-teachers).

An additional criterion is that errors have insufficient intention of the speaker; otherwise they may be classified as deviances (James, 1998). The classification of an utterance as deviant is further confused by the distinction between unacceptability and

ungrammaticality; e.g. a grammatical utterance may be unacceptable because of non-linguistic factors (de Beaugrande & Dressler, 1981). Acceptability is judged by use in a particular context, while grammaticality can be judged by a native speaker of the language and a grammatical utterance is necessarily acceptable as well (Lyons 1968).

In this study, all errors of pronoun produced by Thai university English majors will be counted entirely as *errors* in different categories, including underuse or avoidance, identified in section 2.3 of this chapter.

2.1.2.4 Conclusion

Despite the above criticisms and methodological difficulties, there is evidence of a more positive approach to Error Analysis in some recent writing. There are two reasons for the continued use of EA when investigating SLA data. Firstly, the empirical design is simple, with a clear indication of an error if a particular norm is chosen. Secondly, teachers play this normative role and encourage their students to achieve these target norms. Ignorance of errors may be held by some teachers, but many SLA theorists tend to regard errors in a much more positive way because they regard them as signs of creative hypothesis construction and testing.

In conclusion, James (1998) feels that those doing IL studies and those engaged in EA analyses have different goals, the former concerned with developing a theory of acquisition and the latter with pedagogic goals. However, EA is an effective way of dealing with data given the lack of an appropriate analytical framework in IL studies (Cook, 1993). It is thus for this reason that the EA idea has been carried out on the data in the present study with the innovative approach of Computer-aided Error Analysis (CEA). CEA alone does not provide a sufficient description or explanation of learner language, but it has a significant contribution to make as part of an analysis

of this type of language because it can offer insights into the sequence of pronoun acquisition, the patterns of pronoun acquisition and the types of pronoun which learners find difficult. On top of that, CEA could reduce limitations of the traditional approach, especially errors in context; however, the picture of learners' pronoun errors in a specific group will be informed here. Interlanguage (IL), therefore, supplements explanation of pronoun used by Thai EFL learners over time in the study.

2.1.3 Interlanguage and Language Development (IL)

Based on the research question number 3 *"How do language development patterns of pronouns occur among the groups over time?"*, the researcher assumes the more English courses the students have taken, the more English pronoun knowledge they have probably acquired. Therefore, the results among the subjects should be diverse and the researcher should find various patterns of language development. Interlanguage and language development are discussed as aspects of Second Language Acquisition (SLA).

The term Interlanguage (IL), which refers to the language produced by learners acquiring an additional language, was coined by Selinker (1972), although studies of the IL phenomenon and the notion that learner language should be studied as an autonomous system predate the term by several years (e.g. Corder 1967). Some writers (e.g. Lalleman 1996) classify IL as part of the Creative Construction Hypothesis (CCH) paradigm which has as its central tenet that the learner is an active participant in the learning process and produces a system which is not too different from other natural languages (Long & Sato 1984). On the other hand, Sharwood Smith (1994) argues against IL falling under the CCH because this branch of research

was not concerned with interim grammars but rather with finding common patterns of development.

This discussion of Interlanguage will be presented from the perspectives of composition of the *Interlanguage system, features of Interlanguage, Interlanguage development,* and *development of Interlanguage system.*

2.1.3.1 Composition of the Interlanguage System

Bialystok & Sharwood Smith (1985) pointed out that there are two ways of viewing IL. On the one hand, it can be seen as the product of a set of highly structured hypotheses making up the underlying competence. The product is revealed in performance. This product can be studied as a diachronic series of systems which develop as second-language acquisition (SLA) progresses or as a synchronic language system at a particular point in time (Cooreman & Kilborn, 1991). On the other hand, IL can be seen as the system of underlying competences which needs to be investigated in terms of the psychological processes at work. It seems then that both the product and the system need to be investigated.

There are different views of the "composition of the IL system" (Bialystok & Sharwood Smith 1985:102-103). Some, like Selinker, see it as a "single system composed of rules which have been developed via different processes", while others, like Adjémian (1976), see the system as composed of "a combination of separate knowledge sources" consisting of the L1 and an L2-based system. Robert (1989:219-220) views a language as being composed of two aspects: a "computational system", which is the set of rules for structuring syntax, phonology and semantics, and a "conceptual system" which relates to thematic relations. Whereas full-fledged languages draw on both systems to a great extent, it is claimed that IL users rely on

the conceptual system more than on the computational system. The conceptual system is more universal whereas the computational system would be language-specific.

Bialystok & Sharwood Smith (1985) argued instead for a knowledge component (underlying linguistic system) and a control component (retrieval procedures) in the description of IL (see also Cook cited in Towell 1987). The control component is the set of retrieval procedures used to access the underlying linguistic system. Either or both of these components need to be considered in the analysis of IL data and may be different from or similar to those of the native speaker (see also Towell 1987). Their definition of IL is therefore:

> ... the systematic language performance (in production and recognition of utterances) by second-language learners who have not achieved sufficient levels of analysis of linguistic knowledge or control of processing to be identified completely with native speakers. (Bialystok & Sharwood Smith 1985:116)

Second-language acquisition development in this view requires increased knowledge and/or better control.

Universal Elements in Interlanguage

The IL may also consist of a universal component, connected to the Chomskyan notion of a Universal Grammar (UG). Despite extensive writing in the area (e.g. Cook 1996; Felix 1995; Flynn 1984; Gass 1984, 1995; Schachter 1992; Sharwood Smith 1994; White 1992, 1998), theorists cannot agree whether IL is a natural language which is UG-based. It is possible, if interlingual identifications (IIs) are not based on UG (which might be the case if UG is unavailable to older learners), that the new system is entirely separate from the native language (NL) and the TL. On the other hand, another argument is that the system is partially separate, has to be based on the NL and the TL to some extent because of transfer, is linked via IIs and

should be investigable in a UG framework (e.g. White cited in Selinker 1992). A further possibility is that IL develops from UG with a "learning mechanism which incorporates a theory of markedness" (Zobl 1992:176).

Formulaic Elements

Formulaic utterances are a major component of the IL system. They serve a vitally important communicative function and may be unanalyzed chunks of language at early stages (Altman 1997; Klein 1986). For example, Altman (1997) points out that inflected verbs may be produced as unanalyzed wholes but this does not mean that the learner can use the rules productively to form other inflected verbs. Apart from their communicative function, Fillmore (cited in Myles et al. 1998) suggests that formulaic expressions may also be useful data on which language learning is based because they can be used as frames for further development. Similarly, Myles et al. (1998, 1999) who studied the development of pronominal systems, conclude that formulaic expressions aid communication in the early stages and that they are used as the basis of hypothesis testing. Hakuta's (1974) discussion of formulaic expressions shows that learners perceive formulaic expressions as units, both syntactically and semantically.

Towell & Hawkins (1994) discuss the different origins of the elements of IL and conclude that some of them have UG origins, some have transfer origins, some are formulaic, and some are a result of learned linguistic knowledge. It appears then that there is a knowledge system which may consist of hypotheses about the TL, universal elements such as a conceptual system relating to thematic relations, and formulaic utterances. In addition, there is a control system which consists of

procedures to access the knowledge system. The product of the knowledge and control systems will be discussed in the next section.

2.1.3.2 Features of Interlanguage

There have been a variety of attempts at describing IL features and comparisons have been made between IL, child language and pidgins, because some features appear to be similar (Andersen 1983). Kachru (cited in Williams 1989) found that many features of non-native institutionalized varieties of English (NIVES) can be found in learner language and suggests that they have their roots in individual SLA. Theorists comparing IL to "reduced" forms of language like child language and pidgins see it as "a variety of language which is both formally and communicatively reduced when compared to languages used as native languages by adults" (Faerch et al. 1984:271). However, comparing IL with other non-native forms of language has met with criticism because some (e.g. Bickerton 1983; Myhill 1991) believe that the social contexts in which each occur are incomparable, although Adamson (1989) suggests that a common bioprogram guides development in these different contexts. Mitchell & Myles (1998) also conclude that pidgins and ILs do not share all their syntactic features and that pidgin systems are more stable than most ILs.

However, Givón (1979) provides a framework which describes the features of different forms of language, in which he distinguishes a pragmatic mode of communication which occurs in "reduced" languages and a syntactic mode of communication which occurs in fully-developed languages. Table 2.1 presents some of the features of the two modes.

Table 2.1 Features of the pragmatic and syntactic modes

Pragmatic Mode	Syntactic Mode
1. Topic-comment structure	1. Subject-predicate structure
2. Free coordination	2. Fixed coordination
3. Slow rate of delivery	3. Fast rate of delivery
4. Small chunks under one intonation contour	4. Large chunks under one intonation contour
	5. Higher noun/verb ratio
5. Lower noun/verb ratio in discourse	6. More complex verbs
6. More simple verbs	7. Extensive use of grammatical
7. No use of grammatical morphology	morphology

(Adapted from Givón 1979:98)

Sato (1988) reports that only the lack of grammatical IL morphology has been studied extensively and that the evidence of a low noun/verb ratio is not conclusive, although Perdue's (1993) study supports both lacks of morphology and low noun/verb rations. Pienemann (1992) believes that ILs have a common lexicon, part of the rule system of the TL and a set of idiosyncratic rules. However, these aspects have not been fully explored and a full set of lexical items and rules is not available. A more detailed account of the features of IL (or "simple codes") is supplied by Corder (1977:2):

> ... a simple or virtually non-existent morphological system, a more-or-less fixed word order, a simple personal pronoun system, a small number of grammatical function words and grammatical categories, little or no use of the copula, absence of an article system (less often the absence of deictic words). The semantic functions of these and other systematic systems such as tense and aspect are typically performed, when at all, by lexical means, e.g. adverbs, or some "imperial form". The basic syntactic relations are expressed by word order.

The Interlanguage does not therefore consist of the same phonological, morphological and syntactic categories of the target language (TL) or the source language (SL), but it is still unclear to what degree the system is a simplified version of full-fledged languages and to what degree it is based on universal categories.

The present study aims to identify and analyze English pronoun errors occurring in written contexts by Thai EFL learners; however, the error categories defined in the study will not differentiate between syntactic or discoursal errors. More explanations will be made in section 2.3.

2.1.3.3 Interlanguage Development

a) General Development

The hypothesis of the starting point of learner language is controversial in SLA theory. Contrastive analysts believed that the L1 is the starting point with successive restructuring until the learner's language approximates the TL (Lado 1957). Development occurs when the basic system is elaborated by incorporating the rules of the TL. Corder (cited in Gass 1984a) believes that this is the correct position for phonology, but Zobl (1984:85) criticizes this perspective because he feels that "(i)n restructuring distinctions are lost and subsequently elaborated along different lines ... unmarking or despecification of the L1 grammar would have to take place for any movement to occur in the direction of the target". He also refutes the idea that the L1 in its full form is the starting point of IL because of empirical data showing that certain structures that one would expect to be transferred because of their equivalent positions in the L2 are in fact not transferred.

He joins others (e.g. Sharwood Smith & Rutherford cited in Selinker 1992) in believing that the initial hypothesis is an independent system (core grammar, IL base)

which resides in the learner, especially for syntax (Corder cited in Gass 1984a). This core grammar might be made up of universal rules of human language (Selinker 1984, 1992) and, according to Corder (cited in Gass 1984), it is based on the system used when the L1 developed. This may be akin to the notion of Universal Grammar (UG), although Corder does not commit himself on this point. His 1977 paper suggests that semantic aspects such as agency, animacy and spatial location may be sufficient to analyze IL data. He says that the common features found in the ILs of learners with diverse L1 backgrounds may be due to a return to the basic system, rather than a need to use the L1 to any great extent. The pull towards the basic code is therefore stronger than any L1 influences. However, Selinker (1992:33-34) feels that a combination of approaches is the correct one:

> My hypothesis is that the NL is part of where the L2 learner has to be
> on day one of exposure to input from the TL, because of the pervasive
> reality of language transfer and, therefore, interlingual identifications.
> But it cannot be the entire starting point because of the reality of early
> fossilization of non-L1-like structures.

b) Simplification

Simple systems and simplification processes are controversial issues in IL writing. The core system is a simple system (but not a simplified version of the NL or the TL because the learner does not know the TL, according to Valdman, 1977). This simple system is elaborated as the learner learns more about the TL, with the possibility that development is arrested at some point, resulting in fossilization. Explaining the simplification which has taken place in simple codes, Corder (1977) discusses two views of simplification. The first is that IL appears simple because it is compared to a native language (NL), which means that ILs are described in terms of

the complexities of these fully-fledged languages. This is the view taken by Silva-Corvalán (1991:330) when she says linguistic simplification includes "reduction of the inventory of linguistic forms, semantic range, or language functions, and the elimination of alternative structures at certain levels". She also sees it as generalization or overgeneralization (see also Preston cited in Silva-Corvalán 1991), which involves expanding the use of one form at the expense of another form, resulting in a loss of variety in the forms used. In SLA, forms are not being lost, but the overgeneralization of a form is the starting point with progressive acquisition of further forms in a process which is the mirror-image of language loss. Simplification in ILs may therefore be the result of universal simplification processes, which removes the link to a specific native language.

The second way of looking at simplification, which Corder (1977) also examines, is that NLs are complex forms of simple codes and that there are language-specific complexification strategies or rules (see also Klein & Perdue 1992 for a similar conclusion). According to Baker (1979), from a learnability perspective, the idea of a simple code which is elaborated is a better one.

2.1.3.4 Development of the Interlanguage System

When discussing further development of the IL system, it is essential to remember that internal systematicity is a central principle of IL. Ellis (1985a) points out that IL is systematic from both horizontal (synchronic) and vertical (diachronic) perspectives. Although it does not necessarily detract from the systematic nature of IL, it is also necessary to posit a degree of variability. However, Sharwood Smith (1994) sees systematicity and variability as incompatible, and he argues that absolute systematicity and independence of the system has to be an idealization

because of the dynamic quality of the learner's IL. Development may be viewed in two ways: either as leaps from stage to stage, or as the gradual spread of a rule to different areas and contexts of the IL. Both views lead to a role for variability which, although it may be random at times, is more often contextually derived, with influence from the L1 (permeability) and internal development having an effect on the choices made. The task may also have an effect on variability (Tarone 1982, 1983, 1989). Any analysis of IL therefore requires an account of situational, contextual and linguistic factors as well as their interaction, as these factors may influence the product of the system (Sato 1988, 1990).

The focus on L1 and developmental errors is a combination of a contrastive analysis (CA) approach, cognitive processes (Gass 1979; Kellerman 1977, 1979; Sharwood Smith 1979), and developmental processes (Andersen 1983). Learning happens gradually as the process involves learning forms, meanings and functions (Larsen-Freeman 1991) while rules change over time (Ellis 1987a and b; Selinker & Douglas 1985). Rules need to cover all aspects of language from the phonological to the discoursal levels (Klein 1991), and these rules must allow for interaction, since each level of language is not learned in an isolated fashion. Form-meaning relationships need to be developed since

> ... there are also forms whose relationship with meaning is difficult to access in the L2. These forms carry little semantic weight or have little perceptual salience, or the form meaning relationship may be difficult to grasp. ... Learners also need data as they construct or set their interlanguage. They need to know how their interlanguage differs from the L2. It might be said that they need to know what is ungrammatical, but since interlanguage is systematic and, therefore, grammatical in its own way, one might simply say that learners need to know what in their interlanguage is inconsistent with the L2. Finally, learners need to have

data on the potential of their interlanguage for expressing relationships of
form and meaning as well as the extent to which they can modify and
restructure their interlanguage toward L2 morphosyntax. (Pica 1998:11)

As new knowledge is gained and integrated into the underlying system, the functions of other items in the system are narrowed or broadened and elements of the system are restructured or rejected. Interlanguage does not develop in a linear fashion but is recursive and continually restructured (Corder 1992). Carroll (1984) finds in her study of English learners of German that IL data shows periods of focus where there is an inconsistent number of one type of structure which has been newly acquired at the expense of other forms which may have been used earlier. The notion of staged development from one IL to another might not be a feasible one, as restructuring may not necessarily move the learner forward but may in fact move him/her further away from the TL if an incorrect hypothesis is made. Alternatively, it may create confusion, so that the learner temporarily reverts to an earlier stage (backsliding). In addition, the degree of variation apparent in an IL at any one time may preclude the notion of discretely staged development. Corder (1992) prefers to use this model for phonological development, but he does not see this as necessarily feasible in the case of other subsystems of language, partly because of the lack of clarity about the starting point of IL. As a result, it is not always possible to predict that learners will progress in exactly the same way and improve their competence in a particular element of the system. For example, Clahsen (1995), discussing the development of German plurals, found that although some learners increased their accuracy over time, some learners stayed at about the same rate of accuracy even though other elements of their grammar did develop. Over the two year period, none of the learners he investigated reached 100% accuracy.

Cazden (1968) offers a classification into *four periods* in L1 acquisition: no inflection is apparent; formulaic utterances are produced (there are no errors but little communication beyond the set of known utterances); much communication with many errors and overgeneralizations; and 90% correct use of a form.

These L1 acquisition stages seem to be true of L2 development as well (Klein 1986). With regard to the grammaticalization of utterances, Skiba & Dittmar (1992) conclude that the first stage is where learners put words next to each other but they are not explicitly related. The second stage occurs when syntax starts to develop and relations between words are shown more explicitly. The third stage is where the IL and TL start to converge, and syntactic and morphological relations are fairly target-like. Bardovi-Harlig & Bofman (1989) conclude that advanced learners are better at syntax than at morphology because they focus on what is communicatively important rather than on what is less important (and often redundant) to communication.

Fodor & Crain (cited in Zobl 1992) feel that the conservativeness of the learning process is a result of hypothesis-testing which relies on the input and unmarked versions of a grammar. Less conservative development only occurs once marked versions of a structure are noticed in the input. Beck et al. (1995) and Schwartz (1993) argue that positive evidence is the main factor in L2 grammar development. Although they do not rule out a role for negative evidence in SLA, they do not believe that it plays a role as evidence in constructing L2 grammar systems.

Schachter (1992) says that hypotheses are formulated on the basis of experience with the language and input and that the formulation-testing-acceptance/rejection process is cyclical and continuous, with the learner focusing at different times on different hypotheses.

Common to both versions of the cognitive approach is the idea that L2 learners initially decode, analyze, store and produce - i.e. process - material from the new language in ways which are determined by general cognitive factors like the 'perceptual saliency' of the material, the 'continuity' of elements in that material, the basic 'conservatism' of earners in not extending hypotheses to domains not warranted by the input. This approach considers that people perceive events in terms of 'actors', 'actions' and 'persons or things acted upon', and that these are more 'salient' than the place where the event took place, or the time it took place, or the manner in which it took place. By extension it is considered that L2 learners will attend to and acquire new ways of expressing 'actors', 'actions' and 'people or things acted upon' before they will attend to and acquire adverbials dealing with the place, time and manner of the event. (Towell & Hawkins 1994:46)

Tarone (1988) and Ellis (1992) believe that noticing is the crucial aspect for inclusion of a new structure in the grammar. Ellis (1992) sees development spreading according to tasks or contexts. However, Towell & Hawkins (1994) feel that this view implies a random entry point by the learner rather than an entry point which can be related to the L1. They cite as evidence studies of learners of English and French which show that these learners differ systematically in where they place the object pronoun, with French learners easily learning to place the pronoun postverbally in English, but English learners going through a stage of placing the French pronoun postverbally rather than preverbally (the same sequence is found for English learners of Spanish according to Andersen, 1991).

This section has discussed the general features of IL development which seem to result from a series of cycles of hypothesis-making and hypothesis-testing, possibly limited by a universal set of constraints on what kind of hypotheses can be made. The performance displayed by the learner is subject to variability according to task and

level of development. As this section has focused mainly on the general features of development, an area which is particularly relevant to the present study. English pronoun IL development might be found over time among the Thai university English majors; both pronoun accuracy and errors will be taken in account.

2.1.3.5 Conclusions regarding Interlanguage

Selinker's notion of Interlanguage is a valuable one which has had an enormous impact on the way in which SLA theorists conceive of learner language. For example, interlanguage is no longer widely viewed amongst most theorists as a poor approximation of the TL, but a system in its own right which is regular and rule-governed. It is therefore equal to other languages, although there are notable differences between ILs and fully-fledged languages. Firstly, it is a system which can change quite rapidly and it may have a higher degree of variability at any one time or over time. Secondly, it is idiosyncratic to the learner to some extent, although there seem to be common developmental routes across learners from the same and different language backgrounds. Thirdly, some learners fossilize before they have a fully functional system which could be used in as wide a range of contexts as their L1s.

To the present study Interlanguage will be employed as fundamental framework for analyzing learners' pronoun and pronoun errors over time, associated with Computer-aided Error Analysis approach discussed later in sections 2.3.

2.1.4 Conclusion

The tripartite SLA approaches: Contrastive Analysis (CA), Error Analysis (EA), and Interlanguage (IL) are employed in the present study as the central hypothetical framework for the data interpretation and discussions. CA supports ideas for predetermined tentative errors of pronoun performed by Thai EFL learners; EA

assists categorizing errors of pronoun in writings and; IL provides explanations of pronoun development over time among the Thai EFL learners. All three fundamental approaches will be expected to explain how performance of English pronoun is so in the students' narratives, as well as how the errors occur. In addition, differences across the groups of English majors are anticipated to clarify tentative students' language development of English pronoun over time. The approaches are employed in the computer learner corpus-based approaches as discussed in the next section.

2.2 Computer Learner Corpora

As discussed in the prior section, Computer Learner Corpora (CLC) originated from such SLA approaches as CA, EA, and IL; however, CLC studies have been considered as distinct from SLA on two bases: better functions (that is, better descriptions of interlanguage and better understanding of factors influencing L2 learning) and better pedagogical application (Granger, 2008; Nesselhauf, 2004; Tono, 2003). This section focuses on Computer Learner Corpora and its current situation around the world. In addition, CLC approaches of Contrastive Interlanguage Analysis (CIA) and Computer-aided Error Analysis (CEA) are provided in the final part of the section.

2.2.1 Definitions and Typology

2.2.1.1 Definitions of Computer Learner Corpora

The definitions used in the literature can be chronologically identified (Bowker and Pearson, 2002; Dagneux et al, 1998; Granger, 2002; Leech, 1998; Nesselhuaf, 2004 and 2005; McEnery et al, 2006) as presented below.

Firstly, Leech (1998: XV) addresses learner corpus as 'a computerized textual database of the language produced by foreign language learners'

For Dagneaux, learner corpus is "a collection of machine-readable natural language data produced by L2 learners" (Dagneaux et al., 1998: 165).

Following Granger (2002: 124), learner corpora are defined as 'electronic collections of spoken or written texts produced by foreign or second language learners in a variety of language settings.'

Another definition of learner corpora from Bowker and Pearson (2002: 9) is 'a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria'.

Nesselhuaf (2004: 125; 2005: 40) defines learner corpora as 'systematic computerized collections of text produced by language learners.'

Finally, McEnery et al (2006: 65) state that learner corpus is 'a collection of the writing or speech of learners acquiring a second language (L2).'

In brief, since the definitions given by the scholars are nearly identical; a conclusive definition of leaner corpora can be merged as '*systematic* collections of *spoken* or *written language*, in a computerized textual form, produced by *ESL or EFL learners* in and *variety of language settings*' which provide pictures of various factors possibly found in studies of learner language. The definition truly relates to research design since it involves language, tasks, and learners. Researchers should take these criteria into considerations when designing studies of learner corpora, which will be extensively addressed in the next section, in section 2.2.3, and again in Chapter 3, Research Methodologies.

2.2.1.2 Typology in Computer Learner Corpora

In worldwide perspective of how learner corpora are collected around the world (Pravec, 2002), learner corpora studies investigated with the same characteristics (Granger, 1998, 2003, 2008; Granger et al, 2002). The current situation of typology of learner corpora studies are: (a) objective (commercial or academic); (b) size of corpora (big or small); (c) target language (English or non-English); (d) language medium (writing or speech); and (e) methodology (longitudinal or cross-sectional).

(a) Objectives of Learner Corpora:

Computer learner corpora can be grouped by objectives, whether it is academically or commercially conducted. There are only a few commercial corpora, often investigated by business firms, most of them publishing companies; while there are a greater number of learner corpora projects conducted with academic purposes. However, the commercial ones tend to provide greater data than do the academic ones. The most well-known commercial corpora are Longman Learners' Corpus and the Cambridge Learner Corpus, providing more than 10 million words and presenting numerous learners' L1 backgrounds. In contrast, the academic corpora generally provide much smaller size of corpora with a limited range of learners' L1 backgrounds. The exception is the International Corpus of Learner English (ICLE) with its notable large corpus size and multiple backgrounds of learners from a large number of L1s.

(b) Size of Corpora

Due to learner corpus data collecting and storing, a size of corpora in each project can grow speedily. Accordingly, most of the learner corpora projects approximately hold a million words rather than hundreds or thousands; however, a small number of words in a corpus is always a starting point for a larger corpus for better generalization. It is worthwhile to investigate small corpora; Ragan (2001) mentioned that the size is not as important as the planning and tracking of the language product and its application. In addition, to conduct a longitudinal study with a corpus of small size can be valuable since the information obtained will trace back how individual language develops over time.

(c) Target Language

Although English clearly dominates the outlook of learner corpora studies, other non English corpora have become more in focus according to an increasing number of such projects of non-English learner languages. So far there have been at least 65 projects conducting learner corpora of English around the world. Probably the largest projects are gathered in ICLE, with 2.5-3 million words produced by learners of English with 11 different mother tongues (Granger et al, 2002; Granger, 2003), excluding sub-projects independently carried out in other countries such as The Br-ICLE corpus in Brazil, The GICLE corpus in German, and The PICLE corpus in Poland. Additionally, Hong Kong University of Science and Technology Learner Corpus with 25 million words was obtained from Chinese learners of English (Milton, 1998). Other smaller projects are, for examples, EVA corpus of spoken language produced by Norwegians (Hasselgren, 1997), APU Spanish learner corpus (Ife, 2004), and Japanese speech in NICT JLE corpus (Izumi et al, 2004).

Aside from the popularity of English corpus projects, some corpus projects investigate languages other than English, e.g. French (Debrock et al, 1996), Norwegian (Tenfjord et al, 2004), Dutch (Degand and Perrez, 2004) Spanish (Ife, 2004), and German (Lüdeling et al, 2005). Interestingly, Multilingual Learner Corpus has collected data from learners with a single L1 background (Brazilian Portuguese) learning different target languages such as English, German, and Spanish (Tagnin, 2003). Such studies have become more popular; at least five projects have obtained data from learners with a common L1, for instances, Polish, Catalan, and Italian.

(d) Medium

Learner written corpora have clearly dominated over learner spoken corpora (Pravec, 2002). According to the current situation of learner corpora around the world reported below in the section 2.2, there are 70 English corpora projects in total. 53 projects out of 70 examine learner corpora in written texts; while only 17 projects investigate learner spoken corpora. Mixed medium of both written and spoken corpora have been studied in 7 projects. Remarkably, there are 2 projects examining learner language by employing information and communication technologies obtaining data from CMC (computer-mediated communication) courses (Howard, 2009; Howard and van Moere, 2002) and telecollaborative communication in 5 years, from 2000-2005 (Reder et al, 2003).

(e) Methodological Design

Cross-sectional methodological design is used for studying interlanguage data collected from different subjects with different proficiency levels at the same time; while real longitudinal corpora constitutes only 10% of known corpora or 7 out of 70 projects (for example, the Barcelona English Language Corpus,

the Corpus of Young Learner Interlanguage (CYLIL), and the LONGDALE project) collecting data from the same subjects over time.

2.2.2 Computer Learner Corpora around the World

To research information about CLC around the world, the information collected by Pravec (2002), with nine major centers of learner corpora, and one provided in ICLE website (http://www.uclouvain.be/en-cecl-lcWorld.html) are truly the best comprehensive starting point. Studies of learner corpora have been conducted for more than twenty years since the late 1980s (Granger, 2008); however, the studies in this field are relatively at an infant stage; as presented in the table 2.2 below that there have been 70 main projects so far actively conducting learner corpora. The corpora are categorized in terms of: the learners' L1 backgrounds; language medium; text and task type; learner level & English environment; size in number of words; and methodological design & objectives.

Actually, there are 105 projects examining learner corpora around the world; 35 of them conduct research of corpora in languages other than English. As a result, 70 projects of learner corpora in English will be presented in this section. Table 2.2 below summarizes English learner corpora around the word.

Table 2.2 English Learner Corpora Projects around the World

Corpus	L1	Medium	Text type / task type	Proficiency level/ Learners	Size in words	Methodological Design: Objective
1) Asao Kojiro's Learner Corpus Data	Japanese	written	essays and stories written or reproduced	EFL Japanese college students		academic
2) The Barcelona English Language Corpus (BELC)	Spanish Catalan	written	4 tasks: written composition	ESL children and young adults learning English	2 m	longitudinal data: academic
		spoken	oral narrative oral interview role-play			
3) The Bilingual Corpus of Chinese English Learners (BICCEL)	Chinese	written	in-class assignments	ESL		academic
		spoken	National Oral English test			
4) The Br-ICLE corpus (Brazilian component of ICLE)	Brazilian Portuguese	written	Argumentative and literary essays	ESL		academic
5) The British Academic Spoken English (BASE) corpus	Mainly L1 + data by native speakers	spoken	160 lectures and 40 seminars VDO + audio recorded in a variety of university departments	University students	1.6 m	academic/ commercial
6) The British Academic Written English (BAWE) corpus	Mainly L1 + data by native speakers	written	ESP papers	4 levels of study (undergraduate levels to final year and taught masters level)	6.5 m	academic
7) The BUiD Arab Learner Corpus (BALC)	Arabic	written	school examination essays	EFL various	287,227	academic
8) The Cambridge Learner Corpus (CLC)	various	written	exam scripts	various	30 m - still expanding	commercial
9) The Corpus of Academic Learner English (CALE)	German	written	various written text types: term papers/ reports/ research plans, abstracts/ reviews /summaries	advanced ESL		academic
10) The Corpus of English Essays Written by Asian University Students (CEEAUS)	various	written	student essays	EFL various	200,000	academic

Table 2.2 English Learner Corpora Projects around the World (Continued)

Corpus	L1	Medium	Text type / task type	Proficiency level/ Learners	Size in words	Methodolo gical Design: Objective
11) The Chinese Academic Written English (CAWE) corpus	Chinese	written	dissertations	EFL Chinese undergraduate s English or linguistic related majors	407,960	
12) The Chinese Learner English Corpus (CLEC)	Chinese	written		EFL various	1 m	academic
13) The City University Corpus of Academic Spoken English (CUCASE)	Chinese + data by Eng native speakers	written & spoken		EFL	2 m	academic
14) The Cologne-Hanover Advanced Learner Corpus (CHALC)	German	written	term papers and essays	Advanced ESL	210,000	academic
15) College Learners' Spoken English Corpus (COLSEC)	Chinese	spoken	National spoken English test	EFL university students: non-English majors	700,000	academic
16) The Corpus Archive of Learner English in Sabah/ Sarawak (CALES)	Malay	written	argumentative essays	EFL various	400,000	
17) The Corpus of Young Learner Interlanguage (CYLIL)	various: Dutch French Greek Italian	spoken		ESL various levels of European School pupils	500,000	longitudinal data: academic
18) The English of Malaysian School Students corpus (EMAS)	Malay	written	student essays	EFL various	500,000	academic
19) The English Speech Corpus of Chinese Learners (ESCCL)	Chinese	spoken	dialogue reading-aloud	EFL middle school and college	35,000	academic
20) The EVA Corpus of Norwegian School English	Norwegian	spoken	picture-based tasks	ESL		academic
21) The GICLE corpus (German component of ICLE)	German	written	mainly non-academic argumentative essays	ESL advanced	234,000	academic
22) The Giessen-Long Beach Chaplin Corpus (GLBCC)	German	spoken	Transcribed interactions between native English speakers & students	ESL students	350,000	academic
23) The Hong Kong University of Science & Technology (HKUST) learner corpus	Chinese - mostly Cantonese	written	Untimed assignments: EFL courses /take home exam	EFL university and advanced high school students	25 m	academic
24) The Indianapolis B Business Learner Corpus (IBLC)	various	written	business documents: letters, CVs, communication			

Table 2.2 English Learner Corpora Projects around the World (Continued)

Corpus	L1	Medium	Text type / task type	Proficiency level/ Learners	Size in words	Methodological Design: Objective
25) The International Corpus of Cross-linguistic Interlanguage (ICCI)	various	written				
26) The International Corpus Network of Asian Learners of English (ICNALE)	Chinese Indonesian Japanese Korean Malay etc.	written	short argumentative essays: control of topics, time, and dictionaries	EFL various	300,000 (estimated goal: 1 m)	academic
27) The International Corpus of Learner English (ICLE)	various	written	argumentative and literary essays	high-intermediate to advanced	3 m	academic & commercial
28) The International Teaching Assistants corpus (ITAcorp)	various	spoken	a variety of spoken classroom tasks: office hours role plays, presentations, discussions		500,000	academic
29) The ISLE speech corpus	German Italian	spoken	Recoded students' reading, using minimal pairs with multiple choice tests	Intermediate ESL		academic & commercial
30) The Israeli Learner Corpus of Written English	Hebrew	written	argumentative and descriptive essays	EFL	750,000	academic
31) The Japanese English as Foreign Language Learner (JEFLL) Corpus	Japanese	written	student essays	beginning to intermediate EFL	700,000	academic
32) The Janus Pannonius University (JPU) Corpus	Hungarian	written	essays and research papers	ESL university students	500,000	academic
33) Lancaster Corpus of Academic Written English (LANCAWE)	various	written	IELTS academic writing tests (descriptive and argumentative tasks); assignments.			longitudinal data: academic & commercial
34) The LeaP Corpus: Learning Prosody in a Foreign Language	German	spoken	4 types of speech style: 1.nonsense word lists 2. readings of a short story 3.retellings of the story 4.free speech in an interview situation	ESL various		academic

Table 2.2 English Learner Corpora Projects around the World (Continued)

Corpus	L1	Medium	Text type / task type	Proficiency level/ Learners	Size in words	Methodological Design: Objective
35) The Learner Corpus of English for Business Communication		written	Different types of business correspondence written		117,500	academic
36) The Learner Corpus of Essays and Reports		written	Essays and project reports: topics controlled		188,000	academic
37) A Learners' Corpus of Reading Texts	French	spoken	Unprepared reading of English texts: abstracts of fictions or dialogues	ESL		academic
38) The LONGDALE project: LONGitudinal Database of Learner English	various	written / spoken	Range of text types	intermediate to advanced		longitudinal data
39) The Longman Learners' Corpus	various	written	Essays and exam scripts	various	10 m	commercial
40) The Louvain International Database of Spoken English Interlanguage (LINDSEI)	various	spoken	Interviews and picture descriptions	high-intermediate to advanced	800,000	academic & commercial
41) The Malaysian Corpus of Learner English (MACLE)	Malay	written		EFL		academic
42) The Michigan Corpus of Academic Spoken English (MICASE)	Mainly L1 + data by native speakers	spoken	Transcripts of academic speech events		1.8 m	academic
43) The Michigan Corpus of Upper-level Student Papers (MICUSP)	semi-balanced sample of NS and NNS	written	ESP papers A-grade papers or ungraded papers assessed and accepted (such as research proposals), but not published		2.6 m	academic
44) The Montclair Electronic Language Database (MELD)	various	written	Student essays	various	100,000	academic
45) The Multimedia Adult ESL Learner Corpus (MAELC)		multi-media	Video of classroom interaction and associated written materials	beginning to upper-intermediate ESL		academic
46) The Neungyule Interlanguage Corpus of Korean Learners of English (NICKLE)	Korean	written / spoken	student essays / student interviews and oral speech tests transcriptions	beginning to intermediate EFL	890,000 / 100,000	academic

Table 2.2 English Learner Corpora Projects around the World (Continued)

Corpus	L1	Medium	Text type / task type	Proficiency level/ Learners	Size in words	Methodological Design: Objective
47) The NICT JLE (Japanese Learner English) Corpus	Japanese	spoken	English oral proficiency interview test	Various EFL	2 m	academic & commercial
48) The Łódź Polish English Learner Corpus (LPELC)	Polish	written	Argumentative, descriptive, narrative and quasi-academic essays; formal letters	beginning to post-advanced ESL	2.8 m	academic
		spoken			200,000	
49) The PICLE corpus (Polish component of ICLE)	Polish	written	student essays	Advanced ESL	330,000	academic
50) The Qatar learner corpus	Arabic	spoken	interviews with Qatari learners of English	EFL		academic
51) The Québec learner corpus	Arabic	written	argumentative essays	intermediate and advanced EFL	250,000	academic
52) The Santiago University Learner of English Corpus (SULEC)	Spanish	written	compositions or argumentative essays	ESL		academic
		spoken	semi-structured interviews, short oral presentations and brief story descriptions			
53) The Scientext English Learner Corpus	French	written	academic argumentative texts	ESL		academic
54) The Seoul National University Korean-speaking English Learner Corpus (SKELC)	Korean	written	student essays	various EFL	900,000	academic
55) The SILS Learner Corpus of English	various (mainly Japanese)	written	student essays	Basic, intermediate and advanced EFL		academic
56) The Soochow Colber Student Corpus (SCSC)	Chinese	written	student essays	EFL	227,000	academic
57) The Spoken and Written English Corpus of Chinese Learners (SWECCL)	Chinese	written	argumentative and narrative essays	EFL	2 m	longitudinal data: academic
		spoken	National Spoken English Test			
58) The Taiwanese Corpus of Learner English (TLCE)	Chinese	written	Journals and essays (descriptive, narrative, expository, argumentative)	intermediate to advanced EFL	2 m	

Table 2.2 English Learner Corpora Projects around the World (Continued)

Corpus	L1	Medium	Text type / task type	Proficiency level/ Learners	Size in words	Methodological Design: Objective
59) The TELEC Secondary Learner Corpus (TSLC)	Chinese	written		EFL	1.5 m	
60) The Telecollaborative Learner Corpus of English and German Telekorp	German	written	Computer-mediated interactions collected 6 different tele-collaborative parthnerships (2000-2005)	Bilingual (NS & NNS): ESL	1.5 m	longitudinal database: academic
61) The Tswana Learner English Corpus (TLEC)	Tswana	written	argumentative essays	advanced ESL	200,000	academic
62) The Uppsala Student English Corpus (USE)	Swedish	written	student essays	Various ESL	1.2 m	academic
63) The UPV Learner Corpus	Catalan	written	essays	Various ESL	150,000	academic
64) The Varieties of English for Specific Purposes dAtabase (VESPA) learner corpus	various	written	ESP texts (term papers, reports, MA thesis)	various		academic
65) The WriCLE (WrittenCorpus of LearnerEnglish) corpus	Spanish	written	essays	various ESL	750,000	academic
66) The ESF (European Science Foundation Second Language) Database *	various: Punjabi Italian Turkish Arabic Spanish Finnish	spoken	Spontaneous communication with native speakers	ESL 40 adults immigrant workers living in western Europe	various	academic
67) The MiLC Corpus*	Catalan	written	letters, summaries, CVs, essays, reports, translations, real-time and delayed communication exchanges,	ESL		academic
68) The Multilingual Learner Corpus (MLC) *	Brazilian Portuguese	written	argumentative and narrative essays	ESL		academic
69) The Padova Learner Corpus *	Italian	CMC (Computer-Mediated Communication)	FirstClass conferencing software: diaries, debate contributions, formal reports, résumés etc	ESL		longitudinal data: academic
70) The PAROLE corpus (corpus PARallèle Oral enLangue Etrangère) *	various	spoken	5 oral production tasks	various		

*projects (66-70) conducting studies of learner corpora from learners with multiple mother tongue backgrounds; the subjects are learners of multiple languages including English.

In accordance with the table, 70 projects of learner corpora of English have been investigated; however, a shorter note of learner corpora in written mode and additional report of learner corpora studies in writing are stated below.

2.2.2.1 Review of Learner Corpora in Written Texts

The review in this section has been completed based on the accessibility of the corpus centers and on the present study objective to explore learner language in written mode. In terms of accessibility, learner corpora objectives will be considered as the first criterion; later the review is narrow down into projects of learner written corpora.

In a view of accessibility, certainly the learner corpus projects with commercial objectives like the Cambridge Learner Corpus and the Longman Learners' Corpus do not allow outsiders to use the learner corpus database freely; a few projects are conducted with academic and commercial purposes such as the Louvain International Database of Spoken English Interlanguage (LINDSEI), the International Corpora of Learner English (ICLE), and the ISLE speech corpus offer a commercial channel to order CD containing learner corpus database online. Others are conducted with academic purposes and most of their research papers are searchable or downloadable online (e.g. Asao Kojiro's Learner Corpus Data, BELC-The Barcelona English Language Corpus, the EVA Corpus of Norwegian, and others). Some projects are limited to researchers in the projects and some provide accessibility for researchers allowed to use the learner corpora database after registering online such as the Br-ICLE corpus (Brazilian component of ICLE), the Giessen-Long Beach Chaplin Corpus (GLBCC), and the Chinese Learner English Corpus (CLEC). At this point, the present study obviously aims to explore learner

corpora in written mode; learner corpora studies with academic purposes will be reviewed next based on task types, language, and learners.

There are 53 projects of learner corpora of English include learner database in written mode; the projects conduct research by using different text and task types. Of the 53 projects, 7 projects combine learner corpora in written and spoken form. In addition, a variety of written task types provided in these projects is essay-oriented since 35 projects out of 53 define the task eliciting learner data in the forms of essays and compositions. The essays can be categorized into various styles: argumentative, descriptive, expository, narrative, and literary (Myles, 2003); however, the most common is argumentative (e.g. Abe, 2003; Arts and Granger, 1998; Agerström, 2000; Cobb, 2003; and Diez-Bedmar, 2009). Other tasks used in such project are stories written or reproduced (Asao, 1997); term papers: reports, research plans, abstracts, reviews, summaries, and thesis (Brezolin, 2008; Gabrielatos and McEnery, 2005). The other projects provide some different tasks when collecting learner data. They obtain the learner data from some business documents: letters, CVs, diaries (Flowerdew, 2000) EAP materials (Flowerdew, 2001), and translation (Granger, 2008; Spence, 1998; Uzar, 2002, 2003a, 2003b). However, such studies conducted in different English learning environments, whether ESL or EFL learning environments, might come up with results in another way.

Based on the information in the table, consideration of learner context, whether it is an ESL or EFL environment, should be highlighted as well. There are 52 projects that clearly reveal learners' language learning context; from this, it can be determined that the number of projects of ESL and EFL corpora are both 26. The others do not identify the learners' learning environment in total. However, of the 26

EFL corpora projects, some Asian learners of English corpora (e.g. Japan, Korea, Taiwan, China, and Malaysia) are actively implemented and publicized. None of the projects explore Thai EFL learners in particular. Some corpus-based studies of learner language have revealed some of main problems that non-native learners of English have when writing.

2.2.2.2 Learner Corpora Studies in Written Language

Corpus-based studies have shown many problems non-native learners of English encounter when writing: errors in *collocational patterning of words* and *phraseological inappropriateness* (Flowerdew, 2000; Giliquin et al, 2007), *pragmatic inappropriacy*, e.g. unsuitable use of modals or hedges (Flowerdew, 2000; Granger and Rayson, 1998; Hewings and Hewings, 2002; Neff et al, 2004); *semantic misuse* (Guliquin et al, 2007), *discourse features* such as misuse, underuse, overuse of connectors; tendency to put connectors in initial position of sentences; signaling noun use (de Cock, 2003; Flowerdew, 2001; Granger and Tyson, 1996), *register awareness deficiency* such as using some grammatical or lexical features of speech in writing, and underuse of some grammatical or lexical features of writing (Guliquin et al, 2007; Guliquin and Paquot, 2007; Granger and Rayson, 1998). However, among those grammatical, lexical, semantic, and discoursal errors, pronoun errors, a grammatical feature accepted as one of cohesive device in writing discourse (Johnstone, 2002) is not examined in particular. The present study, therefore, seeks to explore how Thai EFL learners produce pronouns in writings, what types of pronoun error occur when they write, and what language patterns of pronoun occur over time.

2.2.2.3 Conclusion

To put it briefly, an increasing number of studies of learner corpora have been investigated around the world in terms of the number of projects, number of words collected, and a growing number of L1 languages examined. Approximately a hundred projects still actively carry on their exploration of learner language in different English learning contexts, ESL or EFL environments. The studies of learner corpora have been set with different considerations or variables: learner levels (e.g. proficiency level and L2 context), language (e.g. medium, English or non-English), and task type (e.g. authentic or prepared, timed or untimed, and source free or source control). These truly help researchers in this field set the research design more systematically; the criteria will be comprehensively conferred for tracing the present study design of learner corpora.

2.2.3 Computer Learner Corpora Design

As stated earlier in section 2.1.1, Nesselhauf (2004:125; 2005:40) identifies learner corpora as 'systematic computerized collections of text produced by language learners'; Granger (2008) supports that the word '*systematic*' is essential for a strict design of learner corpora collection. The most accepted design criteria have been proposed by Tono (2003:800) with three main categories: language, task, and learners, as explained below.

Table 2.3 Criteria for a learner corpus design (Tono, 2003: 800)

Language-related	Task-related	Learner-related
Mode: [written/spoken]	*Method of collection*: [cross- sectional/longitudinal]	*Internal-cognitive*: [age/cognitive style]
Genre: [letter/diary/fiction/essay]	*Method of elicitation*: [spontaneous/prepared]	*Internal-affective*: [motivation/attitude]
Style: [narration/argumentation]	*Use of references*: [dictionary/source text]	*L1 background/L2* *environment*: [ESL/EFL]/[level of school]
Topic: [general/leisure/etc.]	*Time limitation*: [fixed/free/homework]	*L2 proficiency*: [standard test score]

The criteria consist of three main issues: language, task, and learners.

a) Considerations of language-related issues (mode/genre/style/topic)

To take language-related issues into consideration, a researcher has to decide which language of learner corpora he or she aims to explore. Domination of English is undeniably found among the designs of corpus-based; however, some, 35 out of 105 projects, research designs define non-English learner corpora as a target language; and only 5 projects describe their target language in the designs as multi-lingual. As a result, the target language can be English, non-English, and multi-languages. Next, a researcher has to consider mode or medium of the learner corpora provided it is writing, speech, or multi-media. Then genre eliciting the language medium will be considered.

It may be worth clarifying what is meant by genre first of all. Genre for most people is associated with the world of fiction writing, and categories such as thriller,

science fiction or gothic horror spring immediately to mind (Zaytseva, 2011). What most of our students need to produce when they need to function in English, however, are things like a letter to a prospective employer, a business report or the write-up of a scientific experiment. The key to the concept of genre is the 'purpose' the piece of writing serves. Genre in writing can be essays, diaries, fictions, articles, and business documents (e.g. correspondences, CVs, and other synchronic or asynchronic communications); while spoken genres tends to be recorded dialogues, conversations, interviews, read-aloud, storytelling or recalling, and other oral communication paths (Widdowson, 2000). However, for more precise interpretation and discussions, styles of genre need to be clarified.

Additionally, styles of genre, a characteristic of particular genre (Biber and Concord, 2009), have to be reviewed as one of variables for more valid and specific conclusions. For example, a written essay can be categorized as recount, narrative, information report, discussion, exposition, explanation and procedure (Zaytseva, 2011). Finally, the topic will be selected.

In short, language-related issues to be taken into considerations when designing learner corpus studies are mode, genre, style, and topic. These criteria are determined so that research instruments can be designed with high content validity (the instrument tests what the study aim to explore) (Cresswell, 2005).

b) Considerations of task-related issues (data collection/elicitation/use of reference/time limitation)

The data collection method normally relates to research questions; the possible ways generally occupied are *cross-section*, collecting data from different subjects at a

time, and *longitudinal*, collecting data from same subject at different time (Cresswell, 2005).

Another issue to be considered is data elicitation methods, whether it is in natural way or in a prepared setting. This truly depends of what data type a researcher seeks. The data collected in a natural way clearly does not require subjects' preparation before administering the research instruments such as writing outside classroom setting and writing during social network chatting; while the other requires subjects' preparation before collecting the data, examination or interview, for examples. Then, other criteria can categorize the data elicitation methods as well. First, permission to use reference textbooks or dictionaries has to be cautiously borne in mind if it influences the subjects' performance or not. If not, a researcher might allow textbooks or dictionaries, or identify dictionaries possibly to be used when the data is collected. Finally, time controlling usually implies a prepared setting; for example, a take-home assignment to write an essay for two week will give subjects some time to prepare, research, review, and revise before submitting their work. In this case, all references are freely employed.

In summary, task-related criteria directly involve research theoretical framework since they imply how the research is held, whether in cross-sectional or longitudinal paradigm. Besides, research instruments and how the instruments are managed are included.

b) *Considerations of learner-related issues (internal-cognitive/internal-affective/L1 background & L2 environment/L2 proficiency)*

The last criteria for building a learner corpus are about learners: internal-cognitive, internal affective, L1 background & L2 environment, and L2 proficiency

levels. These help a researcher identify subjects in the study. Firstly, learners'

internal-cognitive or learners' individual characteristics: age, aptitude, motivation &

attitude, personality, cognitive style, and learning strategies (Larsen-Freeman and

Long, 1991; Lightbrown and Spada, 1993). Effects of each individual difference has

to be considered since it influences L2 learning: for examples, older children can

learner more rapidly than younger children; with regards to morphology and syntax,

teenagers can learn best; adults can learn grammar faster than do children; but

children have better progress than adults where pronunciation is concerned (Ellis,

1994). In addition, other internal cognitive issues affect L2 learning in similar ways;

positive motivation & attitude, personality, style, and learning strategies tend to

support better L2 learning. Thus, learners' individual characteristics should be

clarified as factors in L2 learning.

Additionally, learners' L1 background or mother tongue has to be identified so

that comparisons between L1 and L2 can be made if needed. Moreover, L2

environment of learners can be defined as ESL, EFL, or EOL (Granger, 2002). ESL

and EFL are widely acknowledged; EOL is not. ESL (English as a Second Language)

is most often referred to when English is required in an English-speaking environment

(e.g. England, US, and Australia), often affecting immigrant populations. EFL

(English as a Foreign Language), means English that is learned in a classroom

environment in a non-English-speaking country (e.g. Japan, Korea, Thailand, etc.).

Finally, English as an Official Language (EOL) covers indigenized varieties of

English, such as Indian English and Nigerian English and used as an official language

of communication (such as affairs of government). This information will help

researchers depict how and why learners of English perform language in some particular ways.

Finally, Tono (2008) proposed learners' L2 proficiency as another criterion in terms of learner-related issues. The learners' L2 proficiency might be classified as beginner, pre-intermediate, intermediate, upper-intermediate, and advanced; however, classifying learners' L2 proficiency is not simple in reality. Learners often gather with various L2 proficiency levels. To categorize the learners according to their proficiency, therefore, a placement test is required. Although positioning learners' L2 proficiency might support more reliability of conclusions, generalizability might be questioned. In this case, Granger (2008) suggested an alternative way to use learners' institutional status so that the conclusion is more applicable; plus, no tests are needed.

All three main criteria have been proposed systematically by Tono (2003), as the most applicable criteria for building learner corpus research these days. The criteria are related to language, task, and learners; after all are set, research approach to process the data obtained has to be defined next.

2.2.4 Conclusion

The number of learner corpora studies has been increasing around the world in recent years due to the advantages of storing and processing larger learner databases associated with SLA perspectives to analyze and interpret the data. The learner corpus-based studies have been categorized by different ways related to language (modes, genre, styles, and topic); tasks (data collection methods, elicitation methods, use of reference, and time control); and learners (internal-cognitive, external affective, first language, and second language proficiency). To build up such studies, therefore, a researcher has to review and decide criteria of language, task, and learners. Then

the study approach to deal with the data elicited from learners will be highlighted after this section.

2.3 Computer-aided Error Analysis (CEA)

The present study aims to explore second language learners' IL errors using learner corpora; it mainly employs Computer-aided Error Analysis (CEA) approach to explore learners' corpus development. However, until now the results seem to be still incomplete in nature, but there is a growing body of research into specific areas of IL errors: for example, collocation (Granger, 1998b; Chen, 1998), connectors (Milton and Tsang, 1993; Granger and Tyson, 1996), and irregular past tense (Tono and Aoki 1998). Additionally, learner corpora studies rarely investigate interlanguage in different levels of learners, English pronoun in particular.

In this section, the background of the approach is reviewed, including general steps of CEA and error tagging systems. A few learner corpus studies employing the approach are reviewed in the last section.

2.3.1 Background of Computer-aided Error Analysis (CEA)

In accordance with the drawbacks of error analysis mentioned earlier in section 2.1.2, to collect a number of learner language data with more systematic error annotation might be a more effective approach to examine learner language (Granger, 2002). The present study applies the approach to explore English pronouns used in essay writing by Thai university English majors of different language levels attending English courses, mainly based on computer error analysis (CEA) as well as accuracy in focus and partially based on contrastive interlanguage analysis (CIA). In addition,

interlanguage will be explored for correlations, if any, among the subjects with different number of English courses the learners have taken in university level.

IL studies in learner corpora are still a very small number, though some studies have shown some meaningful information of learner corpora to describe IL characteristics at different developmental stages (Tono, 2000b). ICLE projects (de Haan 1997; Aarts and Granger 1998) have investigated characteristics of learner language in different stages by examining sequences of part of speech (POS) tags. The frequency of use of English articles was lowest among Finnish EFL learners (de Haan 1997); and this corresponds with the findings of Mazon and Uzar (2000) that Japanese and Polish EFL learners hardly used articles when communicating in English, since there is no article system in Japanese and only a minimal one in Polish. Aarts and Granger (1998) found similar tag sequence frequencies among Dutch, Finnish, and French EFL learners; based on POS tagging, nouns were underused and pronouns were overused in sentence-initial sequences. Other research results, such as Tono (2000b), showed very different sequence patterns compared to Aarts and Granger; modals and prepositional phrases were consistently underused by Japanese learners.

CEA analysis system, associated with Error Analysis approach, implemented in FreeText project and developed by Dagneaux et al (1998) comprises 5 steps as following:

1) Manual correcting of L2 French corpus

2) Elaborating error tagging system for L2 French

3) Inserting error tags and correction in the text files

4) Recovering specific error types and statistics list

5) Analyzing with linguistically concordance-based methods for error types

Similar to the present study, after POS tagging with the pronoun tagset designed the 5 steps for error tagging are processed. Firstly, manual correcting of the tagged pronouns produced by the Thai subjects will be processed. Next, the error tagset based on Thai learners' performance is systematically designed and detailed (see Chapter 3). Thirdly, the error tagset and correction will be applied into the text files. Then, specific error types and statistical list are administered. Lastly, both quantitative and qualitative data analysis will be completed with linguistically concordance-based methods for error analysis and discussions. The 5 steps will be processed with the tagset designed based on the errors possibly produced by Thai EFL learners; however, the error tagset is possibly adapted to pronoun errors performed by other learners of English according to suggestions for error tagging system provided by Granger (2003).

Additionally, to design more effective error tagging systems, the annotators have to consider the following four required characteristics suggested by Granger (2003, 2008):

1) Informative and manageable

2) Reusable for variety of languages

3) Flexible for addition or deletion of tags at annotation and post-annotation stages

4) Consistent error tagging principles between annotators

The error tagset in the study has been designed with these considerations. The error tagset symbols with corrections are informative enough to identify types of error, positions, and corrections of the errors. In addition, the error categories are clear

enough to direct. For the error tagset reusability, it is able to be reused in studies of Thai learners of English or other language learners of English. On top of that, the error tagset can be added or deleted at any stage; its flexibility is rather high. Finally, though the annotators attempted to design better error annotations for applicable annotation to a wide range of languages, there is no automatic error annotation at present (Granger 2008; Tono 2000b); reliability of error tagging is very vital. Thus, there are three annotators trained for reliable error tagging system; at least two thirds of the annotators' decisions are final.

In brief, CEA has been innovative approach, utilized most in Asian contexts, to reduce some weak points in the traditional EA; however, the fundamental idea of EA has been applied along with analysis processes. Applying CEA into learner language studies, researchers have to be cautious in designing the error tagset by considering its usefulness, reusability for a choice of languages, flexibility, and reliability of error tagging rules employed by annotators. To keep on these tracks, error tagging systems developed by some CLC researchers must be taken into considerations for designing particular tagsets of pronoun and pronoun errors for Thai EFL learners.

2.3.2 Error Tagging Systems

As mentioned earlier there is no completely automatic error annotation; annotators of many languages have to find out the most appropriate annotation systems for their studies. Advantageously, some error annotation systems designed in some previous studies can be adapted as they are flexible enough to add or delete some tags manually during tagging and after tagging processes (Granger 2003; Negrillo and Domínguez 2006). The error tagging system, therefore, must be kept in

mind as it truly helps tracing and arranging learners' errors; error tagging systems have been developed among CEA research reviewed below.

Firstly, error tagging system for English developed by Cambridge University Press employs 'a two-letter coding system', in which the first letter refers the general type of errors (e.g. incorrect forms, omission, and overuse); while the second letter refers to word class (Nichollos, 2003). For example,

Timmy <#UP>he<#UP> hits his brother.... (Nichollos, 2003: 573)

The letter 'U' refers to overuse and the letter 'P' refers to pronoun. However, the two-letter coding system is rather restricted since there are plenty of error types. Moreover, some word classes begin with the same letter, e.g. pronoun and preposition; thus the system might be too limited for extending.

Another initial error tagging system suggested by Granger (2003) and drawn for French as a Foreign Language is based on 'a three-tiered annotation system' consisting of error domain (form, morphology, grammar, lexis, and others), error category (number, tense, person, and others), and word category of the error (verb, noun, pronoun, and others). For example,

L'héritage du passé est très <G><GEN><ADJ> #fort$ forte </ADJ></GEN></G>

Error tagging systems normally includes the correction of the error detected. In the sentence above, the error detected after '#' is 'fort', with a grammatical error <G> of an adjective <ADJ> in genitive case <GEN>. The correction is provided right after '$' as 'forte'. The error tagging system is effective enough in terms of application and extension to further studies. However, some other taxonomies are possibly used in other CEA studies.

The other two major descriptive error taxonomies proposed by Dulay, Burt, and Krashen (1982) are taxonomy based on linguistic categories (e.g. general linguistic categories like morphology, lexis, and grammar; and more specific ones like nouns, articles, and pronouns) and taxonomy based on surface structures (e.g. omission, addition, malformation, and misordering). The error annotation system has been developed with alternative tools for computerized autoimmunization of error annotation such as *Université Catholique de Louvain Error Editor* (UCLEE) and TagEditor (Izumi et al, 2003). Some other studies of specific errors were not named, but generally referred to the name of university or counterpart project they worked for (e.g. Izumi and Isahara 2004; Tono 2000b). The most frequently mentioned as best representatives for error annotation systems are:

1) The Cambridge Learner Corpus project (CLC)

2) The FreeText project

3) The Université Catholique de Louvain

4) The National Institute of Information and Communications Technology Japanese Learner of English (NICT JLE)

The first one is commercially available since it provides tagsets for a variety of languages (Cambridge University Press, www, 2006). The second and third have been developed in the Centre of English Corpus Linguistics, Université Catholique de Louvain (http://cecl.fltr.ucl.ac. be). They have been the most repeatedly applied in some studies (e.g. Dagneaux et al 1998; Granger 1999; L'haire and Vandevender Faltin 2003). Finally, the last one, NICT JLE has been consistently applied in Japanese learner corpora studies (Izumi et al 2004, 2005; Tono 2004).

This section will discuss error annotation systems mainly based on dimensions of the error taxonomies and the structure of error taxonomies and error tags in the three well-known error annotation systems for academic purposes mentioned above.

a) Dimensions of Error Taxonomies and Structure of Error Tags

Normally, there are two dimensions of error taxonomies: error categories (linguistic levels) and error subcategories (error type) presented in one tag, for example:

[…] barons that (GVT) lived $had lived$ in those (FS) castels $castles$.

(Dagneaux et al, 1998: 16)

The first tag, (GVT), shows a grammatical (G) error of the word-class verb (V), and the grammatical category of tense (T) including correction of the error inside the symbol '$'. The second tag, (FS), shows form (F) error of spelling (S) followed by the correct form of the misspelling word.

Similarly, NICT JLE provides two linguistic errors at two levels: major categories (i.e. POS classification) and error categories (i.e. noun case, number of adjective, adverb inflection). The tagset pattern is POS classification, error category, correction of the error tagged, and it ends with the similar pattern right after the word annotated with the symbol '/' at the beginning of the pattern, for example:

I belong to two baseball <n_nnum crr= "teams">team</n_num>. (Izumi et al, 2005: 75)

However, the error annotation system of FreeText has three levels of annotation: domain (i.e. linguistic level of analysis), error category (i.e. homonymy, voice, prefab, word order), and word category (i.e. POS classification and sub-classification), for example:

[...]barons that <G><TPS><VSC> #lived$ had lived </G></TPS></VSC> [...]

(Adapted from Dagneaux et al 1998: 16; Granger 2003: 468)

The tagset consists of error domain of grammar, <G>; error category of tense, <TPS>; and word category of verb, <VSC>. It is also inserted with correction of the error tagged preceded with the symbol '#' and ended with the symbol '$'. The similar tagset with the symbol '/' in front of each annotation level is provided at the end of the error found.

The tags by FreeText and NICT JLE are based on XML (extensible markup language) syntax. The sequence of opening tags and the correction are inserted in front of the erroneous data, and the closing XML comes right after the error. The strength of using XML is that it can clearly identify the structure of the text and it is also very beneficial when corpus data is utilized for web-based pedagogical tools or databases as a hypertext (Izumi et al, 2005).

2.3.1.1 Other Approaches of Error Tagging

An error annotation approach known as *MELD* has been proposed by Fitzpatrick and Seegmiller (cf. Negrillo and Domínguez, 2006). The system depends on reconstruction of the error rather than on description of errors by coding; as the annotators themselves stated that a list of errors "[...] limits the errors recognized to those in the tagset [and] introduces the possibility that annotators will misclassify those errors that do not fit neatly into one of the tags on the list" (Negrillo and Domínguez 2006: 97). Over classifying reconstruction is considered strength of the study somehow as it was not time consuming and its assist of tagging and parsing learner corpora. However, the approach has been criticized as questionable in terms of reliability and consistency of error annotation.

According to Granger (2008) and Tono (2006), error annotation is very time and budget consuming; therefore the present study can focus only on a part of learners' written language, English pronouns, and based on the limits of time, labor and financial support. The error annotation system has to be informative and manageable, reusable for many languages, flexible for adaptation, and consistent between annotators. XML format will be applied throughout to be useful in the future for all corpora and qualitative studies (Lee, 2008), because of its strength of comprehensible format. Additionally, the study will employ a partial adaptation from the FreeText error annotation system according to its linguistic taxonomies of grammatical and discoursal errors. In addition, some structural features of error taxonomies and error tags are based on Tono (2006) since Japanese pronoun error types are quite similar to Thai ones.

2.4 English and Thai Pronouns

According to Jacobs (1995, p. 124), English pronominals can cause serious communication problems for ESL/EFL learners. Because pronominals are often ambiguous, especially in utterances considered separately from their context, non-native speakers need to draw for their interpretation on grammatical principles and pragmatic knowledge to derive meaning. Inaccurate use of pronouns in speech or writing poses problems in syntax which if excessively present will also become obstacles to effective communication. Pronoun, therefore, is the main language feature to be examined.

To explore pronouns produced by L2 learners, pronouns in the learners' L1, Thai, and L2, English, are reviewed and compared for better comprehensive understanding of the language differences and similarities.

Before moving to review of pronouns in both languages, definitions of pronoun are stated here.

Common Definitions of Pronoun

The traditional definition of a pronoun is a word taking the place of a noun or noun phrase; modern grammarian regarding its position and function as the decisive factors in classifying a part of speech usually consider pronoun as a subclass of noun (Frank, 1972; Wales, 1996).

Additionally, other reference books such as:

- The American Heritage Dictionary of English (2009) defines pronouns as "the part of speech that substitutes for nouns or noun phrases and designates persons or things asked for, previously specified, or understood from the context";

- Collins English Dictionary (2003) defines pronoun as "one of a class of words that serves to replace a noun phrase that has already been or is about to be mentioned in the sentence or context"; and

- Merriam-Webster (2011) also states a pronoun is "any of a small set of words in a language that are used as substitutes for nouns or noun phrases and whose referents are named or understood in the context."

In short, a pronoun can be defined as a small unit of words in a language which is able to take the place of a noun or a noun phrase likely to be stated or understood previously in the context.

Information of English and Thai pronouns will be presented in the next sections. Forms and rules governing the pronouns in each language will be discussed.

2.4.1 English Pronouns

The following information of English pronoun includes its paradigm and usage. The later one will be presented in terms of forms and function, and cognitive process.

2.4.1.1 Pronoun Paradigm in English

In this study, we are interested in the acquisition of those pronouns which indicate the notion of person. They make up the central class of pronouns in English. Table 2.4 is an inventory of the pronouns of Modern English (Borjars and Burridge, 2001).

Pronouns are words that replace a noun or noun phrase, but they commonly refer to persons and things, Kolln (1991: 331) has said, when the word pronoun comes to mind, we generally label them on the basis of person and members. There are three case forms of personal pronoun to indicate different sentence function: Subjective case, objective case and possessive case. Personal pronouns change their form for person (First, second and third), for case (subject, object, possessive), number (singular, plural) and gender (masculine, feminine, neuter), except for reflexive pronoun making the same kind of changes.

Table 2.4 Pronouns of modern English

	Personal Pronoun		Possessive Pronoun
	Subject Case	Object case	
1st person			
Singular:	I	me	mine
Plural:	We	us	ours
2nd person			
Singular:	You	you	your
Plural:	You	you	your
3rd person			
Singular:			
- Feminine	She	her	hers
- Masculine	He	him	his
- Neuter	It	It	its
Plural:	They	them	theirs

The forms of pronoun in English are systematic according to the control of person, number, gender, neuter, and case. However, usage of pronouns in English is rather complicated as presented in the next section.

2.4.1.2 Usage of Pronouns in English

A. Forms and Functions

a) Subject Pronouns

According to Teresa (1991), a pronoun in the subject group (I, we, you, they, he, she, and it) may be used in two ways:

1) The subject of a verb

He is my brother. (*He* is the subject of the verb is)

We girls gave a party. (*We* is the subject of the verb gave)

He is taller than I.

Table 2.4 Pronouns of modern English

	Personal Pronoun		Possessive Pronoun
	Subject Case	Object case	
1st person			
Singular:	I	me	mine
Plural:	We	us	ours
2nd person			
Singular:	You	you	your
Plural:	You	you	your
3rd person			
Singular:			
- Feminine	She	her	hers
- Masculine	He	him	his
- Neuter	It	It	its
Plural:	They	them	theirs

The forms of pronoun in English are systematic according to the control of person, number, gender, neuter, and case. However, usage of pronouns in English is rather complicated as presented in the next section.

2.4.1.2 Usage of Pronouns in English

A. Forms and Functions

a) Subject Pronouns

According to Teresa (1991), a pronoun in the subject group (I, we, you, they, he, she, and it) may be used in two ways:

1) The subject of a verb

He is my brother. (*He* is the subject of the verb is)

We girls gave a party. (*We* is the subject of the verb gave)

He is taller than I.

The sentence is not written out in full, it means He is taller than I am (I is the subject of the verb am).

She plays as well as he.

(In full, she plays as well as he does. *He* is the subject of the verb does).

2) A word that means the same as the subject.

That boy in the blue jeans is *he* (*He* is the same as the subject boy).

It was *she* all right

(*she* is the same as the subject *it*. Therefore, the pronoun from of the subject group is used).

Personal pronouns appearing after the verb be (usually for identification of a person) take the nominative case in formal English (*it is I, it is we*), although many native speakers are not comfortable with this form and modern spoken English allows some exception for that rule: *it is me* and *it is us* are widely used particularly in informal speech.

b) Object pronoun

Object pronouns appear after verbs or after prepositions with a certain type of verb- preposition combination. In such phrasal verbs, a object pronoun appears between the verb and preposition, for example *call him up*. Object pronouns or pronouns in the non-subject group (me, us, you, them, him, her, and it) can be used in four ways as:

1) Direct object of verb:

They invited *me* (*me* is the direct object of verb *invited*)

2) Indirect object of verb

They gave her a book.

(*her* is the indirect object of the verb *gave*. It comes before a direct object).

3) Object of preposition

They taught it to *him* (*him* is the object of preposition)

4) Subject of infinitive

They asked *her* to move (*her* is the subject of infinitive).

In a series of two (or more) subjects or objects, the pronoun *I* comes last for the sake of politeness, for example:

My brother and *I* go to the movie.

In American English, two personal pronouns do not usually occur together as an indirect and direct object combination, for this reason, the sentence. *I gave it to him* would be preferred to. *I gave him it* (but, *I gave him some, I gave him that*).

c) *Possessive Pronouns*

Possessive pronouns are not followed immediately by a noun and they often stand alone. They show possession in the same way as the similar possessive determiner (*my, our, your, their, his, her,* and *its*). The possessive pronouns are *mine, yours, theirs, his, hers, ours,* and *its.*

This book is *mine.*

(*mine* is the possessive pronoun, indicating possession of the subject *book*).

B. Pronoun principles

To keep cohesion in writing, a writer needs to realize the appropriateness of particular forms of reference by appealing to the cognitive status of the referent in the mental representations of the discourse participants (Wales, 1996).

For the most part, a pronoun is used when the writer believes that the referent is already available in the readers' consciousness (i.e., when the entity is "given"). However, many things are possibly in the readers' consciousness, and the writer needs to use a form that will let the readers easily decide the correct one. At any one moment, discourse referents differ in terms of their salience in the mental representations of discourse participants, and that less-specified forms are only used when the referent is sufficiently salient. It is worth to review how pronoun as one of reference is expressed and what control the use of native speakers of English.

Some principles concerning cognitive process of how native speakers express pronominal to keep readers' comprehension are also important, for instances the most acceptable discourse hypotheses are Gradation of Referring Expression (Givons, 1980) and Accessibility Hierarchy (Areil, 1990). The others e.g. Chafe (1994) and Gundel et al. (1993) are similar to Ariel's Accessibility Hierarchy, but are not as comprehensive.

Givon (1980) proposed a list of referents used in communication based on the degree of continuous or accessible topic as follows.

For the most part, a pronoun is used when the writer believes that the referent is already available in the readers' consciousness (i.e., when the entity is "given"). However, many things are possibly in the readers' consciousness, and the writer needs to use a form that will let the readers easily decide the correct one. At any one moment, discourse referents differ in terms of their salience in the mental representations of discourse participants, and that less-specified forms are only used when the referent is sufficiently salient. It is worth to review how pronoun as one of reference is expressed and what control the use of native speakers of English.

Some principles concerning cognitive process of how native speakers express pronominal to keep readers' comprehension are also important, for instances the most acceptable discourse hypotheses are Gradation of Referring Expression (Givons, 1980) and Accessibility Hierarchy (Areil, 1990). The others e.g. Chafe (1994) and Gundel et al. (1993) are similar to Ariel's Accessibility Hierarchy, but are not as comprehensive.

Givon (1980) proposed a list of referents used in communication based on the degree of continuous or accessible topic as follows.

Low topic continuity/accessibility

(1) Referential indefinite NPs

(2) Cleft/focus constructions

(3) Y-moved NPs (contrastive topicalization)

(4) L-dislocated Def-NPs

(5) Neutral-ordered DEF-NPs

(6) R-dislocated DEF-NPs

(7) Stressed/independent pronouns

(8) Unstressed/bound   pronouns   or   grammatical
agreement

(9) Zero anaphora

High topic continuity/accessibility

Figure 2.1: Givón's Gradation of Referring Expressions

In a collection of cross-linguistic text analyses, Givón et al (1983) correlated the forms of reference in 10 with three measures of topicality:

a) referential distance (how recently the entity has been mentioned);

b) potential interference (how many other potential antecedents of the referring form there are); and

c) persistence (how long the entity will remain in the discourse).

By using both referential distance and persistence, Givón included two features in his conception of topicality. First, topicality reflects the status of the referent according to the discourse thus far. Second, the way in which a speaker refers to an entity reflects the speaker's intentions about the role of that entity in the remainder of the discourse.

Givón's measure of "potential interference" addresses the issue of the ambiguity of the referring form, relative to the discourse situation. One problem with

this measurement, however, is defining exactly what constitutes possible interference. Givón's description is somewhat unclear:

"An interfering topic was counted only if it was just as semantically compatible (most commonly in terms of animacy, humanity, agentivity or semantic plausibility as object or subject) with the predicate of the clause of the topic under consideration."

(Givón ,1983:14)

This characterization, even though free, reflects the importance of constraining information from the predicate as a whole, and not just the anaphor itself. For examples, if one reads

(a) The butterfly$_i$ saw the cat$_j$ before it$_i$ flew away.

(b) The butterfly$_i$ saw the cat$_j$ before it$_j$ pounced.

The interpretation of "it" in sentence (a) is assisted by the compatibility of "the butterfly" and "flew". On the other hand, in the sentence (b) the compatibility between "pounce" and "cat" would force a different interpretation of "it".

Even though Givón's measures of topicality (referential distance, potential interference, and persistence) are too rough to accurately reveal the processes of language comprehension and production.

However, by counting "interference" as one of the measures of topicality, Givón seemingly has stepped beyond what was originally intended by the term "topic". The topicality of an entity, even if it is a continuous notion, seems to be a characteristic that exists in the role an entity plays in a discourse. On the contrary, interference from other discourse entities is only relevant as long as it may hold back the interpretation of referring forms. Accordingly, Givón's idea of "topicality" has

more in common with other scales of salience or accessibility than with a traditional conception of "topic".

Givón's topicality scale differs from other accounts of reference form in that his measures of topicality concern the referring expression itself, as opposed to the cognitive status of the referent. Other scholars, such as Ariel, have been concerned with the topicality of the referent, while Givón's measures are meant to indicate the topicality of the referring expression. Nevertheless, Givón assumes that the text properties are associated with the cognitive status of entities, such that "What is continuing is more predictable", and "What is predictable is easier to process" (1983a:12). Based on Givón's, to a first approximation, the three measures of topicality can be interpreted as indices of the cognitive status of the conceptual referent, and in that sense are comparable with other approaches that identify degrees of topicality with the referent and not with the referring expression.

The other is 'The Accessibility Hierarchy' proposed by Areil (e.g. 1988; 1990; 1994), one of the most comprehensive in proposals of referent; the hypothesis is employed for discussing the results of the present study. Areil suggested different distribution patterns for different forms of reference, which she termed "accessibility markers"

Several researchers have suggested that the cognitive status of referents can be characterized in terms of a graded scale. One of the most comprehensive proposals is Ariel's Accessibility Hierarchy (e.g., 1988, 1990 and 1994). Importantly, Ariel (1990) suggested that the accessibility of a referent entity is determined by multiple factors. She proposed that the four most important are those listed below.

Factors affecting the Accessibility status of an antecedent:

a) Distance: The distance between the antecedent and the anaphor (relevant to subsequent mentions only)

b) Competition: The number of competitors on the role of antecedent.

c) Saliency: The antecedent being a salient referent, mainly whether it is a topic or a non-topic.

d) Unity: The antecedent being within vs. without the same frame/ world/point of view/segment or paragraph as the anaphor.

(reproduced from Ariel, 1990:28)

Ariel's third factor, "Saliency", addresses the difference between topical and nontopical antecedents. Her discussion reflects an assumption that "topic" is defined in terms of the grammatical subject. She also mentioned Levy's (1982) claim that topicality is influenced by the number of anaphoric references to an entity, in particular pronominal references. Although the Accessibility hierarchy is inherently a graded scale, it appears that "topic" is treated as an all-or-nothing phenomenon, implying that an entity either is the topic or not, and that a given discourse segment has one and only one topic.

The fourth area that Ariel lists is that of "Unity". This factor reflects the effect that discourse structure can have on "Working Memory", and thus reference form. Ariel suggests that choices in reference form are influenced by the discourse structure, which can be influenced by things like the passage of time within the discourse or paragraph breaks in written text. She links this factor to Fox's (1987) claim that in English, 'by using a pronoun the speaker displays an understanding that the preceding sequence has not been closed down' (1987:18). Thus, pronouns are more natural for

references to things from the same discourse segment, and fuller forms are used when the referent was last mentioned in a different segment.

Through text analysis in Hebrew and English, Ariel (1988, 1990) demonstrated different distribution patterns for different forms of reference, which she termed "accessibility markers". Her full Accessibility Marking Scale is reproduced in 12, with examples of English accessibility markers.

Ariel's Accessibility Marking Scale (1990:73), with examples (1988:84)

Marking Scale	Examples
Full name + modifier	Joan Smith, the president
Full ('namy') name	Joan Smith
Long definite description	The tall and authoritative president
Short definite description	The president
Last name	Smith
First name	Joan
Distal demonstrative + modifier	that hat we bought last year
Proximal demonstrative + modifier	this hat we bought last year
Distal demonstrative + NP	that hat
Proximate demonstrative +NP	this hat
Distal demonstrative	that
Proximate demonstrative	this
Stressed pronoun + gesture	SHE (plus gesture)
Stressed pronoun	SHE
Unstressed pronoun	she
Cliticized pronoun	(no examples in English)
Extremely High Accessibility Markers	gaps, including pro, PRO and
wh - traces, reflexives, and Agreement	

According to the scale above, Areil showed that zero anaphora and unstressed pronouns co-occur with high accessibility of referents, while stressed pronouns and full name with modifier signal low accessibility of referents. This co-occurrence can easily be understood in terms of cognitive process of activation.

Other scholars have proposed similar scales of accessibility, such as Chafe (1994) (given > accessible > new), or Gundel et al. (1993) (in focus > activated > familiar > uniquely identifiable > referential > type identifiable). These scales are similar to Ariel's Accessibility Hierarchy, but are not as comprehensive.

In the present study, pronoun use of Thai EFL learners will be explained what influenced a pronominal choice in a pronoun form in contexts. When the student writers express pronominal choice with low accessibility degree, with cognitive process it seems that they realize the readers might not understand the referent well. Plus, when they express or omit personal pronouns in their writings it seems that the pronominal choices might be easily understood by the readers.

2.4.2 Thai Pronouns

Thai pronouns here will be discussed in terms of forms and usage. Only those forms and usage of personal and possessive pronouns directly involved the study will be emphasized.

2.4.2.1 Thai Pronoun Forms

Although Thai personal pronouns and possessive pronouns explained here exclude personal names, such words as occupational titles ('Teacher'), and kin terms ('older sister', 'mother') function as pronouns; if expanded to include these non-pronominal parts of speech, Thai has a staggering number of words used as pronouns used in different situations and social parameters.

Thai is very different from most other languages in terms of the confounding number of pronouns that are available and used in everyday speech. With over a dozen words that can describe a first person singular and a similar number for second person plural or singular, knowing which one to use and when can seem like a daunting task (Campbell & Shaweevongse, 1957; Noss, 1964) proposes that Thai pronouns include the parameter of 'Situation', or social context, together with person, gender, and plurality/singularity. Palakornkul (1972) highlights the complex social factors that help govern pronoun choice in Thai including "situation" as described by Noss. The choice of which one to use depends on many socio-cultural factors – who you are talking to, how well you know them, how old you are relative to them and, most importantly, the relationship between the interlocutors and who is of 'higher status' (Iwasaki and Ingkaphirom, 2005).

While some pronouns like 'chăn' for 'I' can be, classified as 1st person pronouns (PP1), there is no very clear distinction between 2nd and 3rd personal pronouns (PP2 and PP3) in the word 'thəə,' often translated as 'you' and 'he' or 'she'. Moreover, some Thai pronouns have no gender. 'He' and 'she' can be problematic because the most common term equivalent Thai term – kháw —has no gender and can be used to refer to a male or female. This is a choice that has to be made countless times a day in every conversation, which is effortless for Thais.

Cooke (1968) lists 27 first-personal pronouns, 22 second-person pronouns, and 8 third-person pronouns including some specialized terms like 'àattàmaa' (อาตมา) used as a first-person pronoun by a Buddhist monk when speaking to non-intimate layman or lower ranking monks and some borrowed terms like 'úa' (อั๊ว), a first-person

pronoun for 'I' borrowed from Chinese. In addition, he classified other forms which are possibly used pronominally to refer to addresser, addressee, or referent into three types: *personal pronouns, kin-type nouns*, and *name nouns*. Palakornkul (1972) classified Thai pronouns into eleven groups; however, the groups relatively fall into Cooke's categories. Besides, Hatton (1978) further categorized pronominal elements which can be used to refer to the speaker: *personal pronoun proper, names, titles, kinship terms,* and *zero*. However, Iwasaki and Ingkaphirom (2005) classify the three personal pronouns with 9 first-person pronouns, 8 second-person pronouns, and 5 third-person pronouns according to Thai common use and consideration of formality and gender similar to the tables below. These include reciprocal pronouns ranked by formality level.

SPEAKER	Male	Male/Female	Female
Higher formality	/kraphŏm/ (กระผม)	/khâaphachâw/(ข้าพเจ้า)	
			/dìchán/ (ดิฉัน)
	/phŏm/ (ผม)		
		/chán/ (ฉัน)	
		/raw/ (เรา)	
			/kháw/ (เขา)*
			/tua een/ (ตัวเอง)*
Lower formality		/kuu/ (กู)	

(Adapted from Iwasaki and Ingkaphirom, 2005: pp.50)

Figure 2.2 Thai First-Person Pronouns

SPEAKER =	Male	Male/Female	Female
Higher formality		/thân/ (ท่าน)*	
		/khun/ (คุณ)	
		/thəə/ (เธอ)*	
		/naay/ (นาย)	
		/raw/ (เรา)*	
			/tua/ (ตัว)*
			/tua eeŋ/ (ตัวเอง)*
		/kɛɛ/ (แก)*	
Lower formality		/muŋ/ (มึง)	
		(Adapted from Iwasaki and Ingkaphirom, 2005: pp.51)	

Figure 2.3 Thai Second-Person Pronouns

SPEAKER =	Male	Male/Female	Female
Higher formality		/thân/ (ท่าน)*	
			/thəə/ (เธอ)*
		/kɛɛ/ (แก)*	
		/kháw/ (เขา)*	
		/tua eeŋ/ (ตัวเอง)*	
Lower formality		/man/ (มัน)	
		(From Iwasaki and Ingkaphirom, 2005: pp.52)	

Figure 2.4 Thai Third-Person Pronouns

Interestingly, some important forms of Thai personal pronouns can function in both as first- and second-; second- and third-; and third- and first-person pronouns. For instances,

	Speaker	Addressee
Higher formality	/khâaphachâw/ (ข้าพเจ้า)	/thân/ (ท่าน)
	/kraphŏm/ (กระผม)	/khun/ (คุณ)
	/dìchán/ (ดิฉัน)	
	/phŏm/ (ผม)	
	/chán/ (ฉัน)	/thəə/ (เธอ)
	/raw/ (เรา)	/naay/ (นาย)
	/kháw/ (เค้า)	
		/tua/ (ตัว), /tua eeŋ/ (ตัวเอง)
	/tua eeŋ/ (ตัวเอง)	
Lower formality	/kuu/ (กู)	/muŋ/ (มึง)

(From Iwasaki and Ingkaphirom, 2005: pp.54)

Figure 2.5 Thai personal pronouns as speakers & addressees in different formality

Additionally, the singular pronouns above can be pluralized by adding 'phùak', or a group, in front of a pronoun. For instances, 'phùak raw' means we; 'phùak kháw' means they (people); and 'phùak man' means they (animals or things). However, the word is not regularly added for pluralizing, especially in spoken

language which usually it is omissible when the pluralized meaning has been shared between the speaker and the hearer.

Also, a prefix 'kɔ̌ɔŋ', or to belong, will be added to personal pronouns for showing possession, for examples 'kɔ̌ɔŋ dìchán', 'kɔ̌ɔŋ raw', 'kɔ̌ɔŋ thəə' and 'kɔ̌ɔŋ kuu'. Remarkably, these possessive words without a preceding noun function as same as possessive pronouns in English, for example:

นี้ เป็น กระเป๋า ของ ฉัน และ นั่น ของ เธอ

ni pen krapáw *kɔ̌ɔŋ chán* lɛ́ɛ nán *kɔ̌ɔŋ thəə*

'this is bag *possession-I* and that *possession-you*'

As seen in the example above, possession in Thai can be either preceded by a noun phrase or omissible noun phrase since the noun phrase is known between interlocutors.

Briefly, English pronouns and Thai pronoun forms are undoubtedly different. English pronoun forms vary due to person, case, gender, and number. Forms of personal and possessive pronouns in English, therefore, are fixed and limited due to subject case (I, we, you, they, he, she, and it); object case (me, us, you, them, him, her, and it); and possession (mine, ours, yours, theirs, his, hers, and its). Excluding the enormous number of pronominal terms, Thai personal pronouns are formed in alternative ways due to person, social status, formality and relationships between the interlocutors. As a result, several forms of first, second and third personal pronouns are generally found in Thai; plus, some forms can be used as first, second, or third persons depending on the speaker's attention and contexts. In addition, the word

'phùak' can be added in front of a pronominal term to show pluralization; the word 'kɔ̌ɔŋ' is added before a pronominal term to show possession in Thai. However, the pronominal can be omitted when it is old information; it functions similar to possessive pronouns in English though without inflection of forms.

As mentioned earlier, the rule governing English pronouns is rather fixed, as per the principle B in Government Binding Theory proposed by Chomsky. However, Thai pronouns are much more flexible and the information below will explain how they are ruled when compared to English ones.

2.4.2.2 Thai Pronoun Usage

Thai personal pronoun is under the Principle B of the Binding theory; however, Thai topic including pronoun is controlled by some other different rules allowing resumptive subject, dummy subject, and pro-drop. These differences might be related to the Thai EFL writings in English.

1) Principle B of the Binding Theory

In accordance with the Principle B of the Binding theory, which says a pronoun must be free in its governing category; Thai pronouns are under the same umbrella, for instance:

(a) Malee$_i$ knew [Wannee$_j$ hated *her*$_{i/*j/k}$].

(b) Malee$_i$ told [Wannee$_j$ to buy *her*$_{i/*j/k}$ the book].

In both (a) and (b), the object pronoun '*her*' can refer to 'Malee' or another person (female) outside the minimal clause, but not to 'Wannee' which is in the same minimal clause.

If Thai EFL learners follow only the rule, they should not have any problems when using English pronoun. However, the learners might be influenced by

some other rules, which will be discussed below, governing Thai subject in contexts; they sometimes produce incorrect or inappropriate pronouns, especially in discoursal aspect (Iwasaki and Ingkaphirom, 2005).

2) Resumptive pronoun (Shadow pronoun)

Resumptive pronoun or shadow pronoun is frequently used in informal speech in Thai. The pronoun refers back to its full NP appearing immediately before it within the same clause (Iwasaki and Ingkaphirom, 2005). Generally, the pronoun appears after an NP in subject or topic position and it is regularly mentioned as /kháw/, /kɛɛ/, and /man/, for examples:

(c) พ่อเด็กคนนี้เขากระดูกหักแล้ว

ph3 dék khon ni *kháw* kradùk hàk lɛɛw

'The father of this child, *he* had broken bones.'

(Iwasaki and Ingkaphirom, 2005: 369)

Resumptive pronoun is used for giving importance to the NP referent. The listener or reader will be aware of the weigh being provided to the NP. Although English allows the shadow pronoun in the same way, some Thai EFL learners produce such sentence without a comma between the NP referent and the pronoun, '*The father of this child he had broken bones,*' which considered ungrammatical mainly in academic writing.

3) Dummy subject

Thai uses dummy subject for evaluative expressions, and /man/ is commonly used such as /man dii/ or 'it's good'. In the excerpt below, the writer

explained that her parents do not criticize her for her involvement with her boyfriend, but she must finish her study. Otherwise, it (/man/) would ruin everything. The NP referent of /man/ in the sentence (e) is nonrepresentational if it exists at all.

(d) ถ้าเราอยากจะมีแฟน พ่อแม่ไม่ได้ห้าม

tha raw yak cà mii fɛɛn phɔ́ɔ mɛɛ mǎy dai hâam

If we want to have a boyfriend

(e) แต่ต้องเรียนให้จบก่อน มันจะทำให้เสียหมดได้

tɛɛ tɔ̂ŋ lian hǎy còp kɔ̀n *man* cà sǐa mòt dai

But we have to finish study first *it* will ruin everything.

(adapted from Iwasaki and Ingkaphirom, 2005: 373)

In the examples, the NP referent, having a boyfriend, for the pronoun '*it*' is not clear-cut since there is new information in between before the writer mentioned the pronoun. If Thai EFL learners produce such sentences in English, they undoubtedly produce unclear sentences and cannot communicate effectively at first place.

4) Pro-drop

Although Thai has large sets of words that can function as first, second and third person pronouns, it generally follows SVO constituents in clauses and sentences, and the constituent order is much more flexible in real discourse. In other words, overt subject or topic might not be used in every sentence (Hoonchamlong, 1991). In a binary interaction, Campbell (1969: 23) stated that personal pronouns are omissible on condition that the speaker and addressee are aware of the 'actors' and 'goals' involved in the discourse. Remarkably, Hatton (1978) proposed that

pronominal forms are used only when the speaker considers himself as "new information". Once the speaker identifies himself, the pronominal forms are possibly omitted.

In addition, Thai is not real null-subject; according to Chomsky (1995), null subject or pro-drop sentence will be followed by an inflectional agreement of a verb based on the missing subject so that the reader or listener can interpret who or what the subject is. Examples of possibly problematic constructions with English equivalent versions are following:

(f) ไม่สบาย โรงเรียนก็ปิด ไม่ได้มาเรียน รู้สึกเสียเวลาไปเป็นอาทิตย์

măy sabăy loŋlian kɔ̌ pít măy dăy ma lian lúsᵾk sĭa welaa pai pen athít

was sick. The school was also closed. (I) think we wasted time for a week.

(adapted from Iwasaki and Ingkaphirom, 2005: 375)

(g) ถ้าต้องการมาเที่ยวอุบล อาจมาทางรถไฟ เครื่องบิน หรือรถทัวร์

tha toŋkâan ma tiaw u bon âat ma tâaŋ rotfai krᵾŋbin lᵾ̌ rottɔ̌ɔ

If (one) wants to travel in Ubon, (s/he) may come by train, plane, or bus.

There is no inflection of words in Thai; it is rather difficult to interpret the sentences with pro-drop usage. Pro-drop is prohibited in English; it is considered ungrammatical and incomprehensible both in communicative and academic levels.

Some of the rules governing Thai pronouns (principle B) are similar to English, while others (resumptive subject, dummy subject and pro-drop) are very

distinct from English. These rules possibly influence Thai EFL learners' pronoun errors when writing English. The annotation of pronoun errors in the present study is designed partially in accordance with the possibly that errors have been influenced by Thai rules.

2.4.3 Predetermination of Pronoun Errors by Thai EFL learners

According to the discussions of Thai pronouns above, the researcher has predetermined the categories of pronoun errors into six main error types: omission, overuse, case error, gender error, number error, and mis-coreference (both within a sentence and across sentences). These errors can possibly change any time due to the data obtained. Some examples of each error type are given below.

Since every pronoun must have a clear reference; in other words, every pronoun must refer to an unquestionable antecedent. An antecedent is the noun phrase to which the pronoun refers. However, pronoun errors and unclear pronoun reference, no matter they are ambiguous, general, weak, or indefinite reference, found in writings by Thai EFL learners might be categorized under the following six error types.

Error Type 01: Pronoun omission occurs as the writer leave out pronoun in any positions which makes ungrammatical sentences. For examples:

1) We did not attend the meeting because *[] visited the new branch last week.

Possible correction:

We did not attend the meeting because *we* visited the new branch last week.

2) Sally cried whenever her mother hit *[].

Possible correction:

Sally cried whenever her mother hit *her*.

3) In the forest *[] has a small cottage.

Possible correction:

In the forest, *she* has a small cottage.

Error Type 02: Pronoun overuse occurs as the writer provides unnecessary subjects or objects, which makes implausible sentences. For examples:

1) My cat *it likes dog instant food.

Possible correction:

My cat likes dog instant food.

2) A girl *she went to the cottage *it was in the forest.

Possible correction:

A girl went to *the cottage* in the forest.

3) The step sister told *her Cinderella to clean up the bedroom.

Possible correction:

The step sister told *her* to clean up the bedroom.

Error Type 03: Pronoun case error occurs as the writer misuses the pronoun form in subject or object cases. For examples:

1) *Me want to visit my grandmother.

Possible correction:

I want to visit my grandmother.

2) I do not like *she because she always complains.

Possible correction:

I do not like *her* because she always complains.

3) Sally provides some refreshment for *they whenever the kids back home.

Possible correction:

Sally provides some refreshment for *them* whenever the kids back home.

Error Type 04: Pronoun gender error occurs as the writer misuses the pronoun gender referring to its antecedent. For examples:

1) Jack hit his daughter$_i$ because *he$_j$ messed up his documents.

Possible correction:

Jack hit his daughter$_i$ because *she$_j$* messed up his documents.

2) The businessman$_i$ walked to the office and *she$_j$ was rather tired.

Possible correction:

The businessman$_i$ walked to the office and *he$_j$* was rather tired.

3) My sister$_i$ visits home every year because *he$_j$ works for a company in England.

Possible correction:

My sister$_i$ visits home every year because *she$_j$* works for a company in England.

Error Type 05: Pronoun number error occurs as the writer misuses the pronoun number refereeing to its antecedent. For examples:

1) Peter does not like kids$_i$ because *it$_j$ is too noisy.

Possible correction:

Peter does not like kids$_i$ because *they$_j$* are too noisy.

2) Thai people$_i$ require more earnings because *she$_j$ has encountered higher cost of living.

Possible correction:

Thai people $_i$ require more earnings because *they$_j$* have encountered higher cost of living.

3) Enjoy the party$_i$! *They$_j$ provide you free drinks.

Possible correction:

Enjoy the party$_i$! *It$_j$* provides you free drinks.

Error Type 06: Pronoun mis-coreference occurs as the writer does not provide clear antecedent or there are multi possible antecedents in the sentences. For examples:

1) Jeff$_i$ told Tony$_j$ that *he$_k$ got a promotion.

Possible correction: (assuming Tony got promotion)

- *Tony told Jeff that Tony got a promotion.*

- *After getting promotion$_i$, Tony told Jeff about it$_j$.*

2) Peter spoke Chinese well enough on his trip to China; *it was encouraging.

Possible correction:

Peter spoke Chinese well enough on his trip to China; *his fluency* was encouraging.

-When Peter spoke Chinese well enough on his trip to China, it was encouraging.

3) In the new atlas, *it shows the recent changes in countries of Europe and Asia.

Possible correction:

The new atlas shows the recent changes in countries of Europe and Asia.

The possible correction might need a pronoun, a repeated noun phrase or a pronominal word with the same co-reference so that the sentence could be grammatical and comprehensible.

2.5 Summary of the Chapter

The chapter recalls the heyday of the traditional approaches of SLA, Contrastive Analysis and Error Analysis, during 1960s. Since then, the new age approaches of computer interlanguage analysis (CIA) without requirement of tagging processes and computer error analysis (CEA) with requirement of error tagging have emerged. Studies of Computer Learner Corpora have been young but stronger in terms of increasing number of the studies with computerized data associated with the theoretical root in SLA. These corpora have been conducted and collected around the world so far. Unfortunately, none of academic error tagging is universally and automatically applied. The uniqueness of the approach is now for its manual workload on error tagging with particularized design of tagsets to specific EFL learners of the same L1 background.

The error tagging can be designed in many ways; there is none of standard error tagsets generally employed by other research in the same area. Therefore, error tagging is introduced. Normally, the error tagset is blueprinted according to linguistic level studied. Ones designed in the present study are planned according to Thai EFL learners' background of pronoun in L1 and the previous errors made by the Thai EFL learners. It has been found that pronoun errors can be produced in many ways; consequently, the error tagset to be employed in the next chapter as a tool for CEA

informing pronoun types studied including pronoun errors categorized as omission, overuse, malformation (case, gender, and number errors), and misco-reference.

The further chapter will explain how the present study including the pronoun error tagset for Thai EFL learners at university level was designed. Associated with the main approaches mentioned earlier, the study is expected to figure out how Thai university English majors produce English pronouns in contexts; which pronoun error types occur; and pronoun development over the period they study in the university.

CHAPTER 3

RESEARCH METHODOLOGY

This chapter discusses how the study will be carried out. It consists of two main sections: 3.1) *the pilot study* and 3.2) *the present study* including subjects of the research; research design; research instruments; and data analysis and statistical techniques.

As a background the pilot study will inform the methodology of the present study before adjusting the research methodology which is presented in section 3.2.

3.1 The Pilot Study

The pilot study was conducted to examine possibilities and problems of the research methodology. Furthermore, pilot studies help the researcher adjust the study to be more appropriate based on the research questions. The present study had been piloted before the research methodology was designed for subjects, instrument, data collection, and data analysis.

3.1.1 Design of the pilot study

For the most part, the design considerations were based on Tono (2003:800) as mentioned in the previous chapter, section 2.2.3. However, some suggestions from Granger (2008) are employed, in learner-related issues in particular. The design considerations for building corpora consist of three main criteria of *language, task* and *learner*. Each criterion is comprised of sub-issues. The pilot study was planned

and reviewed type by type, both before and after piloting processes. Table 3.1 below shows information in brief, followed by detailed explanations of the design elements for the pilot study.

Table 3.1 Design considerations for building the present study (adapted from Tono 2003: 800 and Granger, 2008: 129-130)

Language-related	Learner-related	Methodology-related
Mode: Written	*Internal-cognitive*: 19-22 year- old learners	*Method of collection*: Cross-sectional
Genre: Essay	*L1 background*: Thai	*Method of elicitation*: Spontaneous
Style: Narration	*L2 environment*: EFL in university levels (Y1-Y4)	*Use of references*: Dictionaries permitted
Topic: General without preparation by the writers	*Institutional status:* English Majors (Y1-Y4)	*Time limitation*: free

Table 3.1 informs how the present study was designed. Criteria used in the design were relatively different from what was suggested by Tono in that they consisted of *Language-related, Learner-related, and Methodology-related* elements.

The pilot study examined the genre of narrative essay, based on its nature eliciting more pronouns as mentioned in the previous chapter section 2.2.2, with general topic without preparation by the writers. The narratives were *Little Red Riding Hood, Three Little Pigs,* and *Cinderella.* It was a cross-sectional study which will be clarified later in section 3.2. The data were obtained from 19 to 22-year-old English majors, Year 1- Year 4, with Thai mother tongue and in a Thai EFL context.

The pilot data were collected in natural way; in other words, the data collection was not set as an examination. The student writers were allowed to use dictionaries as they wanted. They were asked to complete the task within one hour; the appropriate duration of the task for the main study would be defined or adjusted from this duration after completing the pilot.

3.1.2 Results and Adjustments

The results of the pilot study verified that the overall research design was feasible. However, there were some minor points to be adjusted.

During the data collecting process, writing was timed and the appropriate time was set as 60 minutes. In addition, some of the pilot subjects informed researchers they could not recall the whole story and a few had not known the stories before; some pictures and CDs of the three stories are included so that the student writers have better recall when writing.

After collecting the data, all written texts were tagged and analyzed as planned. The adjustments that are described were essential for the actual research plan in terms of instruments and interpretations.

Table 3.2 Adjustments after piloting

	Pilot	**The present study**
Subjects	40 (10:10:10:10)	231 (70:56: 56:49), as available
Instrument	Narrative essays : Little Red Riding Hood/ Three Little Pigs/ Cinderella	Narrative essays : Little Red Riding Hood/ Snow White & the Seven Dwarfs/ Cinderella ⇨ CD in Thai to help the students better understand the stories
Tag-sets	XML tagsets: personal pronouns / errors	XML tagsets: pronouns including three cases (subject, object, and possessive ones)/ errors

The number of subjects was adjusted from 120 to 231; the present study included 231 subjects from all 4 groups due to availability of students in the population of English majors at the University level. Before writing their essays, the subjects would watch animated cartoons of each story so that they could easily recall the story. It is noted that all animated cartoons are in Thai to prevent imitation or quotation from the stories.

In addition, the narrative 'Three Little Pigs' was replaced with 'Snow White and the Seven Dwarfs' for better comparable to the other two narratives, which mainly involve human characters and not anthropomorphized animals.

In short, the pilot study, which had been conducted to confirm the main features of the research methodology before the actual implementation, helped adjust the design of the present study.

3.2 The Present Study

The final plan for the present study, including subjects, research design, research procedures, instruments, data collection, data coding and tagging processes, and data analysis, are presented below.

3.2.1 Subjects

In keeping with the results of the pilot study reported earlier, information of actual subjects has been defined and additional information about the subjects is provided.

The subjects were Thai EFL English majors at Ubon Ratchathani University, Thailand. They were grouped according to their institutional status; the subjects were organized into 4 groups out of a total number of 231 English majors: 70 students from Year 1; 56 students from Year 2; 56 students from Year 3, and 49 students from Year 4. L2 proficiency is presumably increasing with each year the subjects have studied in the university. Table 6 summarizes the information of subjects in the present study.

Table 3.3 Information of the subjects of the present study

Group	English major	Number of subjects (n)	Presumable L2 proficiency
1	Year 1	70	Year 1 English Majors: Beginner
2	Year 2	56	Year 2 English Majors: Pre-intermediate
3	Year 3	56	Year 3 English Majors: Intermediate
4	Year 4	49	Year 4 English Majors: Upper-intermediate
Total		231	

Although the group definition is given, the terms of language ability are not really verifiable, and are based on some necessary assumptions. Primarily, the assumed language ability of the subjects is based on the years of exposure to L2 at the university level and at Ubon Ratchathani University. Language development in terms of inter-language of pronoun proposed by Selinker (1972) might explain the position of the subjects within the distance between L1 and L2:

L1<---------------------------Y1----------------Y2------Y3-------------Y4---------------> L2

The beginner, Y1 English Major, was considered somehow far away from L1, but the subjects in this group were not yet half way to the destination, while other groups were closer to their destination, L2.

Additional Information about the Subjects

The subjects were English and Communication majors in Ubon Ratchathani University (UBU), a midsized university considered as a comprehensive university in Northeast of Thailand. They were like other Thai EFL learners at the same level in up-country universities, with similar background of English learning in out of town provinces before entering the universities. Most Thai EFL learners in tertiary institutions face similar contexts under their respective English Programs.

The English and Communication major is under the English Program, Department of Western Languages and Literature and Faculty of Liberal Arts. 75% of the permanent Thai teaching staff in the English Program holds PhDs from English-medium Universities. In addition, some native speakers of English work there as additional teaching staff. In a curriculum that consists of four years of study in a graded program at the university level, the subjects will be exposed to and instructed in the target language. The information presented below includes English courses taken by the students along four year curriculum

For instances, some information involves the subjects in general as well as some writing courses over time the subjects study in the university might provide a clearer picture for further discussion.

The subjects, therefore, have studied English as a foreign language in their home L1 society; they do not apply English in daily life, as described by Gass and Selinker (2001). For them, the EFL learning environment does not have much access to the language being learnt. All of the subjects speak Thai as their mother tongue, none have English-speaking family and none had lived outside Thailand.

There were seven writing –related courses taken during their four years. In the first year, the courses called Foundation English for Liberal Arts Students I (FE1) and

Foundation English for Liberal Arts Students II (FE2) are taken by the incoming freshmen in the major in semester 1 and semester 2, respectively. Each course provides 10 hours a week of instruction, 4 hours of instruction by Thai instructors and 6 hours by native instructors. Thai instructors dealt with reading, grammar and sentence level writing; the native instructors provide activities for listening and speaking. After taking the courses providing sentence level writing, in the second semester of the second year the students take writing courses entitled Written Expression I, II, and III. The three written courses were designed in series; the courses end at the end of the third year. Written Expression I covers with paragraph writing; Written Expression II expands to include essay writing and Written Expression III introduces research writing. Finally, Year 4 students are required to take a course called Language Style and Communication, which entails the students learning different communication methods, e.g. research papers, news, articles and oral communication such as radio and video documentaries. Then, in the second semester of the fourth year, the students are required to write an Independent Study thesis involving a paper of 2,500-3,000 words in length.

The dropout rate for English majors was approximately 5% between Year 1 and Year 2; however, among Year 3 and Year 4 students, the dropout rate fell down at almost zero.

To sum up, the subjects of the present study comprise 4 groups of Year1, Year 2, Year 3, and Year 4 English majors in UBU, Thailand. The subjects are native speakers of Thai. None of them have experiences overseas. During the time of their study in the university, the subjects are expected to take more courses involving

writing. Accordingly, all of the subjects were expected to write during their study, from the sentence level up to the essay level.

For the purpose of subject confidentiality, before collecting the data the subjects would be certainly asked for permission, and informed about their confidential protection of confidentiality and about the restrictions on any people who request to use the data.

3.2.2 Research Design

This study investigated the students' use of English pronouns in three different written tasks. It was a *pseudo-longitudinal study* emphasizing second language development, with data being collected at a single point in time, but with different proficiency levels represented (Gass and Selinker, 2001); the proficiency levels represented in the present study are based on the years of English exposure, which assumes a level of proficiency that is untested. This research method would be applied to see how pronouns are used by all students, and whether there are changes in pronoun use from Year 1 to Year 4 students. This acts as a surrogate for examining changes in pronoun use in the same subjects over time. The data were collected from the four subject groups three times during the academic year 1/2011 at UBU (June – October 2011) and analyzed as planned after piloting, if necessary. The research design largely follows the outline originally proposed by Tono (2003).

3.2.3 Design Considerations for the Present Study

The research design was based on considerations described by Tono (2003) and Granger (2008) as reviewed in chapter 2. The following table presents how the research had been planned.

Table 3.4 Design considerations for building the present study (adapted from

Tono 2003: 800 and Granger, 2008: 129-130)

Language-related	Learner-related	Methodology-related
Mode: Written	*Internal-cognitive*: 19-22 year- old learners (231 subjects)	*Method of collection*: Cross-sectional
Genre: Essay (at least 250 words)	*L1 background*: Thai	*Method of elicitation*: Spontaneous
Style: Narration (3 narratives)	*L2 environment*: EFL in university levels (Y1-Y4)	*Use of references*: Dictionary permitted
Topic: General without preparation by the writers	*Institutional status*: English Majors (Y1-Y4)	*Time limitation*: fixed (1 hour)

The integrated versions suggested by Tono and Granger for design building of learner corpus study shown in table 3.4 above reveals the whole picture of the present research design including what aspects are controlled in the study. The information on the table for building up the present research design in learner language are described, but not too detailed as some have already been discussed before in the section 3.1 about the pilot study.

The table concerns issues related to language in focus, learner background and research methodology for blueprinting learner language studies. The followings are design considerations in brief.

a) Considerations of language-related issues

The data consists of a narrative essay without prior preparation by the participants. The narrative task is a picture story telling task in which the meaning of the story is both well known and easily understood so that the subjects will not find

comprehension an obstacle to write. Given that pronoun use among Thai EFL learners is the language focus of the study, other difficulties, such as vocabulary, will be minimized by providing some essential vocabularies for the writings.

b) Considerations of learner-related issues

The last three categories for consideration are about the learner: internal-cognitive, internal-affective, first language background, second language environment and second language proficiency. The learners' internal-cognitive state in the present study is partially defined by the subjects' age. Other internal affective issues, such as anxiety, distress, frustration and resistance (Cohen, 1998, and Ellis, 1994) are not really explored in this study. However, any negative internal factors will be reported by the subjects at the interview after writing.

It is a primary supposition of the researcher that the L1 background of the learners is likely the primary culprit of any pronoun errors. Thai, the learners' L1, possesses some differences in pronouns from English both in syntactic and discoursal levels as mentioned in Chapter 2. On top of that, the L2 environment must also be borne in mind; the subjects have learned English in classroom settings with little access to the target language in reality; they are all EFL learners at a university with nearly 100% of the students speaking only the L1.

Finally, a quantitative assessment of the learners' English proficiency in the study is not applied in the study; instead, an approximate based on the term 'institutional status' defined as years of exposure to English acts as a differentiating term.

c) Considerations of methodology-related issues

The second consideration is task-related: method of collection, method of elicitation and time limitation. The data collection method planned for the present study is cross-sectional or pseudo-longitudinal so as to examine language developmental patterns of pronoun performance among the subjects differentiated according to the different institutional status as a correlation to different years of exposure to the target language at university level.

In addition, the method of elicitation is rather natural than prepared as the subjects will not have to prepare anything before the data collection. Moreover, the method is managed outside the language classroom with permission granted for using dictionaries; also some essential terms are provided to facilitate the writers. The present study does not examine the student writers' vocabulary knowledge; with time limitation, all potential interferences of writing comfort are reduced.

Lastly, a time limitation for the task is an hour. In total, each group takes 3 hours; however, the writing can be administered with more than one group at a time. Accordingly, the tentative maximum time for collecting the data from the whole group is 12 hours, based on the subjects' availability.

In short, the study design includes issues related to language in focus, learners as subjects, and methodology. Pronouns in narratives performed by EFL learners at four different university levels are to be studied. The data collection was administered outside language classrooms with relatively natural conditions. In addition, with dictionary permission and without content preparation and course evaluation involvement, the subjects, as a result, are possibly more relaxing when writing.

3.2.4 Research Procedures

The following flowchart describes how the results were processed.

1) **Selecting** an essay type appropriate to the study

⇩

2) **Selecting narrative topics:**   *Little Red Riding Hood, Snow White and Seven Dwarfs,* and *Cinderella*

⇩

3) **Piloting** and **adjusting** where needed

⇩

4) **Collecting** and **storing** the data in a computerized form   (Non-detected data)

⇩

5) **Detecting** pronouns in the data, manually   (PRO data)

⇩

6) **Tagging all pronouns**, based on the designated pronoun tagset, manually   (XML tagged PRO data)

⇩

7) **Tagging all PRO errors**, based on the designated error tagset, manually   (XML tagged PRO Error data)

⇩

8) **Counting** and **concordancing** tagged pronouns and tagged pronoun errors by   using WordSmith Tool

⇩

9) **Analyzing** the final data by statistical techniques and content analysis

In the research procedure above, steps (1) – (3) were explained in 3.1, piloting and adjustment. After that, steps (4) – (9) were possible to arrange. Section 3.2.5, research instruments, presents procedures in item (4). Later, in section 3.2.6, data analysis & analytical techniques, focuses on the rest steps.

3.2.5 Research Instruments

The choice of research instruments is determined by the research questions. Table 3.5 explains relationships between research questions and instruments eliciting data as expected.

Table 3.5 Relationships among Research Questions, Instruments, Data elicited, and Data Coding

	Instrument(s)	Data Elicited	Data Coding
Research question 1*	1) Narrative writing task with pictures	1) Essays prompted by the stories	⇨Non-detected PRO data ⇨PRO tagset for accuracy rates
Research question 2*	1) Narrative writing task with pictures	1) Essays prompted by the stories	⇨PRO error tagset
Research question 3*	1) Narrative writing task with pictures	1) Essays prompted by the stories	⇨PRO tagset ⇨PRO error tagset

*RQ1) How are pronouns produced by Thai university English majors in narrative writings?

RQ2) What are the differences of pronoun errors occurring in the narrative writing produced among the students in different four groups?

RQ3) What developmental patterns of pronoun acquisition can be seen to occur over time?

The instrument in this study included:

Narrative writing task consisted of three picture story telling: *Little Red Riding Hood, Snow White and the Seven Dwarfs, and Cinderella.* The students were assigned to write at least 250 words a story (for research questions 1-3). The total tokens of each story written by 231 students were at least 57,750 words and the whole tokens are approximately 173,250 words (see appendix B).

The Narrative Writing: Picture Story Telling

The narrative writings provided in the study consisted of three famous fables, *Little Red Riding Hood, Snow White & the Seven Dwarfs* and *Cinderella.* Before writing, the subjects watched a short animation of the stories so that they were able to recall them. Also, to help recalling and to reduce interference of vocabulary limitation, additional pictures of the stories arranged in sequence and some essential terms were given.

Data Collection:

The instruments were chosen and developed in consideration of the writing topic, time duration, coding and analyzing methods. The first page conforms to the guidelines suggested by ICLE, a project coordinated by University of Louvain (http://cecl.fltr.ucl.ac.be/Cecl-Projects/Icle/icle.hml), (see appendix B).

The instrument eliciting the data was a picture story-telling task prompted by three popular fables: Little Red Riding Hood, Snow White and the Seven Dwarfs, and Cinderella. The subjects were assigned to write three stories sequencing according to the given pictures. With dictionary permission and essential vocabulary attached, the subjects did not encounter word expression problems when writing. They were allowed to write within an hour for a story. In addition, animated CDs of the three fables were provided at the beginning to facilitate students recalling the stories with least difficulty as possible

After collected, the texts written by the students were archived in electronic forms requiring every single data originally obtained. This step took time since the researcher had to type each story; however, the next steps truly involved time-consuming and troublesome tasks, pronoun tagging and pronoun error tagging.

When stored in electronic forms, the data were detected for pronouns occurring in the students' writings. Later, they were annotated with pronoun tagset as planned. In this step, it is noted here that pronoun omissions could not be tagged. They would be tagged in the next step, pronoun error tagging. In this step, two inter-raters were needed to provide reliability of the tagging. One was the researcher and the other one was a native speaker of English who is an American holding bachelor degree in business management with Spanish as his minor. Before tagging both pronouns and pronoun errors, the raters had to be trained in order to maintain reliability of the tagging. Besides, whenever there were any different ideas of tagging, the discussion would be made online or on phone with some suggestions and comments from other native speakers in linguistic or applied linguistic fields.

All tagged pronouns and pronoun errors were categorized as the research plan; they were counted and analyzed with the statistical techniques discussed below.
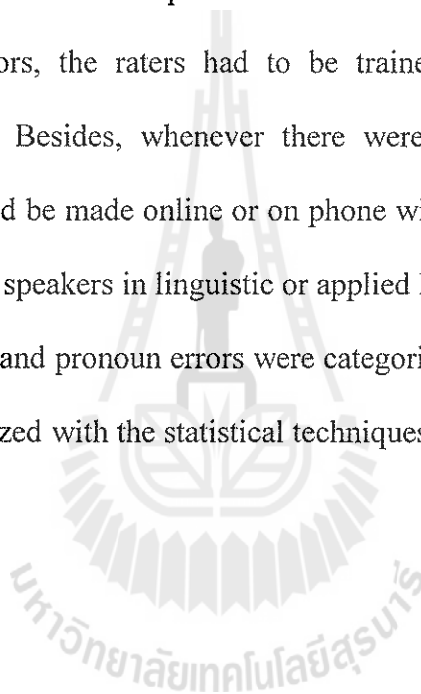
3.2.6 Data Analysis and Statistical Techniques

Table 3.6 provides a quick overview of the data analysis that was conducted.

Table 3.6 Relationships among the data coded, counting, and analytical techniques

Coded Data	Counted data by WS	Quantitative analysis (SPSS)	Qualitative analysis
1) Non-detected PRO data	⇨PRO (numeral data)	⇨Descriptive results (means / percentages)	
2) PRO accuracy rates	⇨ PRO accuracy rates (numeral data) ⇨ PRO in contexts	⇨Descriptive (means/ANOVA)	⇨ Text analysis
1) Tagged PRO errors	⇨ PRO error: in each group, within groups (numeral data) ⇨ PRO errors in contexts	⇨Descriptive results (mean/ANOVA)	⇨Text analysis (coding PRO errors in context)
1) Tagged PRO errors of six error types	⇨ Six PRO error types: in each group and within groups (numeral data) ⇨ Six PRO error types in contexts	⇨Descriptive results (mean/ANOVA)	⇨Text analysis (coding PRO errors in context)

Numerical data was statistically calculated and compared, while the data in contexts was explored by text analysis and concordance-based technique for the qualitative part.

In short, this section gives overview of how the written data were managed due to the research questions. The data were stored in electronic form and tagged, which will be discussed in the next section.

3.2.7 Data coding and tagging processes

1) Manual detecting for the occurrence of English pronouns in the data obtained

2) Manual error tagging of the errors regarding to error types

3) Word-Smith version 5.0 was the software used for counting and detecting errors occurring in the subjects' narrative writing

The data was tagged in two steps:

1) *XML text formatting*: categorizing the pronouns occurring in the task as planned in the pronoun tag-set

2) *FreeText & Tono set tagging*: annotating the errors according to the error taxonomies (syntactic and discoursal errors) as predetermined.

The data obtained was typed in Microsoft Word format and were converted to XML text format (Ide, 2000) so that they can be used for corpus processing software. The format appears like this:

<3PP_S> it </3PP_S >

XML tags have been proposed by Ide for part of speech categorizing. To convert the data to XML format, some formatting (e.g. fonts and margin) is lost; but some data remain, such as font style and paragraph breaks. Codes are inserted after the converting of pronouns produced in the essay writing.

In the present study, only English pronouns were tagged as P for personal pronouns in both subject and object cases, and PS for a possessive pronoun. The

numbers 1, 2, and 3 preceding the initial set of letters inform personal case features (1 is first person, 2 is second person and 3 is third person). The letters S and P following the initial P or PS (that is, the third position for personal pronoun or the forth position for possessive pronouns) provide the numerical value of the pronoun: S is for singular and P is for plural. Finally, the letters in the last position inside the angle brackets provide the functions or relative positions of the pronoun in the sentence: S is for a subject and O is for object, direct and indirect. The pronoun tagset employed in the present study is shown below.

Data coding in the present study are designed as following:

Pronoun Tagset

PRO	Subject position	PRO	Object position
I	<1PS_S> I </1PP_S>	me	<1PS_O/> me </1PS_O>
We	<1PP_S> we </1PP_S>	us	<1PP_O> us </1PP_O>
You	<2PP_S> you </2PP_S>	you	<2PP_O> you </2PP_O>
They	<3PP_S> they </3PP_S>	them	<3PP_O> them </3PP_O>
He	<3PSM_S> he </3PSM_S>	him	<3PSM_O> him </3PSM_O>
She	<3PSF_S> she </3PSF_S>	her	<3PSF_O> her </3PSF_O>
It	<3PS_S> it </3PS_S>	it	<3PS_O> it </3PS_O>
mine	<1PSS_S> mine </1PSS_S>	mine	<1PSS_O> mine </1PSS_O>
ours	<1PSP_S> ours </1PSS_S>	ours	<1PSP_O> ours </1PSS_O>
yours	<2PSS_S> yours </2PSS_S>	yours	<2PSS_O> yours </2PSS_O>
yours	<2PSP_S> yours </2PSP_S>	yours	<2PSP_O> yours </2PSP_O>
theirs	<3PSP_S> theirs </3PSP_S>	theirs	<3PSP_O> theirs </3PSP_O>
his	<3PSS_M_S> his </3PSS_M_S>	his	<3PSS_M_O> his </3PSS_M_O>
hers	<3PSS_F_S> hers </3PSS_F_S>	hers	<3PSS_F_O> hers </3PSS_F_O>
its	<3PSS_S> its </3PSS_S>	its	<3PSS_O> its </3PSS_O>

(adapted from McEnery, Xiao, and Tono, 2006: 254)

Some examples of the tagged data, originally performed by a year 1 pilot subject, are presented below.

Once upon a time, there is a pretties girl who wears a red cloak everybody called "Red Riding Hood". <3PSF_S> She </3PSF_S> lives with her mother at the edge of the forest...... After that, Little Red Riding Hood set off her path to her grandmother house. <3PSF_S> She </3PSF_S> walks along the way and <3PSF_S> she </3PSF_S> is happy with the nature around <3PSF_O> her </3PSF_O>.....Then <3PSM_S> he </3PSM_S> dressed himself with her night cloth and lie down on her bed. <3PSM_S> He </3PSM_S> is waiting for the Little Red Riding Hood. The Little Red Riding Hood arrived the cottage. <3PSF_S> She </3PSF_S> knocks at the door "Tock Tock Tock." "Little Red Riding Hood grandma" said Little Red Riding Hood. "Come in" said the wolf. <3PSF_S> She </3PSF_S> goes inside and go to the bed. "What big ears <2PP_S> you </2PP_S> have" said Little Red Riding Hood. "The ears to hear <2PP_S> you </2PP_S> with" said the wolf. "What big eyes <2PP_S> you </2PP_S> have" said Little Red Riding Hood. "The eyes to see <2PP_O> you </2PP_O> with" said the wolf. "What big mouth <2PP_S> you </2PP_S> have" said Little Red Riding Hood. "The mouth to eat <2PP_O> you </2PP_O> with" said the wolf. After finished his words, <3PSM_S> he </3PSM_S> ate <3PSF_O> her </3PSF_O> up. <3PSM_S> He </3PSM_S> sleeps on the bed and he snores so loud. He goes into the cottage and he sees the ugly wolf lies down on the bed. He takes the knife and he cuts his bully. Then the Little Red Riding Hood and her grandmother came out. The hunter puts the stones in the wolf's bully and <3PSM_S> he </3PSM_S> saw <3PS_S> it </3PS_S>. The wolf wakes up and <3PSM_S> he </3PSM_S> runs away, but his body is too heavy. <3PSM_S> He </3PSM_S> falls down and die.

The process mentioned earlier makes the data appropriate as input in the POS corpus-processing software, WordSmith Tool version 5.0. The output is called a non-error tagged corpus (Muehleisen, 2006); in this case, the emphasis is on pronoun accuracy in the data obtained. In error tagging, Tono (2002: 804) stated that most kinds of errors cannot be identified automatically; but program will have been developed for some simple kinds of errors possibly occurring in particular context, for

example overuse of a pronoun in a single clause. In addition, a particular error like omission calls for human judgment only.

The present study includes some errors that are likely to occur in Thai EFL learners' essay writing and provides error annotation tagsets adapted from FreeText format and NICT JEL (Negrillo and Domínguez 2006; McEnery, Xiao, and Tono 2006) as follows:

Error type	Sub-Tagset	Example(s)
ER01: Omission	< */ER_>	- If *want to talk to stranger.
ER02: Overuse	<A/ER_>	- Little Red Riding Hood **she** is a small and pretty girl. - A little girl, who **she** is called Little Red Riding Hood.
ER03: Case error	<MF- C/ER_>	- Mother asked **she** to visit her grandma.
ER04: Gender error	<MF-GEN/ER_>	-Little Red Riding Hood$_i$ liked flowers and **he**$_i$ wanted to pick some for grandma.
ER05: Number error	<MF-NUM/ER_>	- The girl picked some flowers$_i$ and gave **it**$_i$ to her grandma.
ER06:Mis-coreference	< Co/ER_>	- Little Red Riding Hood$_i$ and her grandma$_j$ went away from the wolf's tummy. After that, **they**$_{*i/j/k}$ celebrated with food and wine in the basket.

(adapted for Thai EFL learners from McEnery, Xiao, and Tono, 2006: 254)

The present study follows the guidelines of McEnery et al (2006) that omissions are omitted subjects/objects are in account of errors when they are found 'n sentences or fragments with a predicate; fragments without a verb phrase will be ignored.'

As mentioned in chapter 2, a significant limitation in any Error Analysis approach occurs in unclear error categories. To resolve blurred categories, some possibly overlapped categories were managed before the actual tagging. When tagged, malformation errors consisting of ER04 (number error), ER05 (gender error), and ER05 (number error) were at times similar to ER06 (mis-coreference). It was hard to make decision which category the error was under. When they encountered such difficulty, the taggers followed these guidelines:

1) When finding an error possibly under ER03 to ER06, the tagger has to figure out its antecedent in the previous sentence;

2) If the possible antecedent is discovered, the error might be under malformation, ER03-ER05; and

3) If not, the error might be under ER06, mis-coreference, due to an excessively remote antecedent.

For examples:

"Little Red Riding Hood looked at flowers and <3PSF>she</3PSF_S> sure grandmother liked <MF-NUM/ER_3PS_O>them#it</3PS_O>. <3PSF>She</3PSF_S> walked around and picked them. <Co/ER_3PSM_S>The wolf#He</3PSM_S> went to her grandmother's house quickly. Still <3PSF>she</3PSF_S> picked <Co/ER_3PP_O>the flowers#them</3PP_O> and sang song…"

(an excerpt from LRH written by Y1 student#07)

Three errors were tagged in this excerpt. The first one is an error of number since '*it*' refers to '*flowers*' in the previous sentence. In this case, the possible antecedent can be found in the previous clause or sentence, so the error is grouped under 'ER05', number error. In a different way, the third error in the excerpt is under ER06, mis-coreference, because the pronoun '*them*' referring to '*flowers*' is located too remote from its antecedent. Consideration of the distance of the antecedent helps to make more clear-cut the line between malformation and mis-coreference. The error tagset includes error descriptions can ensure more reliable manual tagging.

The error tagset was designed to describe error types and also provide correct and appropriate word(s) for each error instance. It was added into the first POS tagset with Free-Text format providing error types and correction into the tagset (Negrillo and Dominguez, 2006). The first letter set (e.g. X/ER) identifies type of error and the rest is in parenthesis is the original POS tagset. Correct and appropriate word(s) are then presented (YY) and closes the tag with #. Although pronoun omissions had not been POS tagged earlier, pronoun omissions were tagged with possible intended pronouns in accordance with the error position. The tagset applied in the present study is shown below.

Pronoun Error Tagset

PRO	Subject position	PRO	Object position
I	<X/ER_1PS_S> I </1PP_S>	me	< X/ER_1PS_O/> me </1PS_O>
We	< X/ER_1PP_S> we </1PP_S>	us	< X/ER_1PP_O> us </1PP_O>
You	< X/ER_2PP_S> you </2PP_S>	you	< X/ER_2PP_O> you </2PP_O>
They	< X/ER_3PP_S> they </3PP_S>	them	< X/ER_3PP_O> them </3PP_O>
He	< X/ER_3PSM_S> he </3PSM_S>	him	< X/ER_3PSM_O> him </3PSM_O>

She	< X/ER_3PSF_S> she </3PSF_S>	her	< X/ER_3PSF_O> her </3PSF_O>
It	< X/ER_3PS_S> it </3PS_S>	it	< X/ER_3PS_O> it </3PS_O>
mine	< X/ER_1PSS_S> mine </1PSS_S>	mine	< X/ER_1PSS_O> mine </1PSS_O>
ours	< X/ER_1PSP_S> ours </1PSS_S>	ours	< X/ER_1PSP_O> ours </1PSS_O>
yours	< X/ER_2PSS_S> yours </2PSS_S>	yours	< X/ER_2PSS_O> yours </2PSS_O>
yours	< X/ER_2PSP_S> yours </2PSP_S>	yours	< X/ER_2PSP_O> yours </2PSP_O>
theirs	< X/ER_3PSP_S> theirs </3PSP_S>	theirs	< X/ER_3PSP_O> theirs </3PSP_O>
his	< X/ER_3PSS_M_S> his </3PSS_M_S>	his	< X/ER_3PSS_M_O> his </3PSS_M_O>
hers	< X/ER_3PSS_F_S> hers </3PSS_F_S>	hers	< X/ER_3PSS_F_O> hers </3PSS_F_O>
its	< X/ER_3PSS_S> its </3PSS_S>	its	< X/ER_3PSS_O> its </3PSS_O>

(adapted from http://latl.unige.ch/freetext/en/publication.html)

For reliability, two annotators were employed to implement the error tag-set scheme: the researcher and a native speaker of English. Both were trained before tagging and they were able to discuss any disagreement during tagging. This technique, referred to as intra-coder, improves the reliability of the tagging process.

Examples of pronouns tagged in the previous part and retagged with error tag-set are shown below. Words in parentheses after the tagged data identify the noun referents. (LRH = Little Red Riding Hood)

Once upon a time, there is a pretties girl who wears a red cloak everybody called <*/ER_3PSF_O> her#* </3PSF_O> "Red Riding Hood". <3PSF_S> She </3PSF_S> (LRH) lives with her mother at the edge of the forest..... After that, Little Red Riding Hood set off her path to her grandmother house. <3PSF_S> She </3PSF_S> (LRH) walks along the way and <3PSF_S> she </3PSF_S> (LRH) is happy with the nature around <A/ER_3PSF_O> herself or Ø#her <A/ER_/3PSF_O>. Then the wolf appears and said "Where are <2PP_S> you </2PP_S> going Little Red Riding Hood." "I am going to visit my grandmother who lives in the forest" <3PSF_S> she </3PSF_S> said. The cunning wolf tricks <3PSF_O> her </3PSF_O> (LRH) out off her way. Little Red Riding Hood is happy with the colorful flowers. <3PSF_S> She </3PSF_S> picks up and put

<*/ER_3PP_O> them#* </3PP_O> in the basket. <3PSF_S> she </3PSF_S> thinks her grandmother would like <MF-NUM/ER_3PS_O>them#it </3PS_O> . The wolf arrived grandmother house. The wolf knocks at the door "Tock Tock Tock." "Who is there" said grandma. "Little Red Riding Hood" the wolf distinguished his voice. "Come in" <MF-C/ER_3PSF_S> she#her </3PSF_S> said. The greedy wolf go straight on to her bed and eat <3PSF_O> her </3PSF_O> up. .....After finished his words, <3PSM_S> he </3PSM_S> ate <3PSF_O> her </3PSF_O> up. <3PSM_S> He </3PSM_S> (the wolf) sleeps on the bed and <3PSM_S> he </3PSM_S> snores so loud. <Co/ER_3PSM_S> He </3PSM_S> goes into the cottage and <3PSM_S> he </3PSM_S> sees the ugly wolf lies down on the bed. <Co/ER_3PSM_S> He </3PSM_S> takes the knife and <3PSM_S> he </3PSM_S> cuts his bully. Then the Little Red Riding Hood and her grandmother came out. The hunter puts the stones in the wolf's bully and <3PSM_S> he </3PSM_S> saw <Co/ER_3PS_S> NP#it </3PS_S>. The wolf wakes up and <3PSM_S> he </3PSM_S> runs away, but his body is too heavy. <3PSM_S> He </3PSM_S> falls down and die.

The inter-raters coding the text into the error tagset designed had been trained before beginning manual tagging. In the event of any discrepancy, the taggers discussed and arrived at consensus, thus greatly increasing the reliability of the analysis and interpretation (see Cresswell, 2005).

After processing and counting with Word-Smith Tool version 5.0, results of the tagsets were analyzed quantitatively with both statistical analysis and content analysis.

Word-Smith Tools:

Word-Smith Tools (WS) by Scott (1998) is an integrated set of programs for observing the ways words are performed in texts. It consists of three main programs: Concord, Word-List, and Key-Word. The first program, Concord, provides the users accessibility to see any word or phrase in context, so that the users can see what type of context it keeps. The second program is Word-List letting users view a list of all

the words or word-clusters in a text studied, alphabetically or frequently reported as set by the users. The other, Key-Word, supports searching the key words in a text. The WS version 5.0 was applied in the present study.

WS version 5.0 was applied in the study because of its availability and ability to conduct follow-up concordance searches. It can be installed and run on removable drives such as flash drives and USB hard drives. Finally, it is able to convert from .PDF, .DOC, removing all mark-up, etc.

Word-Smith was applied in the study mainly to count frequency of pronouns that occur in the writings and to understand the contexts in which pronoun errors occur. This information was applied after the errors were tagged. The results were quantitatively summarized and then statistically analyzed as described in the next part.

3.2.8 Data Analysis

The present study employs data analysis in two ways: statistical analysis and text analysis for the best appropriate answers to the research questions.

Statistical analysis:

The converted and coded tag-sets were analyzed using the Statistical Package for Social Science (SPSS program) for pronoun accuracy and inaccuracy rates, percentage values and the means. After that, comparisons of the results in focus were subjected to f-test, ANOVA and Scheffe's tests to determine levels of significance of differences across the groups ($P \leq 0.05$).

Text Analysis:

The data from the essay writing was subjected to text analysis. The sentences with pronouns were emphasized, categorized and conceptualized (Cresswell, 2005) to

help determine how and why the Thai EFL learners produced the pronouns they did and in the contexts they did. Examples of interesting pronoun uses will be discussed.

For examples, results from the study were analyzed in terms of text analysis. Some interesting data among groups are underlined for pronouns used and errors. The subjects in group 1, 2, and 3 tended to produce more frequent pronouns in their writings, while the subject in group 4 were inclined to use both pronouns and other noun phrases to replace the referents. For instances, a subject from group 2 wrote:

'<3PSF_S> She </3PSF_S> brought a food basket to her grandma's house in a forest. <3PSF_S> She </3PSF_S> walked through the forest. And <3PSF_S> she </3PSF_S> met the wolf on the way to the house.'

A subject from group 4 wrote:

'*Little Red Riding Hood* held a basket with food, fruits and wine. While *the little girl* walked and sang along the way to the cottage, <3PSF_S> she </3PSF_S> was greeted by a wolf. '

With similar narration, the two students used pronouns differently. More sentences, from the actual study, might lead to a clearer picture and better interpretation for language developmental patterns.

Additionally, there was a team of 3 text coders as inter-raters of the qualitative data so that the analysis and interpretation are more reliable. The raters included Mrs. Lagkana Waters (holding bachelor degree of English and work as an instructor for Regina School, Iowa City, Iowa, the US), Mr. Chanon Jintawet (holding a master degree of applied linguistics), and the researcher. They had to be trained how to code, categorize and conceptualize the text. Consensus among the raters was ideally expected; however, two third of the raters would be in majority agreement if any different coding took place.

3.3 Summary of the Chapter

The chapter consists of two main parts: pilot study report and the present study. The pilot study was briefly reported to reconfirm that the research plan would be doable and analyzable. The plan was adjusted after piloting and analyzing the pilot data to redesign the present study.

The present study was designed according to the research questions by applying criterion proposed by Tono (2003) and suggested by Granger (2008) which provides three main criterion consisting of issues related to language (mode, genre, style, and topic), learners or subjects (internal cognitive, L1 background, L2 environment, and institutional status), and methodology (method of data collection, data elicitation, use of reference, and time limitation).

The present study planed to investigate language, personal pronouns in particular, in written mode as a narrative essay with general topic without writer preparation produced by Thai university English majors from year 1 to year 4, 18-22 years old, who learn English as a foreign language (EFL) by employing crossectional data collection method in natural setting with dictionary permission and within an hour for each essay. The subjects were assigned write three narrative essays: *Little Red Riding Hood, Snow White and the Seven Dwarfs,* and *Cinderella.*

After collecting the data in handwriting format, the original data were typed in computerized format so that the researcher could continue the study procedures: manual pronoun tagging and error tagging. The tagging systems employed in the present study had been designed according to McEnery, Xiao, and Tono (2006) with considerations of suggestions given by Granger (2003 and 2008) that the effective

tagging system must be informative, reusable, reusable, and consistent error tagging rules between annotators.

Finally, the tagged pronouns and tagged pronoun errors were counted for statistical analysis: means and percentages of each group and multi-comparisons (ANOVA) within the four groups to figure out pronoun development over time according to the research questions. In addition, text analysis was employed at the end to investigate how the students in each group produced pronouns in context and how pronoun errors occurred in contexts. This process employed WordSmith Tool to capture the language in focus and arrange in contexts.

In conclusion, the chapter reports the piloting before adjusting to the present study research design. The subjects in the present study were English majors from Year 1 to Year 4; all of them are Thais with EFL environment. The instruments eliciting the subjects' data were three narratives. All raw data were transformed from written texts to computerized texts so that the data analysis would be possible. After the data were tagged with the tagsets designed as planned, the tagged data were both quantitatively and qualitatively analyzed. The results will be discussed in the next chapter.

CHAPTER 4

RESULTS

This chapter provides results of pronoun use as produced in narrative writings by different four student groups, differences of pronoun errors across the four groups, and how the students' pronoun use changes over time. These results will be analyzed according to the research questions:

RQ1) How are pronouns produced by Thai university English majors in narrative writings?

RQ2) What are the differences of pronoun errors occurring in the narrative writing produced among the students in four different groups?

RQ3) What developmental patterns of pronoun acquisition can be seen to occur over time?

Accordingly, both quantitative and qualitative results will be reported. A combination of quantitative and qualitative results will be employed for answering the first and second research questions, while the answers for the third research question will be entirely qualitative ones. All results are presented based on the research plan.

The study aims to investigate pronouns produced by Thai University English majors in narrative writings. Pronouns here include personal pronouns in subject (I, we, you, they, he, she and it), object (me, us, you, them, him, her and it) and possessive (mine, ours, yours, theirs, his, hers and its) cases.

The study hypotheses state that the subjects will produce pronouns with greater accuracy and fewer pronoun errors over time (data occurring from the end of year 1 to the end of year four) possibly due to first language transfer as the main culprit, along with over-generalization and unsuccessful teaching materials and methods. The data were collected from three narratives (*Little Red Riding Hood, Snow White and the Seven Dwarfs,* and *Cinderella*) and managed as the research methodology (see chapter 3).

In accordance with the research methodology, 231 subjects from four different groups wrote three narratives. The pronouns used in the narratives were manually detected and annotated with a pronoun tagset designed for all pronoun cases. Later, the errors were manually annotated with the error tagset for all six error types. Consequently, all tagged data were counted and analyzed with statistical techniques for both individual group results and comparative results.

To answer each research question, the quantitative results will be presented below in a form of tables and graphs with statistical explanations of the results. Firstly, the tables and reports begin with the whole picture of the data obtained: tokens, pronouns occurring, and errors, including comparisons across the four groups. Next, errors will be reported in comparisons across the four groups; all results of the six error types will be presented. Finally, comparative results with developmental graphs will be reported. Qualitative results will be reported as text analysis and some examples in contexts by using WordSmith Tool at the end of each research question.

4.1 Results for Research Question 1: *How are pronouns produced by Thai university English majors in narrative writings?*

4.1.1 Quantitative Results:

The organization of the results will start with descriptive and comparative information of (a) tokens and pronouns and (b) pronoun accuracy and errors, in sequence.

(a) Tokens & Pronouns

Before reporting pronoun information in the narratives, overall information of the writings will start here as a background. Overall information of tokens in the narratives produced by Thai university English majors is shown in table 4.1. Although the students had been assigned to write at least 250 words each narrative, the average tokens of each narrative is 330. Total tokens, therefore, are more than what was anticipated, about 175,000 words.

Table 4.1 Tokens obtained in three narratives by the four groups

Subjects	Narrative 1 *Little Red Riding Hood*	Narrative 2 *Snow White*	Narrative 3 *Cinderella*	Total Tokens
Year 1 (n =70)	17,988	17,504	17,537	53,029
Year 2 (n = 56)	19,033	16,937	18,882	55,033
Year 3 (n = 56)	19,620	18,652	18,696	56,968
Year 4 (n = 49)	21,184	21,022	21,372	63,578
All (N = 231)	77,825	74,115	76,487	228,608

As shown in table 4.1, all groups could produce approximately 76,000 tokens for each narrative; the total token for the learner corpus in the present study is 228,608. Roughly, group 4 was able to write using the highest number of words, followed by group 3, group 2 and group 1, respectively. The raw data, however, cannot tell us anything specific for now; statistically analysis will be present hereafter.

After collecting raw data and counting tokens of each narrative by each group, pronouns were detected and tagged. Later, the pronouns were checked for accuracy and errors. The correct pronouns were not retagged; while the pronoun errors were retagged with the error tagsets22 including possible corrections. Finally, the tokens and tagged pronouns were calculated for means as shown in table 4.2 below. Some abbreviations are used here after: LRH as Little Red Riding Hood, SNW as Snow White and the Seven Dwarfs, and CDR as Cinderella's

Table 4.2 Means and percentages of tokens and pronouns in narrative writings by the English majors (Year 1 to Year 4)

		Tokens	Pronoun	
		$\bar{x}$	$\bar{x}$	%
Year 1 (n = 70)	LRH	302.63	30.50	10.1
	SNW	300.31	35.07	11.7
	CDR	305.31	33.99	11.1
	Total	**908.26**	**99.56**	**11.0**
Year 2 (n = 56)	LRH	321.21	28.62	8.9
	SNW	312.57	31.16	10.0
	CDR	313.16	32.30	10.3
	Total	**946.95**	**92.90**	**9.8**
Year 3 (n = 56)	LRH	350.36	29.61	8.5
	SNW	333.07	30.38	9.1
	CDR	333.86	31.30	9.4
	Total	**1017.29**	**91.29**	**9.0**
Year 4 (n = 49)	LRH	392.12	35.35	9.0
	SNW	345.65	29.67	8.6
	CDR	385.35	33.16	8.6
	Total	**1123.12**	**98.18**	**8.7**
TOTAL (N = 231)	LRH	337.69	30.86	9.1
	SNW	320.84	31.84	9.9
	CDR	331.11	32.75	9.9
	Total	**989.65**	**95.45**	**9.6**

LRH: Little Red Riding Hood / **SNW**: Snow White and the Seven Dwarfs / **CDR**: Cinderella

Table 4.2 shows means and percentages of tokens, pronouns, and pronoun accuracy produced by the subjects in each group (Year 1, Year 2, Year 3 and Year 4) and in total. After being tagged and counted, pronouns were calculated for percentages out of the whole set of tokens occurring in the narratives. In addition, results from each individual narrative by all four groups are included. On average, the students could write more than 250 words as assigned; the mean of total word of all three narratives is 989.65. Year 4 students could write the highest number of words both across all narratives and individually: 1,123.22 mean score in total; 392.12 mean

score of LRH; 345.65 mean score of SNW; and 385.35 mean score of CDR. Among the four groups, Year 1 students scored the lowest both in total narratives and each: 908.26 mean score in total; 302.63 mean score of LRH; 300.31 mean score of SNW; and 305.31 mean score of CDR.

Although Year 1 students produced the lowest mean of total words in the writings (302.63), they produced the highest mean of total pronouns (99.56) or 11.0%. Interestingly, the number is close to the mean of pronouns produced by Year 4 group (98.18), which produced the highest number of tokens in the four groups; however, when considered in terms of the percentage of pronouns out of all the tokens, Year 4 produced the least, 8.7%, among the four groups. In addition, percentages of pronouns out of the tokens are ranked from Year 1, Year 2, Year 3, and Year 4, respectively. These rankings alone might not provide enough information; comparisons of tokens and pronouns across the four groups possibly reveal a better picture of the data.

Table 4.3 to 4.6 will show comparative results of tokens before moving to comparative results of pronouns in table 4.7 to 4.10 among the four groups.

Mean comparisons of tokens, pronouns, pronoun accuracy, and pronoun errors in each type were analyzed with f-test ANOVA and tested with Scheffe's to figure out significant differences between a pair within the groups. Though the numbers of tokens and pronouns occurring are not actually emphasized in the study, they will be reported as background information of comparisons across the four groups.

Tokens produced by the four groups were analyzed with f-test and tested with Scheffe's. The results reveal highly significant differences in the means of tokens among four groups ($p \leq 0.01$). Consequently, multi-comparisons across groups have

been tested with Scheffe's to investigate which pairs have statistically significant differences (higher or lower) from one another, in both total and individual narrative.

Multi-comparison of tokens produced by the four groups (ANOVA) revealed statistically significant differences across these groups; Scheffe's test, therefore, was used to see which pairs had significant differences. Tables 4.4-4.6 reveal results from the test.

Table 4.3 Multi-comparisons of total tokens in *Little Red Riding Hood* tested with Scheffe's

Year		1	2	3	4
	$\bar{x}$	302.63	321.21	350.36	392.12
1	302.63	--	.187	.000**	.000**
2	321.21		--	.015*	.000**
3	350.36			--	.000**
4	392.12				--

* Significantly different at $p \leq 0.05$; ** Significantly different at $p \leq 0.01$

This table shows comparisons of tokens produced in *Little Red Riding Hood* across the four groups. Year 4 group could produce significantly higher numbers of tokens compared to other three groups ($p \leq 0.01$). Also, a comparison between Year 3 and Year 1 show a significantly higher mean of tokens performed ($p \leq 0.01$); Year 3 showed a significantly higher mean score than Year 2 could, at 0.015 ($p \leq 0.05$). However, Year 1 and Year 2 could produce no statistical differences of the number of tokens in *Little Red Riding Hood*.

Table 4.4 Multi-comparisons of total tokens in *Snow White and the Seven Dwarfs* tested with Scheffe's

Year		1	2	3	4
	$\bar{x}$	300.31	312.57	333.07	345.65
1	300.31	-	.301	.000**	.000**
2	312.57		-	.028*	.000**
3	333.07			-	.357
4	345.65				-

* Significantly different at p ≤ 0.05; ** Significantly different at p ≤ 0.01

Comparisons of tokens produced in *Snow White and the Seven Dwarfs* reveal four pairs with significant differences. Year 1 group wrote with a significantly lower mean number of tokens than Year 3 and Year 4 groups (p ≤ 0.01). Plus, Year 2 group used a very significantly lower mean number of tokens than Year 4 group (p ≤ 0.01) and the group had a significantly lower mean number of tokens than Year 3 group (p ≤ 0.05). On the other hand, two pairs without statistically different mean scores of tokens are between Year 1 and Year 2 and between Year 3 and Year 4.

Table 4.5 Multi-comparisons of total tokens in *Cinderella* tested with Scheffe's

Year		1	2	3	4
	$\bar{x}$	305.31	313.16	333.86	385.35
1	305.31	-	.689	.000**	.000**
2	313.16		-	.028*	.000**
3	333.86			-	.000**
4	385.35				-

* Significantly different at p ≤ 0.05; ** Significantly different act p ≤ 0.01

According to table 4.5, Year 4 group wrote a very significantly higher number of words than the other three groups (p ≤ 0.01) in *Cinderella*. Year 3 group had a

significantly higher mean score of tokens than Year 1 (p ≤ 0.01); Year 3 group did significantly higher mean of tokens than Year 2 group (p ≤ 0.05). However, a comparison between Year 1 and Year 2 groups reveals no statistical differences.

Table 4.6 Multi-comparisons of total tokens in all narratives tested with Scheffe's

Year		1	2	3	4
	$\bar{x}$	908.26	946.95	1017.29	1123.12
1	908.26	-	.218	.000**	.000**
2	946.95		-	.005**	.000**
3	1017.29			-	.000**
4	1123.12				-

** Significantly different at p ≤ 0.01

Table 4.6 shows five pairs within the four groups who wrote significantly different numbers of total tokens (p ≤ 0.01). Year 4 group produced significantly higher numbers of tokens in three narratives than Year 3, 2, and 1 did. Other two groups producing very significant lower mean scores of token than Year 3 are Year 1 and Year 2. The other pair, Year 1 (with 908.26 mean score) and Year 2 (with 946.95 mean score), produced no statistical differences in comparisons.

Overall, due to the tests of comparative tokens with Scheffe's, Year 4 group produced significantly higher numbers of tokens measured as mean scores than the rest. Group 3 could write more tokens than the others group in lower year with shorter exposure to English in classroom. However, group 1 and group 2 produced no statistical differences in total tokens.

The next section provides comparative information of the pronouns produced by the four groups. It begins with multiple comparisons before analyzing these results with Scheffe's test.

After comparing pronoun mean scores found in total and in each narrative written by the four groups, ANOVA or multi-comparison test was completed to figure out if any statistical differences within the four groups. The results from ANOVA test showed significantly different mean scores of pronouns performed across the four groups. Later, analysis with Scheffe's was tested, comparative results across the four groups in each narrative are showed in tables 4.7-4.9 and ones of all narrative are showed in table 4.10 below to show pairs with significant differences of pronoun mean scores.

Table 4.7 Multi-comparisons: total pronouns occurring in LRRH tested with

Scheffe's

Year		1	2	3	4
	$\bar{x}$	30.50	28.61	29.61	35.35
1	30.50	-	.000**	.576	.000**
2	28.61		-	.540	.000**
3	29.61			-	.000**
4	35.35				-

** Significantly different at p ≤ 0.01

When individually considered in *Little Red Riding Hood*, each pair across the four groups turns out as follows. Year 4 had higher mean scores than the others (p ≤ 0.01), and Year 1 produced significantly higher numbers of pronouns than Year 2 did.

However, comparisons between Year 3 and Year 1 and between Year 3 and Year 2 show no significant differences.

Table 4.8 Multi-comparisons: total pronouns occurring in SNW tested with Scheffe's

Year		1	2	3	4
	$\bar{x}$	35.07	31.16	30.38	29.67
1	35.07	-	.000**	.000**	.000**
2	31.16		-	.611	.110
3	30.38			-	.716
4	29.67				-

** Significantly different at $p \leq 0.01$

In accordance with table 4.8, Year 1 produced significantly higher mean scores of pronouns in *Snow White and the Seven Dwarfs* than the other three groups ($p \leq 0.01$). Comparative results of the other pairs do not show any statistical differences of pronouns produced in the narrative.

Table 4.9 Multi-comparisons: total pronouns occurring in CDR tested with Scheffe's

Year		1	2	3	4
	$\bar{x}$	33.99	32.30	31.30	33.16
1	33.99	-	.145	.004**	.751
2	32.30		-	.630	.754
3	31.30			-	.136
4	33.16				-

** Significantly different at $p \leq 0.01$

Table 4.9 shows the results of pronouns produced in *Cinderella* across the four groups. The table shows only one pair with a significant difference of pronoun mean scores; Year 1 produced a larger number of pronouns than Year 3 could ($p \leq 0.01$). No other statistical differences were found.

Table 4.10 Comparisons: total pronouns occurring in all narratives tested with

Scheffe's

Year		1	2	3	4
	$\bar{x}$	99.56	92.09	91.29	98.18
1	99.56	-	.000**	.000**	.868
2	92.09		-	.971	.006**
3	91.29			-	.001**
4	98.18				-

** Significantly different at $p \leq 0.01$

In table 4.10, all pronoun frequencies across all narratives and all four groups are collated. , Year 1 had the highest mean score of pronoun frequency in total; the result of Year 1 and Year 4 are very significantly higher than Year 2 group and Year 3 group ($p \leq 0.01$). However, no statistical differences are showed between Year 1 and Year 4 and between Year 2 and Year 3.

Notably, none of the possessive pronouns were found in the students' writing; when showing possession, the students normally used possessive determiners.

In addition to the umbers and frequencies of pronouns produced by the four groups, pronoun accuracy and errors will be reported to further address the first research question.

(b) Pronoun accuracy and errors

Information of pronoun accuracy and errors will be presented firstly with descriptive information before moving to comparative analysis. The descriptive results show means and percentages of pronoun accuracy out of the pronouns tagged; the descriptive results of pronoun errors, means, will follow in sequence. Later, comparative results will be revealed.

Table 4.11 Means and percentages of pronouns and pronoun accuracy in the narratives produced by the four groups

Subjects		Pronoun	Accuracy		Errors*
		$\bar{x}$	$\bar{x}$	%	$\bar{x}$
Year 1 (n = 70)	LRH	30.50	20.23	66.3	12.14
	SNW	35.07	23.99	68.4	13.14
	CDR	33.99	22.59	66.5	13.01
	Total	99.56	66.80	67.1	38.30
=Year 2 (n = 56)	LRH	28.62	18.70	65.3	11.57
	SNW	31.16	22.21	71.3	10.66
	CDR	32.30	22.96	71.1	10.48
	Total	92.90	63.88	68.8	32.71
Year 3 (n = 56)	LRH	29.61	21.70	94.8	8.68
	SNW	30.38	22.61	75.4	7.86
	CDR	31.30	23.00	73.8	8.29
	Total	91.29	67.41	73.8	24.47
Year 4 (n = 49)	LRH	35.35	28.08	79.4	7.90
	SNW	29.67	22.90	77.2	6.84
	CDR	33.16	25.53	77.0	7.73
	Total	98.18	76.51	77.9	22.47
TOTAL (N = 231)	LRH	30.86	21.88	70.9	10.26
	SNW	31.84	22.99	72.2	9.92
	CDR	32.75	23.43	71.5	10.13
	Total	95.45	68.30	71.6	30.32

This data was generated during the tagging process. After pronouns had been tagged and categorized (according to person, number, gender, and case), the tagged pronouns were corrected; if any errors were found, they were tagged with the appropriate error tagset as planned. Pronoun errors were tagged into five categories, which will be discussed in the next section; however, some errors such as omissions were detected in the later step. While the other pronoun errors could be found in the initial step (pronoun tagging), omissions could not.

Total average mean score of pronoun accuracy by all groups is 68.30 out of 95.45 mean score of pronoun produced. In other words, the accuracy percentage of all groups is 71.6% of all pronouns produced. When compared among the four groups, mean score of Year 4 is the highest, with a 76.51 rate of accuracy, while the mean score of Year 2 is the lowest one, with a 63.88 rate of accuracy. Based on percentage comparisons, however, Year 1 could do the lowest pronoun accuracy (67.1%) and the other groups as Year 2, Year 3, and Year four could do better with accuracy percentages of 68.8%, 73.8%, and 77.9%, respectively. That means group 1 produced the lowest rate of pronoun accuracy out of the whole population of pronouns found in their writings.

After detecting and tagging pronouns occurring in the narratives, correcting and error tagging were managed as planned. Error mean calculations are shown in table 4.3; however, the errors could not be calculated for percentages of total pronouns since pronoun omissions were not included at the first stage. The results reported in the table tell that all groups produced pronoun errors with a mean of 68.30. In addition, the results of the errors conform to the accuracy data in that Year 1 produced the highest errors (with a mean of 38.30), followed by Year 2 (with a

mean of 32.71), Year 3 (with a mean of 24.47), and Year 4 (with a mean of 22.47), respectively.

By testing with ANOVA for multi-comparison of pronoun accuracy mean scores produced by the four groups in total and in individual narrative reveal very significant differences in both total and individual narrative (p ≤ 0.01) within the four groups, significant differences within the four groups were found. Testing with Scheffe's, therefore, is essential here to see which pairs are statistically different (See tables 4.12-4.15).

Table 4.12 Multi-comparisons: total pronoun accuracy in LRH tested with Scheffe's

Year		1	2	3	4
	$\bar{x}$	20.23	18.70	21.70	28.08
1	20.23	–	.047*	.063	.000**
2	18.70		–	.000**	.000**
3	21.70			–	.000**
4	28.08				–

* Significantly different at p ≤ 0.05; ** Significantly different at p ≤ 0.01

In *Little Red Riding Hood*, Year 4 produced significantly higher rates of pronoun accuracy than the other groups (p ≤ 0.01). Year 3 had higher mean rates of pronoun accuracy than Year 2 group (p ≤ 0.01); however, their mean score is not statistically different from Year 1. In addition, the table shows significantly different mean scores of pronoun accuracy between Year 1 and Year 2 (p ≤ 0.05) in the narrative.

Table 4.13 Multi-comparisons: total pronoun accuracy in SNW tested with Scheffe's

Year		1	2	3	4
	$\bar{x}$	23.99	22.21	22.61	22.90
1		-	.007**	.058	.226
2			-	.906	.667
3				-	.963
4					-

** Significantly different at $p \leq 0.01$

Table 4.13 looks at "Snow White." The comparisons of pronoun accuracy presented in table 4.13 indicates that only one pair of groups showed any difference in mean rates of pronoun accuracy. Year 1's mean score is statistical higher than Year 2's ($p \leq 0.01$). None of the other group data show any significances of pronoun accuracy in the narrative.

Table 4.14 Multi-comparisons: total pronoun accuracy in CDR tested with f-test, ANOVA

Year		1	2	3	4
	$\bar{x}$	22.59	22.96	23.11	25.59
1		-	.942	.862	.000**
2			-	.997	.002**
3				-	.004**
4					-

** Significantly different at $p \leq 0.01$

Table 4.14 reports that Year 4 group's mean rate of pronoun accuracy in *Cinderella* is significantly higher than the other groups (p ≤ 0.01); however, within those groups, there is no difference in pronoun accuracy produced in the narrative.

Table 4.15 Multi-comparisons: total pronouns accuracy in all narratives tested with

Scheffe's

Year		1	2	3	4
	$\bar{x}$	66.80	63.88	67.41	76.51
1	66.80	-	.184	.975	.000**
2	63.88		-	.096	.000**
3	67.41			-	.000**
4	76.51				-

** Significantly different at p ≤ 0.01

When all results are collated in Table 4.15, Year 4 group showed a higher rate of correct pronouns than the others in total narratives (p ≤ 0.01).

In short, comparative results within and between groups verifies that Year 4's level of pronoun accuracy was the highest among the four groups, but no differences were significant between or among the other groups. The nature of pronoun errors will be reported in the following sections.

Table 4.16 Means and percentages of errors produced by the four groups

	ER01 $\bar{x}$	%	ER02 $\bar{x}$	%	ER03 $\bar{x}$	%	ER04 $\bar{x}$	%	ER05 $\bar{x}$	%	ER06 $\bar{x}$	%	TOTAL
Y1	5.58	14.66	4.93	12.96	4.14	10.88	1.1	2.89	1.86	4.89	20.44	53.72	38.05
Y2	4.5	13.75	3.38	10.33	2.12	6.48	0.64	1.96	1.2	3.67	20.88	63.81	32.72
Y3	0.93	3.76	2.11	8.52	0.91	3.68	0.06	0.24	0.3	1.21	20.45	82.59	24.76
Y4	0.76	3.38	0.76	3.38	0.45	2.00	0.06	0.27	0.18	0.80	20.27	90.17	22.48
TT	3.24	10.69	2.98	9.83	2.09	6.89	0.53	1.75	0.97	3.20	20.51	67.65	30.32

The results in table 4.16 display the tag information of errors. The single most common error type is tagged as ER06, mis-coreference, which occurred about 20% from among all tagged errors across all of the narratives. All groups did approximately the same percentage of the error type, 20%. The rest most common errors among the four groups are ER01 (pronoun omission), ER02 (overuse), and ER03 (case error), respectively.

In terms of group comparisons, generally Year 1 group did the highest error mean scores of 5 error types: ER01 (omission), ER02 (overuse), ER03 (case error), ER04 (gender error), and ER05 (number error). All groups did almost the same mean scores of ER06 (mis-coreference). In addition, Year 4 group did the lowest mean scores of all pronoun error types among the four groups. Table 4.16 roughly tells about how many pronoun errors were produced by the four groups; however, to

figure out statistical comparisons in details, the results need to be tested with ANOVA, a statistical multi-comparisons for more than 3 groups. Also Sheffe's test will be employed if any significant differences within the four groups are found.

To sum up the information in this section, group 4 could do the highest mean scores among the four groups when compared in terms of tokens and pronoun accuracy, while group 2 did the lowest mean score of pronoun accuracy. Other than tokens and pronoun accuracy, the trends go in different ways. When considered pronouns produced by each group, group 1 produced the highest pronouns in the writings. Group 1 has the highest mean scores of error types 01 to 06; while group 4 has the lowest mean scores of errors. Although these results approximately inform the better pronouns are produced due to longer time exposure to English, statistically comparative analysis within the groups will inform better comprehensible results of how each group differs from peers. The results are interesting that how Year 4 students differently used pronouns form the other three and what kinds of words they used to replace noun phrases which were old information in the contexts. Qualitative results in section 4.1.2, therefore, may complete the jigsaw puzzle of how pronouns are produced.

4.1.2 Qualitative Results:

The qualitative results in this section emphasize how pronouns were produced in context. The first step is to use the analytical software program WordSmith to derive concordances of the tagged tokens in focus in order to see how and when pronouns occur. Then, the texts were examined manually where the software program was limited. In particular, the researcher would see the types of substitutions, such as noun phrases, that would be used to replace a personal pronoun. The

following points are from the researcher's observation for examining pronoun choices, pronoun accuracy, and pronoun errors.

(a) Pronoun characteristics found in the student writings

As mentioned earlier that means of pronouns produced by students with shorter time exposure to English in the university were more than ones produced by students with longer time exposure to the target language in the university, it is interesting that how the students in different groups used pronouns in the narratives.

Students with shorter exposure to formal L2 study (in this case, English) at the University level were more likely to replace a noun with pronouns and proper names; while students with longer exposure to formal University level study of English were likely to replace a noun phrase with a variety of forms: a new similar noun phrase, a definite noun phrase (the same noun phrase with the added article 'the'), and a personal pronoun. The following sentences are examples from the students in lower year.

"Once upon a time, Cinderella was her name. <u>Cinderella</u> <u>she</u> used to have mother but <u>her</u> died. Father married new wife. <u>They</u> had two daughters. After that <u>she</u> killed <u>him</u> and took all money and house to own. <u>They</u> gave her little food and ragged clothes. And <u>she</u> must to worked for all everything in house for <u>them</u>. <u>They</u> lived like millionaire but Cinderella <u>she</u> lived like slave....

One day everyone got invitation letter to a party from king. <u>He</u> announce if lady <u>she</u>..."

(Cinderella, original texts by Year1
student #05)

This Year 1 student normally used a personal pronoun after initially mentioning a noun phrase or a proper name, in this excerpt, such as *'Cinderella', 'mother', 'father*

and new wife', 'father', and *'lady'.* The student replaced these noun phrases with a

personal pronoun, such as replacing *'Cinderella'* with *'she'*; *'mother'* with *'her'*;

'father and new wife' with *'they'*; and *'father'* with *'him'.* After initially using the

proper noun, *Cinderella,* however, the proper noun was hardly seen after that.

Similar examples were found in the excerpt from Year 2 student's narrative as
shown below.

> "Once upon a time Cinderella was a girl who <u>she</u> lived with her parents.
> After her mother died. <u>She</u> was sad and her father married again with a bad lady
> who <u>she</u> had two children. <u>They</u> did not like <u>her</u>. Then <u>he</u> died too. <u>They</u> let <u>her</u>
> do everything in house. <u>She</u> was sad because <u>she</u> must to live in a kitchen. And
> <u>she</u> wore old dress but <u>they</u> wore new and beautiful dresses. <u>She</u> must to work all
> day but <u>they</u> were convenient...."

> (Cinderella, original texts by Year2 student#08)

As seen in the excerpt, once the student mentioned a proper noun (*Cinderella*)

and noun phrases (*a bad lady, two children,* and *father*) personal pronouns (*she, they,*

her, and *he*) were used to replace the mentioned noun phrases. *'Cinderella'* was

replaced with *'she'* and *'her'.* *'a bad lady'* was replaced by 'she' The personal

pronoun *'they'* in line 3 could refer to 'father and a bad lady', *'two children'* or *'a bad*

lady and two children'. The pronoun *'he'* takes a place of *'father'.* It is noted here

that the student keep using personal pronouns although the antecedents become

remote.

Even with the same length of writing, students with longer exposure to English

produced pronominals differently.

> "Once upon a time, Cinderella was a pretty girl. <u>She</u> lived with her family
> happily. But after her mother passed away her father married again with a wicked
> lady who had two daughters. <u>The ladies</u> were very cruel. <u>They</u> treated <u>her</u> like a
> slave. <u>Cinderella</u> lived in a small bedroom closed to the kitchen. While <u>the</u>
> <u>stepsisters</u> lived in house with happiness. <u>They</u> dressed beautifully but <u>she</u>

dressed very old and ragged outfit every day. **This pretty girl** had to do all
housework alone ..."

According to the excerpt above, the Y3 student substituted *'Cinderella'* with
pronouns *'she'*, and *'her'*, the proper noun *'Cinderella'* when it might be ambiguous,
and with a definite noun phrase *'this pretty girl'*. In addition, the student took the
place of *'the wicked lady and two daughters'* with a pronoun *'they'* and a definite
noun phrase *'the ladies'*. However, it is rather ambiguous here whether *'the ladies'*
refers to *'the wicked lady and two daughters'* or *'two daughters'*. Also, the student
replaced *'two daughters'* with a pronoun *'they'* and a definite noun phrase *'the step
sisters'*.

The following excerpt is from Year 4 student; her writing reveals similar results.

"Once upon a time in a house, Cinderella lived unhappy. Her mother was
dead. Her father had married another woman, a widow with two daughters. And
they didn't like her. **Cinderella** was lovely and nice but **she** dressed was ragged
and old. Her face was dusty gray from coals. While **the stepsisters** wear dresses
splendid and elegant their clothes, **they** are still ugly. **They** lived in the house
comfortably, but **Cinderella** was not. **The young girl** worked hard all the day.
She lived alone..."

The student replaced noun phrases like *Cinderella,* and *two daughters* with
more varied choices of pronominal forms. The student took the place of a proper noun
'Cinderella' with a pronoun *'she'*; a proper noun *'Cinderella'* when the antecedent is
rather distant; and a definite noun phrase *'the young girl'*. In the same way, the
student substituted a pronoun *'they'* possibly for *'her father and a widow'* or *'a widow
and two daughters'*. Additionally, *'two daughters'* is replaced with a pronoun *'they'*
and a definite noun phrase *'the step sisters'* when its antecedent is remote.

From the excerpts, it can be seen that more pronominal forms such as proper nouns, pronouns, and definite noun phrases were produced by students who had greater exposure to University level English instruction; while pronouns were frequently used by students with shorter exposure to the target language.

(b) No possessive pronouns produced in the student writings

Though a lot of personal pronouns both in subject and object cases were found in the students' writings, no pronouns in the possessive case were found. Instead, the students commonly used 'possessive determiners' when they wanted to express possession in the narratives no matter how many time they had mentioned the possession before, for example:

...She can fit the shoe because it was **her shoe**...

...she asked "why do your teeth are very sharp and long? "**My teeth** very long because I want to eat you!!"...

Instead of using possessive pronoun *'hers'* in the first example and *'Mine'* in the second example, the learner writers used possessive determiner (*her* and *my*) before a noun. Even still, these possessive determiners were not found frequently in these narratives.

This section looked at how pronouns were produced by the students in general. With qualitative analysis, other features used for replacing a noun phrase were found more in writings of Year 3 and Year 4 students; while personal pronouns and proper nouns were commonly used among students in Year 1 and Year 2. It is interesting that these Thai EFL learners did not produced any pronouns in possessive

case; once they had to express possession they normally use possessive determiners. However, due to the low levels of possession in the narratives, the data is insufficient to draw any conclusions.

4.1.3 Conclusions

In brief, students in Year 1 and Year 2 often used pronouns to replace a noun or a noun phrase while students in Year 3 and Year 4 used pronouns, proper nouns, and definite nouns to take the place of the antecedents. Although Year 2 students had the lowest mean score of pronoun accuracy, Year 1 had the highest mean rate of errors. This phenomenon occurs because Year 1 errors included more omissions which were not counted as pronouns at first. A more rigorous analysis of the nature of the errors is presented next in section 4.2.

4.2 Results for Research Question 2: *What are the differences of pronoun errors occurring in the narrative writing produced among the students in different four groups?*

4.2.1 Quantitative Results

The table below shows the mean scores and percentages of pronoun errors in six different types produced by students in Year 1 to Year 4 and by the whole group. In addition, the information is presented both in total and in each narrative.

Table 4.17 Mean occurrences of pronoun errors Type 01 to Type 06 performed by each group

		ER01 $\bar{x}$	%	ER02 $\bar{x}$	%	ER03 $\bar{x}$	%	ER04 $\bar{x}$	%	ER05 $\bar{x}$	%	ER06 $\bar{x}$	%	TOTAL
Year 1	LRH	15.90	15.9	1.67	13.7	1.19	9.8	0.29	2.4	0.71	5.8	6.36	52.3	12.14
(n = 70)	SNW	16.44	16.4	1.57	11.9	1.50	11.4	0.40	3.0	0.53	4.0	6.99	53.2	13.14
	CDR	13.37	13.4	1.69	13.0	1.46	11.2	0.41	3.2	0.61	4.7	7.10	54.6	13.01
	Total	**15.22**	**15.2**	**4.93**	**12.9**	**4.14**	**10.8**	**1.10**	**2.9**	**1.86**	**4.9**	**20.44**	**53.4**	**38.30**
Year 2	LRH	14.17	14.2	1.39	12.0	1.11	9.6	0.20	1.7	0.71	6.1	6.52	56.4	11.57
(n = 56)	SNW	15.95	16.0	1.12	10.5	0.34	3.2	0.23	2.2	0.30	2.8	6.96	65.4	10.66
	CDR	11.07	11.1	0.86	8.2	0.68	6.5	0.21	2.0	0.18	1.7	7.39	70.5	10.48
	Total	**13.76**	**13.8**	**3.38**	**10.3**	**2.12**	**6.5**	**0.64**	**2.0**	**1.20**	**3.7**	**20.88**	**63.8**	**32.71**
Year 3	LRH	8.64	8.7	0.98	11.3	0.52	6.0	0.07	0.8	0.21	2.4	6.14	70.8	8.68
(n = 56)	SNW	1.40	1.4	0.59	7.5	0.16	2.0	0.00	0.0	0.04	0.5	6.96	88.5	7.86
	CDR	0.84	0.8	0.54	6.5	0.24	2.9	0.05	0.6	0.05	0.6	7.34	88.5	8.29
	Total	**3.80**	**3.8**	**2.11**	**8.5**	**0.91**	**3.7**	**0.06**	**0.2**	**0.30**	**1.2**	**20.45**	**82.6**	**24.47**
Year 4	LRH	7.72	7.7	0.24	3.0	0.37	4.7	0.06	0.8	0.14	1.8	6.47	82.0	7.90
(n = 49)	SNW	0.58	0.6	0.14	2.0	0.04	0.6	0.00	0.0	0.00	0.0	6.61	96.8	6.84
	CDR	1.29	1.3	0.37	4.8	0.04	0.5	0.00	0.0	0.04	0.5	7.18	92.9	7.73
	Total	**3.38**	**3.4**	**0.76**	**3.4**	**0.45**	**2.0**	**0.06**	**0.3**	**0.18**	**0.8**	**20.27**	**90.2**	**22.47**

Table 4.17 Mean occurrences of pronoun errors Type 01 to Type 06 performed by each group

		ER01 $\bar{x}$	%	ER02 $\bar{x}$	%	ER03 $\bar{x}$	%	ER04 $\bar{x}$	%	ER05 $\bar{x}$	%	ER06 $\bar{x}$	%	TOTAL
Total	LRH	12.57	12.6	1.13	11.0	0.83	8.1	0.16	1.6	0.47	4.6	6.37	62.1	10.26
(N = 231)	SNW	11.09	11.1	0.92	9.3	0.58	5.8	0.18	1.8	0.24	2.4	6.90	69.6	9.92
	CDR	8.39	8.4	0.93	9.2	0.68	6.7	0.19	1.9	0.25	2.5	7.25	71.4	10.13
	Total	**10.69**	**10.7**	**2.98**	**9.8**	**2.09**	**6.9**	**0.53**	**1.7**	**0.97**	**3.2**	**20.51**	**67.6**	**30.32**

ER01 = Omission, **ER02** = Overuse, **ER03** = Case Error, **ER04** = Gender Error, **ER05** = Number Error, **ER06** = Misco-reference

LRH: Little Red Riding Hood, **SNW**: Snow White and the Seven Dwarfs, **CDR**: Cinderella

Table 4.17 shows that mis-coreference (ER06), contributing mean of 20.51, occurred the most frequently among six error types in

the narratives written by the four groups. The others typically occurred for a few: ER01-omission (3.24); ER02-overuse (2.98); ER03-

case error (2.09); ER05-number error (0.97); and ER04-gender error (0.53), respectively. The better comprehensible picture of error

types produced by all four groups is:

ER06 > ER01 > ER 02 > ER03 > ER06 > ER05

When compared among the four groups, each error type was ranked from the highest mean to the lowest mean as follow:

ER01-Omission: Y1 > Y2 > Y3 > Y4

ER02-Overuse: Y1 > Y2 > Y3 > Y4

ER03-Case Error: Y1 > Y2 > Y3 > Y4

ER04-Gender Error: Y1 > Y2 > Y3 = Y4

ER05-Number Error: Y1 > Y2 > Y3 > Y4

ER06-Mis-coreference: Y2 > Y3 > Y1 > Y4

In rough comparisons, table 4.17 presents the mean scores of each pronoun error type by all four groups in both total and each narrative. The errors produced by the four groups are also reported with standard deviation for consideration of the data distribution. The errors have been categorized into six groups: Error type 01 (ER01) or omission; Error type 02 (ER02) or overuse; Error type 03 (ER03) or case error; Error type 04 (ER04) or gender error; Error type 05 (ER05) or number error; and Error type 06 (ER06) or misco-reference. In total, the students' average error mean score is 30.32: Year 1 had the highest error rates (38.05), followed by Year 2 (32.72), Year 3 (24.75), and Year 4 (22.48), respectively. Mean scores of Error 01 to Error 05 were obviously gradually lower when the students are assumed to have more exposure to the target language. However, means of Error 06, mis-coreference, made by all groups were not lower over time; instead, they are rather stable.

As mentioned earlier, the rankings of error types produced by the group contributed similar trends; further comparisons across the four groups were essential here to see if any were statistically different (a significant difference is at $p \leq 0.05$, and a very significant difference is at $p \leq 0.01$). The means of errors produced by each group were tested with f-test ANOVA. There were some significant differences at $p \leq 0.01$; Scheffe's test was processed after that to identify different pairs. The test results indicated very significantly different means of error types 01, 02, 03, 04, 05, and in total across the groups ($p \leq 0.01$). However, no significant differences were found among the groups for error type 06, mis-coreference. The following tables (table 4.21- table 4.26) are comparative results of six pronoun error types tested with Scheffe's to figure out which groups are statistically different from another one.

Table 4.18 Multi-comparisons: total pronoun errors tested with Scheffe's

Year		1	2	3	4
	$\bar{x}$	38.30	32.71	24.82	22.47
1		-	.000**	.000**	.000**
2			-	.000**	.000**
3				-	.963
4					-

** Significantly different at $p \leq 0.01$

Scheffe's test for all error types of pronoun reported in table 4.22 reveals that Year 1 produced very significantly higher mean rates of the error than Year 2, Year 3, and Year 4 ($p \leq 0.01$). In the same way, Year 2 produced very significantly higher means than Year 3 and Year 4 ($p \leq 0.01$). However, comparative result of error mean scores between Year 3 and Year 4 does not show any statistically significant differences.

Table 4.19 Multi-comparisons: pronoun errors type 01, omission, tested with

Scheffe's

Year		1	2	3	4
	$\bar{x}$	5.83	4.50	0.93	0.76
1		-	.000**	.000**	.000**
2			-	.000**	.000**
3				-	.963
4					-

** Significantly different at p ≤ 0.01

Comparisons of pronoun omission or error type 01 across the four groups are presented in table 4.19. Year 1 produced very significantly higher omissions than Year 2, Year 3 and Year 4 (p ≤ 0.01). Also Year 2 produced very significantly higher mean rates of omissions than Year 3 and Year 4 (p ≤ 0.01). Interestingly, no statistical differences of omissions were found between Year 3 and Year 4.

Table 4.20 Multi-comparisons: pronoun errors type 02, overuse, tested with Scheffe's

Year		1	2	3	4
	$\bar{x}$	4.93	3.38	2.11	0.76
1		-	.000**	.000**	.000**
2			-	.000**	.000**
3				-	.000**
4					-

** Significantly different at p ≤ 0.01

Comparative results of pronoun overuse across the four groups reveal very significantly different rates between all pairs. Year 1's mean of pronoun overuse is significantly higher than Year 2's, Year 3's and Year's 4 (p ≤ 0.01). In the same way, the mean for Year 2 is significantly higher than that for Year 3 and Year 4 (p ≤ 0.01). Finally, the comparative result between Year 3 and Year 4 shows the former is significantly higher than the later one (p ≤ 0.01).

Table 4.21 Multi-comparisons: pronoun errors type 03, case error, tested with Scheffe's

Year		1	2	3	4
	$\bar{x}$	4.14	2.12	0.91	0.45
1		-	.000**	.000**	.000**
2			-	.000**	.000**
3				-	.181
4					-

** Significantly different at p ≤ 0.01

Comparative results of pronoun case error in table 4.21 show Year 1 produced very significantly higher case error mean scores than Year 2, Year 3 and Year 4, Year 4 (p ≤ 0.01). Also Year 2 produced very significantly higher mean score of case errors than Year 3 and Year 4 (p ≤ 0.01). However, no statistical differences of omissions were found between Year 3 and Year 4.

Table 4.22 Multi-comparisons: pronoun errors type 04, gender error, tested with

Scheffe's

Year		1	2	3	4
	$\bar{x}$	1.10	0.64	0.12	0.06
1	1.10	-	.000**	.000**	.000**
2	0.64		-	.000**	.000**
3	0.12			-	.973
4	0.06				-

** Significantly different at p ≤ 0.01

Comparative results of pronoun gender errors in table 4.22 show Year 1

produced very significantly higher gender error mean scores than Year 2, Year 3 and

Year 4 (p ≤ 0.01). In addition Year 2 produced very significantly higher mean scores

of pronoun gender errors than Year 3 and Year 4 (p ≤ 0.01). On the other hand, no

statistical differences of omissions were found between Year 3 and Year 4.

Table 4.23 Multi-comparisons: pronoun errors type 05, number error, tested with

Scheffe's

Year		1	2	3	4
	$\bar{x}$	1.86	1.20	0.30	0.18
1	1.86	-	.000**	.000**	.000**
2	1.20		-	.000**	.000**
3	0.30			-	.899
4	0.18				-

** Significantly different at p ≤ 0.01

Comparisons of pronoun error type 05, number error, across the four groups

are shown in table 4.23. There were no statistically different mean results between

Year 3 and Year 4. However, the other pairs resulted in very significantly different

mean scores of pronoun number errors ($p \leq 0.01$). Year 1 produced very significantly

higher number error mean score than Year 2, Year 3 and Year 4 as well as Year 2

produced very significantly higher mean scores of pronoun number errors than Year 3

and Year 4.

Table 4.24 Multi-comparisons: pronoun errors type 06, mis-coreference, tested with

Scheffe's

Year		1	2	3	4
	$\bar{x}$	20.44	20.88	20.45	20.27
1	20.44	-	.710	1.00	.975
2	20.88		-	.747	.511
3	20.45			-	.977
4	20.27				-

Finally, table 4.24 with comparative results of error type 06 or mis-
coreference reveals no statistically different mean scores of the error type 06. The

approximate number of the mean of each group is about 20.

In brief, comparisons of tokens, pronoun products, pronoun accuracy and

pronoun errors across the four groups from Year 1 to Year 4 were analyzed with f-
test, ANOVA; later, reports of pair comparative analysis with Scheffe's have been

employed and presented. Rather high means of the use of pronouns and pronoun

errors of types 01 to 06 produced by Year 1 and Year 2 are statistically close to each

other; as may be expected, their mean scores of pronoun accuracy are the lowest of the four groups. In contrast, lower means of pronouns and pronoun errors in types 01 to 06 produced by Year 3 and Year 4 are statistically close to each other; their means of pronoun accuracy are higher than the other two groups'.

When considering differences across the groups by each type of error, Year 3 and Year 4 had lower error rates than the other two groups at error types 01 to 06. Results of the error types 01, 03, 04, and 05 produced by Year 3 and Year 4 showed no statistically significant differences between the two groups, thus showing similar results in these 4 types or errors. In error type 02, Year 4 did significantly better than Year 3. The results of the final error type, mis-coreference, revealed all four groups had comparable errors. To examine the students' errors, errors in contexts will be reported in the next section.

4.2.2 Qualitative Results

To give more comprehensible results of errors, the data in contexts will be presented here with some examples excerpted from the student writings.

(a) Error Type 01: Omissions generally found in cases of subject and object of the verb, but not in object of preposition

The underuse of pronoun occurred in both subject and object cases; however, when considered by text analysis, pronouns used as the object of prepositions were never omitted. The following are some examples of pronoun omissions found in the students' performance. It is noted here that the examples were excerpted from Y1 and Y2 papers since most of omissions were tagged in their writings. The first group is the omission in subject case.

... carried a food basket. <*/ER_3PSF_S>She#*</3PSF_S> very happy..

....ready to go. And then <*/ER_3PSF_S>She#*</3PSF_S> get a food...

....the flowers so pretty. <*/ER_3PSF_S>She#*</3PSF_S> picked some..

... to go now". After that <*/ER_3PS_S>it#*</3PS_S> ran to her grandma

..Knock! Knock! Knock! <*/ER_3PS_S>it#*</3PS_S> called loudly "Hi!.

..the hunter must left . So <*/ER_3PSF_S>She#*</3PSF_S> felt scared..

.....After the Queen died <*/ER_3PSM_S>He#*</3PSM_S> marry again

....her father was killed. <*/ER_3PSF_S>She#*</3PSF_S> had many...

...hated her very much. <*/ER_3PP_S>They#*</3PP_S> like to spite her

Next, omissions in object case will be exemplified.

...basket. The girl took <*/ER_3PS_O>it#*</3PS_O> to grandma's house

...grandmother and bit <*/ER_3PSF_O>her#*</3PSF_O> fiercely. And..

.... the door and found <*/ER_3PS_O>it#*</3PS_O>. Hunter and dog ...

.... did not want to tell <*/ER_3PSF_O>her#*</3PSF_O>. Then.....

...a garden Soldier hit <*/ER_3PSF_O>her#*</3PSF_O> very hard

...back home and saw <*/ER_3PSF_O>her#*</3PSF_O> in their beds.

..... because they sent <*/ER_3PSF_O>her#*</3PSF_O> to live in a...

.....ran and forgot tell <*/ER_3PSF_O>her#*</3PSF_O> because time..

Lastly, pronouns as object of preposition were not omitted in the student writings. The pronouns go behind prepositions which collocate with some verbs, e.g.

talk with, talk to, marry to, cook for, dance with, look at, look after, and others. Some examples excerpted from the student data are given below.

..she</3PSF_S> talked to <3PS_O>it</3PS_O> and forgot every..

.. field there and looked at <3PP_O>them</3PP_O> and smiled…

.. the witch got angry with <3PS_O>it</3PS_O> very much. Later

..the house and married to <3PSF_O>her</3PSF_O> and then back

..so Cinderella cooked for <3PP_O>them</3PP_O> every day…..

..sisters wanted marry to <3PSM_O>him</3PSM_O> and became..

Due to the results, objects of preposition seem omissible for the students.

(b) Error Type 02: Some overuse of pronoun found in relative clauses

Although some heads and objects were overused with pronouns after a noun phrase, some of them were found in relative clauses. According to the data produced by students Year 1 and Year 2, some overuses were found in relative clauses. For examples,

.. the girl who <A/ER_3PSF_S>*#she</3PSF_S> was living with..

… wolf which <A/ER_3PS_S>*#it</3PS_S> was very hungry and..

..grandma who <A/ER_3PSF_S>*#she</3PSF_S> tailored her dress

..the lady who <A/ER_3PSF_S>*#she</3PSF_S> was a bewitch…

…dwarfs who <A/ER_3PP_S>*#they</3PP_S> were small and..

…house which <A/ER_3PS_S>*#it</3PS_S> was in deep forest

...the lady who <A/ER_3PSF_S>*#she</3PSF_S> had two
daughters

...a prince who <A/ER_3PSM_S>*#he</3PSM_S> was very clever

...an lady who <A/ER_3PSF_S>*#she</3PSF_S> can fit the shoe

Overuse of pronouns in relative clauses was frequently produced by students in Year 1 and Year 2.

(c) Error Types 03, 04, 05: Some malformation of case, gender, and number possibly just mistakes, produced by Year 3 and Year 4 in particular

Before looking through the examples of malformation, it is noted here that these types of error were found more in Year 1 and Year 2 student writings; but rarely found in Year 3 and Year 4 groups. From text analysis by considering data from the same students, particularly ones in Year 3 and Year 4, making malformation errors, the error types were not systematic in the student writings.

Error 03: Case Errors

The students misused pronoun object forms by placing subject forms, e.g. 'she' for 'her', 'they' for 'them'. In addition, the student misused a pronoun subject form by placing an object form like 'her' for 'she'.

..wanted to fool <MF-C/ER_3PSF_O>her#she</3PSF_O> to spend..

....went to pick <MF-C/ER_3PP_O>them#they</3PP_O> in the gar..

... queen hated <MF-C/ER_3PSF_O>her#she</3PSF_O> so much..

..pushed and hit <MF-C/ER_3PSF_O>her#she</3PSF_O> before ..

....and told that <MF-C/ER_3PSF_S>she#her</3PSF_S> do work in

...remembered <MF-C/ER_3PSF_O>her#she</3PSF_O> at all. So..

Remarkably, overall case errors were found more in object case; ones found in students in Year 3 and Year 4 were also in object case. Very few were found in subject case and they were produced by Year 1 students.

Error Type 04: Gender Error

This error type was not usually found in the narratives written by all four groups; the errors almost disappeared among Year 3 and Year 4 students. The examples excerpted are in contexts so that the antecedent can be seen. The antecedents will be underlined in the examples.

<u>Little Red Riding Hood, written by Year 3 student#44</u>:

"..picked enough flowers and <3PS_S>it</3PS_S> was late now. <u>The girl</u> walked to her grandmother's house and <MF-GEN/ER_3PSF_S> she#he </3PSF_S> though that the house was very quiet...."

From this excerpt, the error produced by the student might be a misspelling of the word '*he*' since there was not such error found in her writings.

<u>Little Red Riding Hood, written by Year 4#08</u>

"<u>The hunter</u> tried to followed it for along time until <MF-GEN/ER_3PSM_S>he#she</3PSM_S> a dog stopped at the house.."

The student wrote '*she*' to refer to '*the hunter*' instead of '*he*' which is the correct form; however this error did not reoccur in other narratives written be the same student, thus the conclusion that it was just an example of momentary carelessness.

Cinderella, written by Year 3 student #23:

"Every ladies wanted to attend the dance party. Cinderella received

letter too. <MF-GEN/ER_3PSF_S>She#He</3PSF_S> also want

to go but <3PSF_S> she </3PSF_S> did not have any new dresses

and.."

Instead of using 'She' for replacing 'Cinderella' in the previous sentence, the

student wrote 'He' which is malformation for the antecedent in terms of gender.

However, when considered the other pieces written by the same student, no other such

errors were found.

Error Type 05: Number Error

The errors of number in Year 1 narratives occasionally occurred; while the

occurrence went gradually down in Year 2. This error type almost faded away in Year

3 and Year 4 students. The following are examples occurring in contexts from the

student writings; the possible antecedents are underlined.

Little Red Riding Hood, written by Year 4 student#25:

"...the little girl enjoy picking flowers. <3PSF_S>She</3PSF_S> knew

that grandma like <MF-NUM/ER_3PP_O>them#it</3PP_O>."

The antecedent is 'flowers' but the student replaced the term with 'it'. The

correct form is 'them' for substituting a noun in plural form.

Snow White and the Seven Dwarfs, written by Year 3 student#23:

"...After work the seven dwarfs came home happily because Snow White
did   bed   and   cooked   for   <MF-NUM/ER_3PSM_O>   him#them
</3PSM_O> every day... "

The student replaced 'the seven dwarfs' with a singular third person 'him' instead of using 'them' which is the correct form. However, only once did the error occur in the student's narratives.

Cinderella, written by Year 3 student#06:

"..the palace servants let all ladies to try the shoe but no one could ware <MF-NUM/ER_3PS_O>it#them</3PS_O> perfectly fit. Until <3PP_S>they </3PP_S> arrived at Cinderella's place. She opened... "

As seen in the excerpt, the student replaced a singular noun 'the shoe' with a plural third person 'them' which is incorrect. She should have used 'it' for the shoe. The student, however, no longer produced such errors in the same and the other narratives.

These three error types were seldom produced by the students in Year 1 and Year 2; but they were rarely produced by the students in Year 3 and Year 4.

(d) Mis-coreference commonly found in the student writings:

Providing the possible antecedents are remote from the pronouns, mis-coreference will occur in the contexts. This error type commonly occurred in the student writing in the present study. Different groups produced similar mis-coreference cases. For examples:

Little Red Riding Hood, written by Year 1 student#06

"(1) Once upon a time, there was a little girl. (2) Her name was Little Red Riding Hood. (3) <Co/ER_3PP_S> People#They </3PP_S> called <MF-C/ER_3PSF_O> her#she </3PSF_O> that,

because <3PSF_S> she</3PSF_S> always went out with the red cape that her grandmother had sewn for <3PSF_O> her</3PSF_O>."

The excerpt consists of three sentences. The antecedent of the pronoun '*they*' in sentence (3) cannot be figured out in the previous sentence. Plus, to correct the error a common noun should have been placed to make the sentence clearer.

Little Red Riding Hood, written by Year 3 student#12

"(1) Because of the wolf's snore was very loud, <3PS_S>it</3PS_S> made a hunter who always comes to have conversation with the grandmother knew that there might be something wrong with his friend. (2) <3PP_S>They</3PP_S> are very good friends. (3) <3PSM_S>He<>/3PSM_S> decided to enter the cottage and saw <Co/ER_3PS_S>NP#it</3PS_S> is sleeping on his friend's bed. (4) <3PSM_S>He<>/3PSM_S> used a scissor to cut the wolf and could help <Co/ER_3PP_O>NP#them</3PP_O>."

It can be seen in the excerpt that the student used a pronoun '*it*' in sentence (3) to refer to '*the wolf*' in sentence (1), which is rather remote due to the length of the sentence. To make the writing clearer, a definite noun phrase might be more appropriate here. Also, the student unexpectedly placed a pronoun '*them*' in sentence (4) possibly referring to '*Little Red Riding Hood and her grandma*'. The sentence looks confusing since the antecedent cannot be anticipated. The better way to repair this error would have been for, the student to have used a definite noun phrase.

Snow White and the Seven Dwarfs, written by Year 2 student#16

"(1) Once upon a time ago in a great castle. (2) Snow White is princess. King married new queen. (3) Snow White is a body that <A/ER_3PSF_S>*#she</3PSF_S> looks very beautiful girl. (4)

<3PSF_S>She</3PSF_S> like singing and dancing because <*/ER_3PP_S>they#*</3PP_S>help her always cheerful. (5) Until <Co/ER_3PSF_S>NP#she</3PSF_S> jealous because <Co/ER_3PSF_S>NP#she</3PSF_S> very beautiful. (6) <Co/ER_3PSF_S>NP#She</3PSF_S> was furious and, wild with jealousy, began plotting…"

Once mis-coreference occurs, other mis-coreferences can be continued in the subsequent sentences since the possible antecedent might be ambiguous. The student mentioned the antecedent '*new queen*' at the very beginning of the story and she placed '*she*' in sentence (5). With no clues, it is too tricky to figure out where the possible antecedent is. The definite noun phrase, possibly '*the new queen*' or '*the stepmother*', should have been placed there for better understanding of the readers.

Snow White and the Seven Dwarfs, written by Year 4 student#09

"(1) Long time ago, Snow White was the princess who was very beautiful. (2) Her skin was white as snow and her lips were red like blood. (3) Later the queen passed away. <Co/ER_3PSM_S> NP#He</3PSM_S> married again with a witch. <3PSF_S>She</3PSF_S> hated the girl because of her beauty. (4) Every time <3PSF_S>she</3PSF_S> asked a magic mirror, <3PS_S>it</3PS_S> would say the girl was the most beautiful. (5) So <Co/ER_3PP_S>NP#they</3PP_S> planned to get rid of her."

The student, without prior notice, used '*He*' in sentence (3) to refer to 'the king'. In addition, the student placed '*they*' in sentence (5) but its antecedent is not possible to be found in the previous sentence. The possible correction is to replace

the pronoun with a definite noun phrase if it is old information in the story or a common now if it is new information of the story.

Cinderella, written by Year 1 student#36

"(1) Once upon a time Cinderella <A/ER_3PSF_S> *#she</3PSF_S> had one step mother and two step sisters. (2) <3PP_S>They</3PP_S> were very cruel and sometimes <3PP_S> they</3PP_S> ordered <3PSF_O>her</3PSF_O> to work all day long. (3) <3PSF_S>She</3PSF_S> was pretty woman. (4) <3PSF_S>She</3PSF_S> cooked and cleaned for <Co/ER_3PSF_O> NP#her</3PSF_O> everyday. (5) But <Co/ER_3PP_S> NP#they</3PP_S> did not good to Cinderella. (6) <Co/ER_3PP_S> NP#They</3PP_S> gave <3PSF_O> her </3PSF_O> old and ragged dresses and dirty shoes. (7) <Co/ER_3PSF_S> NP#She </3PSF_S> did not let her went out except when go to market. "

The excerpt showed that the student replaced '*her*' in sentence (4); its antecedent is unpredictable from the context. Plus, the student writer mixed up the story with pronouns '*they*' in sentences (5) and (6) and '*she*' in sentence (7); however, the possible antecedents are rather unclear. To make the sentences clearer, the student should have placed a definite noun phrase or a common noun.

Cinderella, written by Year 4 student#

" (1) A couple had a little baby whose name was Cinderella. (2) The girl was very pretty and generous. (3) Later, her mother passed away and her father married to a beautiful widow who had got two daughters. (4) Still, <3PP_S>they</3PP_S> had been together happily until <3PSM_S>he</3PSM_S> passed away. (5) <Co/ER_3PSF_S> NP#She </3PSF_S> jealous because Cinderella

was very attractive so <Co/ER_3PP_S> NP#they </3PP_S> did not

take care of <3PSF_O>her</3PSF_O> well. "

Without prior notice, it is difficult to figure out the antecedent of the pronoun

'she' in sentence (5). A definite noun phrase or a noun phrase will be suggested here.

Also, a pronoun 'they' in the second clause of sentence (5) could be better

understandable if replaced with a definite noun phrase or a noun phrase.

Many of the errors found in the student writings in the present study were mis-

coreference because of run-on ambiguity of the pronouns used. Once the pronoun is

not clear which antecedent it refers to, continued errors are likely after that. A

definite noun phrase is essential if it is old information; a common noun is needed if it

is new information of the story. However, possible corrections are optional.

Repeated proper noun might be a good choice in some situations.

4.2.3 Conclusions

Briefly, in the student writings, errors were found most frequently in Year 1

students' narratives. Error type 06, mis-coreference, were produced the most among

six types or pronoun errors and occurred at rather similar rates among the four groups.

The comparative results usually revealed significantly lower means of errors in Year 3

and Year 4 writing than the others. The two groups had no significant differences

between each other. Qualitative results revealed more insightful perspectives of how

students produced errors and what differences among groups there were. Firstly,

ER01 (omission) errors were found in both subject and object cases, but not in object

of preposition. Next, ER02 (overuse) errors were found more in Year 1 and more

often in relative clauses. Thirdly, malformation consisting of ER03 (case error),

ER04 (gender error), and ER05 (number error) showed the errors might be the result

of careless writing, particularly produced by Year 3 and Year 3, since they did not re-occur. Lastly, ER06 (mis-coreference) errors were found often in all groups. Once mis-coreference occurs, ambiguity will be run-on, and more errors are likely to occur.

4.3 Results for Research Question 3: *What developmental patterns of pronoun acquisition can be seen to occur over time?*

Some developmental trends are partially presented in the previous section; however, to give emphasis of the trends the whole picture of pronoun developmental trends will be presented in a form of graphs to show changing of the features in focus over time. To contextualize the word 'over time' based on cross-sectional research method employed in the present study, results of pronouns produced by four groups of university students in different years will be assumed as the results due to time exposure to English. Accordingly, the results from each year will be viewed in timeline pictures, from Year 1 to Year 4. Total information of tokens, pronouns, and pronoun errors in each type will be included in the report. In addition, text analysis results will be attached.

The following figures are information of tokens in narratives, pronouns, and pronoun errors produced by the students in four groups. The six types of pronoun errors are omission (ER01), overuse (ER02), case error (ER03), gender error (ER04), number error (ER05), and mis-coreference (ER06).

Total Tokens &Total Pronouns

[chart: Total Tokens & Total Pronouns over Year 1–Year 4]
TT Token values: 908.26, 946.95, 1017.29, 1123.12
Total PRO values: 99.56, 92.9, 91.29, 98.18
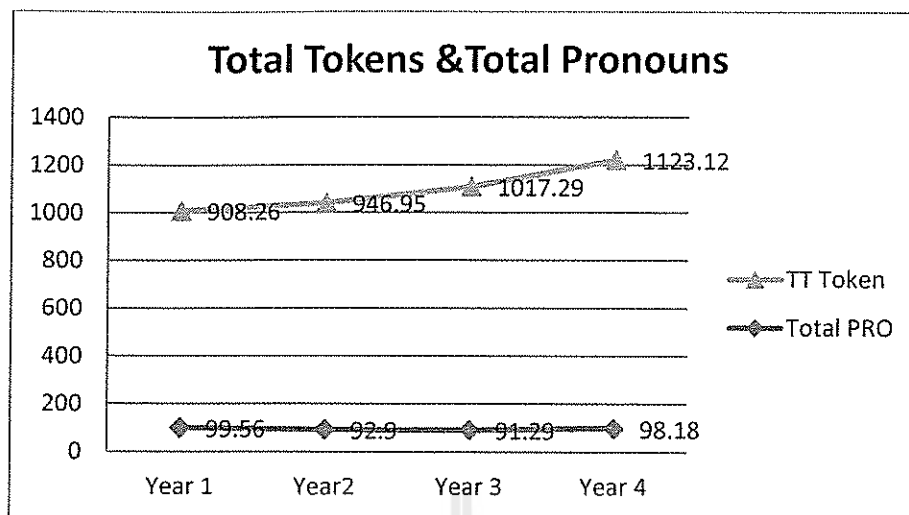
Figure 4.1 Total tokens and total pronouns produced in the narratives over time

According to the trends in figure 4.1, the trend of total tokens in the narratives moderately boosts, while the trend of pronouns produced in the narratives is rather steady from Year 1 to Year 4. In other words, as time passed the students could write more words to narrate stories; however, their pronoun use does not grow together with the token trend.

Total Pronouns & Total pronoun errors

[chart: Total Pronouns & Total pronoun errors over Year 1–Year 4]
Total PRO values: 99.56, 92.9, 91.29, 98.18
Total Error values: 38.05, 32.72, 24.76, 22.48
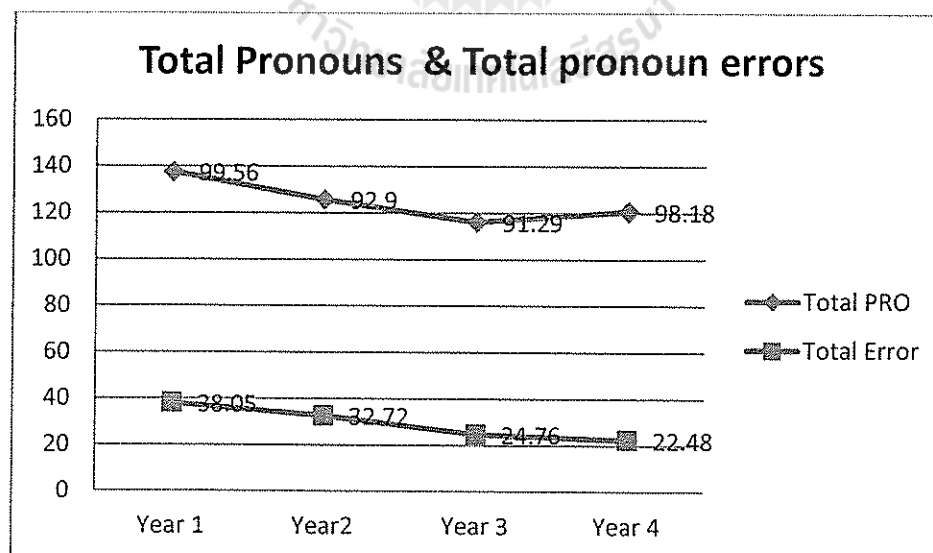
Figure 4.2 Total pronouns and total pronoun errors in the narratives over time

While pronouns had been relatively steady over time, totally pronoun errors had decreased. However, in details pronoun error results revealed different pictures in some ways.
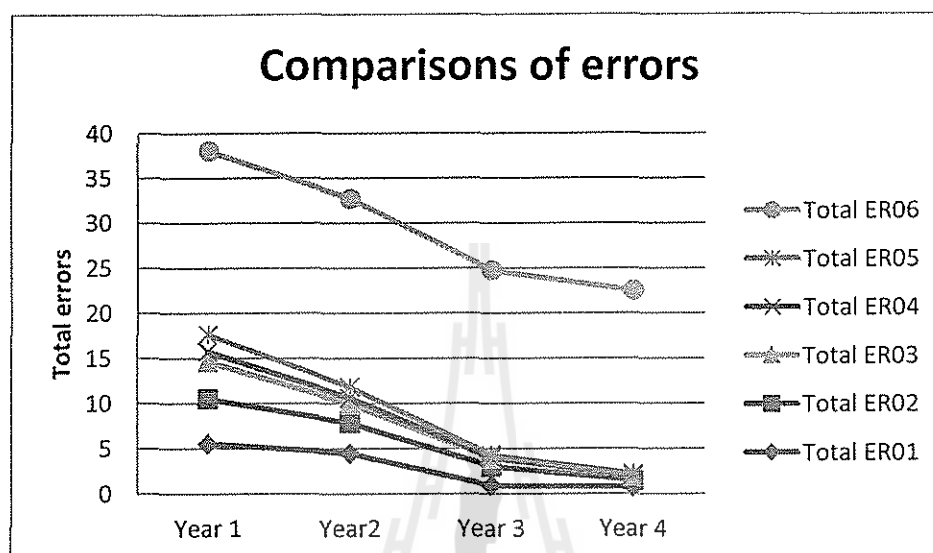
Figure 4.3 Six types of pronoun errors in the narratives over time

Figure 4.3 informs that pronoun error types 01 to type 05 start relatively low and over the period of four year they gradually fall down. It might be implied that the students cannot overcome the error type 06 or misco-reference in contexts.

To give more pictures of how pronouns and pronoun errors occurred over time, characteristics of pronouns used by the students over time will be described below.

Characteristics of Pronouns Used by the Students over Time

As mentioned in the results of the research questions (1) and (2) that the students used fewer pronouns over time; 5 error types decreased while error type 6 or mis-coreference seemed steadily exist over time. When pronouns decreased, other pronominal forms were expressed. The followings are examples or the pronominal used due to saliency of each story.

Table 4.25 Examples of pronominals used by the students over time

Noun phrase	Pronominals			
	Year 1	Year 2	Year 3	Year 4
Little Red Riding Hood	1) she 2) Little Red Riding Hood 3) Ø	1) she 2) Little Red Riding Hood 3) Ø	1) she 2) Little Red Riding Hood 3) The little girl 4) The girl 5) The innocent girl	1) she 2) Little Red Riding Hood 3) The little girl 4) The girl 5) This girl 6) The poor girl
Snow White	1) She 2) Snow White 3) Ø 4) The pretty girl	1) She 2) Snow White 3) Ø 4) The girl	1) She 2) Snow White 3) The pretty girl 4) The girl 5) The pretty princess 6) The cute lady	1) She 2) Snow White 3) The lovely girl 4) The poor lady 5) The beautiful lady 6) The pretty and nice lady
Cinderella	1) She 2) Cinderella 3) Ø 4) The lady	1) She 2) Cinderella 3) Ø 4) The beautiful girl	1) She 2) Cinderella 3) The pretty lady 4) The beautiful girl 5) The poor girl 6) The girl wearing rugged	1) She 2) Cinderella 3) The poor lady 4) The generous girl 5) The lovely girl 6) The girl who worked hard
A wolf	1) It	1) It	1) It 2) The wolf 3) The evil 4) The bad wolf	1) It 2) The wolf 3) The cruel wolf 4) The bad wolf

The students tended to use fewer pronouns to substitute noun phrases over time. At the beginning they tended to use pronouns and full names; during Year 3 to Year 4 they replaced the noun placed with other forms of pronominals such as short definite description (e.g., the girl, the lady, the evil, the wolf); long definite descriptions (e.g., the poor girl, the pretty lady, the beautiful girl, the generous girl, the lovely girl, and the cruel wolf); and a noun phrase with modifier (i.e., the girl wearing rugged)

This seems good news based on more variety of pronominals used in the students' writings; however, when looked through the pronoun errors, some mis-coreference pronoun errors occurred still. The whole picture of how pronouns and pronoun errors produced by Thai university majors over time.

Characteristics of Pronoun Development Over Time

Year 1	-Frequent pronouns used for taking the places of noun phrases/proper names -Some omissions, overuses and case errors -A few of gender and number errors -Mis-coreference = approximately 20% of overt pronouns

Year 2	-Occasional pronouns used for taking the places of noun phrases/proper names -A few omissions and overuses -Very few malformation - Mis-coreference = approximately 20% of overt pronouns

| Year 3 | -Optionally pronouns used for taking the places of noun phrases/proper names

-A few omissions

-Almost zero of overuses and malformation

- Mis-coreference = approximately 20% of overt pronouns |

| Year 4 | -Optionally pronouns used for taking the places of noun phrases/proper names

- Almost zero of omissions, overuses, and malformation

- Mis-coreference = approximately 20% of overt pronouns |

The descriptions of pronoun used by students in each group imply better use of pronouns in terms of form (case, gender, and number) and some surface rules (non-pro drop and no overuse of pronouns). However, pronouns used in contexts need to be reviewed and redesigned for better effective teaching materials and methods to reduce the error of pronoun mis-coreference.

In summary, the results in this section are rather simple when Year 4 could produce the highest pronoun accuracy and the least pronoun errors in overall. Even though Year 2 students did the lowest mean score of pronoun accuracy, Year 1 students did the highest mean score of errors. The trend is likely to be that the pronouns produced by the students in the narratives were better over time. However, in details results of errors show differently somehow. While Error types 01 to 05 tended to be down over time, error type 06 was still almost the same. This implies that Thai EFL learners have got a problem of pronoun co-reference when writing.

4.4 Summary of the Chapter

In conclusion, the chapter provides information in three parts to respond the research questions. First, information of how pronouns in subject, object, and possessive cases were produced in narratives by Thai university English majors. In addition, results of pronoun accuracy and errors are included. The results showed the highest pronoun produced in the narratives by Year 1 group. Year 4 group did the highest mean of pronoun accuracy while Year 2 group did the lowest one. When compared, Year 4 produced the highest tokens but the lowest percentage of pronouns in their writings; while Year 1 produced the lowest tokens but they used the highest pronouns. Text analysis reveals how the students form four groups used pronouns differently. Year 1 and Year 2 groups tended to used pronominal forms of personal pronouns and proper noun when repeating old information in contexts. However, Year 3 and Year 4 groups tended to used more variety of pronominal forms such as personal pronouns, proper nouns, and definite noun phrases. Additionally, pronoun accuracy information is included in this section. The results showed Year 2 did the lowest mean score of pronoun accuracy. Based on the accuracy rate, Year 2 should have produced the highest errors; Year 1 did the lowest mean score of errors since they produced the highest omissions, tagged later. Details of errors will be presented next in order to complete the picture of how pronouns are actually used.

Second, information of error differences among the four groups reveals results of six error types (ER01 = Omission, ER02 = Overuse, ER03 = Case error, ER04 = Gender error, ER05 = Number error, and ER06 = Mis-coreference). As a whole, Year 1 did the poorest results while Year 4 did the best among the groups. However, comparative results in details exposed different outcomes in some ways. Year 3 and

Year 4 did no significantly different means of errors in types 01, 02, 03, 04, and 05. In other words, they produced very similar and low means of the errors in these five types. In particular, error types 04 and 05 were below 1.0 produce by the two groups on the average. On the other hand, Year 1 and Year 2 did rather higher means of the five errors. Lastly, ER06 resulted that all four groups could not figure out when they appropriately use pronouns in contexts.

Finally, information of pronoun development over time elaborates developmental trends of pronouns among the four groups according to their exposure time to the target language in the university. Totally, the trends seem simple since the developmental lines follow the exposure time to English. In other words, the students could do better pronouns and do lower errors over time. However, thoroughly considered, the trend of error type 06 resulted differently. No progress occurs in terms of mis-coreference over time.

The present chapter includes information of what happened in pronouns produced by Thai university English majors both in quantitative and qualitative perspectives; however, the information of why the happening was so has not been yet included. It will be further discussed in the next chapter before going to conclusions and implications of the study

CHAPTER 5

DISCUSSIONS

This chapter mainly includes discussions of the results. There are explanations in accordance with the results occurring and reported in the previous chapter. The explanations are generally based on the research hypotheses that the culprits related to the pronoun used by Thai university English majors could be *first language transfer, overgeneralization, ineffective teaching materials and methods*, as well as *ignorance of restriction, incomplete application of rule* and *false concept hypothesized*.

If the explanations go along well with anticipated results, the hypotheses are retained. On the other hand, in case the results cannot be clarified by the anticipations, the hypotheses are rejected and other possible views will be taken in.

5.1 Discussions

The discussions of pronoun results are separated into three parts: pronouns, pronoun errors, and language development of pronouns. Based on the research hypothesis, first language transfer might be the main interference of pronoun used by the Thai university English majors. However, other potential causes can be related in some points of view.

5.1.1 Discussions of Pronouns Produced in the Study

Two issues underlined in the results that (a) more pronouns used by Year 1 and Year 2 students than ones used by Year 3 and Year 4 students; and (b) all pronouns in object of preposition were not omitted.

(a) Frequency of Pronouns Used in the Student Writings

The results revealed different pronoun use among the four groups of English majors. The Thai EFL learners with shorter time exposure to English tended to more frequently used pronouns in their narratives than ones with longer time exposure to the target language. In addition, text analysis showed that the subjects in Year 1 and Year 2 groups usually used *pronouns* to replace noun phrases or proper names they initially mentioned.

At the same time, Year 3 and Year 4 students used more pronominal forms to replace noun phrases or proper names initially mentioned in the stories. They interchangeably used *pronouns, proper names, and definite noun phrases* in the narratives. The high number of pronoun used among the beginner and pre-intermediate students showed limited choice of referent in communication; while lower number of pronouns with other pronominal forms used in contexts by the students in Year 3 and Year 4 showed more choices including 'pronouns' as one of cohesive devices in the discourse.

To maintain cohesion across units in which the referent occurs, senders (writers or speakers) attempt to coordinate with their receivers (readers or interlocutors) to produce a coherent discourse (e.g., Clark, 1996; Grosz et al., 1995). People do not produce strings of unrelated sentences. Rather, they tend to talk about

the same things for extended periods of time. Consequently, people refer more often to referents that have been recently mentioned than to referents that have not.

To explain why pronouns were differently used between the student writings is possibly explained with Referent Hierarchy consisting of a list of referent expressions in accordance with the topic status of the referent. The most two applicable lists proposed in discourse hypotheses are Givon's Gradation of Referring Expressions (1980) and Areil's Accessibility Hierarchy (1990). The others e.g. Chafe (1994) and Gundel et al. (1993) are similar to Ariel's Accessibility Hierarchy, but are not as comprehensive.

'*The Accessibility Hierarchy*' proposed by Areil (e.g. 1988; 1990; 1994), one of the most comprehensive in proposals of referent will be used for for discussing the results of the present study. Areil suggested different distribution patterns for different forms of reference, which she termed "accessibility markers" as shown in Chapter 2. Her full Accessibility Marking Scale is presented in the figure 5.1 below with some examples of English accessibility markers partially retrieved from the narratives in this study.

Low	Marking Scale	Examples
	(1) Full name + modifier	-Cinderella, the pretty lady
	(2) Full ('namy') name	-Ella Charlie
	(3) Long definite description	-The pretty and gorgeous lady
	(4) Short definite description	-The lady
	(5) Last name	-Charlie
	(6) First name	-Cinderella
	(7) Distal demonstrative + modifier	-that shoe she forgot last night
	(8) Proximal demonstrative + modifier	-this shoe she forgot last night
	(9) Distal demonstrative + NP	-that house
	(10) Proximate demonstrative +NP	-this house
	(11) Distal demonstrative	-that
	(12) Proximate demonstrative	-this
	(13) Stressed pronoun + gesture	-SHE (plus gesture)
	(14) Stressed pronoun	-SHE
	(15) Unstressed pronoun	-she
	(16) Cliticized pronoun	-(no examples in English)
	(17) Extremely High Accessibility Marker	- Gaps, including pro, PRO and *wh*

Figure 5.1: Examples of Pronouns Occurring in Association with Areil's

Accessibility Hierarchy (1990: 73)

Both Givon's and Areil's suggested the lists of referent forms which are chosen in accordance with the sender's evaluation of accessible degree of the referent from low to high accessibility. In other words, the sender will consider which referent and when to use it in communication; accessibility degree means how much the referent continues understanding to the receivers. These lists are a lot in common except that Givon's included indefinite NPs, a form frequently used in initial mention but was left aside by Areil.

As the review in chapter 2, other scholars have proposed similar scales of accessibility, such as Chafe (1994) or Gundel et al. (1993). These scales are similar to Ariel's Accessibility Hierarchy, but are not as comprehensive. Therefore, to give explain why the subjects, between Year 1 and 2 & Year 3 and 4, in the present study used pronoun differently, Ariel's Accessibility Hierarchy will be employed.

It means that Year 1 and Year 2 students might have considered 'pronouns' as the easiest for readers to understand. However, Year 3 and Year 4 students probably considered more choices of referents to continue their narratives: pronouns (item 16), first names (item 6), and short definite descriptions (item 4) to use in the narratives though items (6) and (4) are graded as low accessibility degree in the scale. Only is the scale, however, not able to fully explain why the students used pronouns in different ways. Other factors, not just evaluation of accessibility degree, will affect the students' decision making when expressing referents including pronouns.

Areil (1990) proposed four factors related to Accessibility status of the referent.

a) Distance:	The distance between the antecedent and the anaphor (relevant to subsequent mentions only)
b) Competition:	The number of competitors on the role of antecedent.
c) Saliency:	The antecedent being a salient referent, mainly whether it is a topic or a non-topic.
d) Unity:	The antecedent being within vs. without the same frame/ world/point of view/segment or paragraph as the anaphor.

(Areil, 1990: p. 28)

The first two factors are about distance of the possible referent in contexts. Statistical evidence in written texts investigated by Areil confirmed that the role of accessibility in determining referential choice. With a breakdown of anaphoric expressions along with the distance of the distance of the reference from its antecedence, pronouns (High Accessibility Markers) were found to focus on short distances (in same sentence and previous sentence), demonstratives (Intermediate Accessibility Markers) were found to focus on intermediate distances (in previous sentences and the same paragraph), but definite noun phrases (Low Accessibility Markers) were found to focus on long distances (in the same paragraph or across paragraph).

Ariel's third factor, "Saliency", addresses the difference between topical and non-topical antecedents. Her discussion reflects an assumption that "topic" is defined in terms of the grammatical subject. She also mentioned Levy's (1982) claim that topicality is influenced by the number of anaphoric references to an entity, in particular pronominal references. Although the Accessibility hierarchy is inherently a graded scale, it appears that "topic" is treated as an all-or-nothing phenomenon, implying that an entity either is the topic or not, and that a given discourse segment has one and only one topic.

The fourth area that Ariel lists is that of "Unity". This factor reflects the effect that discourse structure can have on "Working Memory", and thus reference form. Ariel suggests that choices in reference form are influenced by the discourse structure, which can be influenced by things like the passage of time within the discourse or paragraph breaks in written text. She links this factor to Fox's (1987) claim that in English, 'by using a pronoun the speaker displays an understanding that the preceding

sequence has not been closed down' (1987:18). Thus, pronouns are more natural for references to things from the same discourse segment, and fuller forms are used when the referent was last mentioned in a different segment.

These factors might have been considered when the students in the present study decided to express pronouns in their writings. Pronouns produced by the students in Year 1 and Year 2 might have been more affected by Saliency than Distance and Competition expressed previously, since they tended to use pronouns along the narratives no matter how remote the initial noun phrase or proper name mentioned, for examples:

> "Cinderella's parents died after that **she** with stepmother
> and her daughter. **They** ordered **her** do housework and work
> such as cleaning and washing dishes. So sad and very pity. **She**
> does not have friends. **She** played with birds and small animals in
> kitchen. While **they** are very happy. **She** told **her** to take care of
> **them**. **She** cooked food and wore rugged clothes. **They** treated
> **she** very bad. Tell **her** to work, iron, clean room...."

> (Cinderella, original texts by Year1 student#24)

On one occasion in the excerpt, the student in Year 1 mentioned '*Cinderella*' and later the noun phrase is replaced with '*she*' and '*her*'; however, there were some referent candidates in between, '*step mother*' and '*her daughter*'. This might make ambiguity to readers. After the student talked about 'stepmother and her mother', she might have replaced the noun phrase with pronouns 'they' and 'she'. The student usually used personal pronouns for substituting any mentioned noun phrases no matter how far the referents are.

"Once upon a time Cinderella was a girl who **she** lived with her parents. After her mother died. **She** was sad and her father married again with a bad lady who **she** had two children. **They** did not like **her**. Then **he** died too. **They** let **her** do everything in house. **She** was sad because **she** must to live in a kitchen. And **she** wore old dress but **they** wore new and beautiful dresses. **She** must to work all day but **they** were convenient...."

(Cinderella, original texts by Year2 student#08)

Once the student brought up a proper name '*Cinderella*' as new information in the story, she replaced the proper name with '*she*' along the paragraph which might be fine if there were not interferences in between. However, in the student writing she also mentioned '*a bad lady*', the antecedent could be ambiguous. Though the proper name was the head in terms of Salience, the Unity of the story seemed ignored due to other pronouns expressed in between.

However, some examples from students in Year 3 and Year 4 show different paths of pronoun use. In other words, the students tended to use more forms of pronominals instead of solely using pronoun. The examples below are excerpted from performance of the students in Year 3 and Year 4.

"Once upon a time, parents had a child. **They** named **the girl** as Cinderella. **She** was very pretty and attractive. **They** passed away and **she** had to live with cruel **stepmother and two step sisters**. **The step sisters** are very mean to **her**. **She** forced her to work all day so she had to stay in the kitchen. **The poor girl** grew up with suffer but **she** was very pretty and kind. Because of this **they** hated **her** more and more. **She** told **her** to take care of **her daughters** and **they** tried to haze **Cinderella**..."

(Cinderella, original texts by Year3 student#08)

From the excerpt above, the student in Year 3 expressed some more variety of pronominals if compared to the earlier ones in Year 1 and Year 2. She used pronouns, long definite description, and short definite description, and first name to replace the mentioned noun phrases. For examples, the student used a pronoun '*they*' for replacing a noun phrase '*parents*'; used '*she*', '*her*', '*the poor girl*', and '*Cinderella*' for replacing '*Cinderella*' which was mentioned at the beginning of the narrative; and used '*the step sisters*', '*they*', and '*her daughters*' for replacing '*two step sisters*'.

In addition, the students in Year 4 tended to take Distance into considerations when writing. In the excerpt below, the student initially mentioned '*Cinderella*' and repeated the proper name, instead of using a pronoun '*she*' which though in the highest degree of accessibility, when the referent was remote with some competitions in between. In addition, the unity of the sentences in between did not focus only on the same person, *Cinderella*.

> "Once upon a time in a house, **Cinderella** lived unhappy. Her mother was dead. Her father had married another woman, a widow with two daughters. And **they** didn't like her. **Cinderella** was lovely and nice but **she** dressed was ragged and old. Her face was dusty gray from coals. While **the stepsisters** wear dresses splendid and elegant their clothes, **they** are still ugly. **They** lived in the house comfortably, but **Cinderella** was not. **The young girl** worked hard all the day. **She** lived alone..."

(Cinderella, original texts by Year4 student#14)

Although a definite noun phrase '*the young girl*' was evaluated in Low Accessibility Marker, it was expressed right after '*Cinderella*' was mentioned in the previous sentence. However, the referent expression was not ambiguous according to the

single competition and unity in the former sentence. Based on the factors employed when expressing referents, Year 3 and Year 4 students were likely to express more referent forms with various factors taken into considerations. However, no specific patterns can be found in each group; the concluding of the students' pronoun use can be that the beginners, Year 1 and Year 2, tended to use personal pronoun to substitute noun phrases mentioned with little consideration of referent accessibility concerns, while the others with longer exposure to L2, Year 3 and Year 4, tended to use more variety of pronominal choices. The later ones might have considered more about referent accessibility of the readers; they, therefore, used better understandable pronominals to refer noun phrases, ones in long distance position in particular.

The results of the present study are not really supportive for the previous studies (Aikawa, 1991; Kang, 2004; Kim, 1994; Nogushi, 1997; and Siewierska, 2004;) employing Areil's Accessibility Marking Scale of referents since the studies normally investigated pronoun resolution or referring expressions as discourse device in both spoken and written communication. Such studies explored how readers interpret referents including pronouns in contexts. However, backtracking of how and why the senders make decision when expressing referent might be meaningful to explain why different referent occurred in the student writings.

In conclusions, the students with different time exposure to English differently expressed referents including pronouns in the narratives possibly due to their evaluation of accessibility degree of the referents with issues affected to the expressions. According to Ariel's Accessibility Scale and factors (distance, competition, salience, and unity) influencing the referent status in the student writings, the students with less English exposure tended to have limited choices when

evaluating referents according to easiness to understand for the readers; while ones with more exposure to English tended to better evaluate degrees of accessibility. Pronouns and other referents expressed in the narratives by the students in Year 3 and 4, therefore, sounded more comprehensible for readers than ones expressed by the students in Year 1 and Year 2. This information should be suggested to instructors of English to cope with the potential problems for making effective referent choices in communication; however, error information with explanations must be included.

5.1.2 Discussions of Errors Occurring in the Study

Apart from the explanations of pronouns, explanations of error information are also in focus. In accordance with the research hypothesis that first language transfer possibly is the main interference; however, degrees of interference to all four groups when producing pronouns in writing are rather different. Three main errors possibly related to first language transfer are *omission, overuse, malformation (case, gender, and number errors)* and *mis-coreference.* Additionally, some other factors like overgeneralization, ineffective teaching methods and materials, and others might also influence the students' pronoun use in the study.

5.1.2.1 First Language Transfer Influenced Pronoun Erros

In chapter 2, the review of English and Thai pronoun forms and use showed some differences which might be difficult for Thai learners of English to acquire English pronouns rules; the learners might occupy first language rules of pronouns when expressing in English. First language transfer previously reported in the literature review is probably related to *omission, overuse* and *mis-coreference.* The transfer really plays important roles related to errors produced by Year 1 students

in particular. Later, it faded as shown in decreasing such errors produced by students with longer time exposure to English in university.

Omission:

Due to the previous studies (Cambell, 1969; Hatton, 1987; Hoonchamlong, 1991) of Thai pronouns, pro-drop or zero pronouns is acceptable in both spoken and written language. The other linguists (Aroonmanakun, 2000; Muansuwan, 2001; Pingkarawat, 1989) investigated Thai pro-drop characteristics and their findings support the previous study that Thai pro-drop can occur whenever senders and receivers are aware of the 'actors' and 'goals' involved in the discourse. However, the omission in Thai will be avoided when there is uncertainty or ambiguity in the choice of referent.

Due to first language transfer, pro-drop feature in Thai, more pronoun omissions occurred in Year 1 students' narratives in English. Later, the errors produced by Year 2 were significantly falling; first language transfer decreasingly influenced the students' pronoun use. Moreover, the errors produced by Year 3 and Year 4 were even down under 1.0. The examples below are *omissions* in subject and subject positions produced by Year 1 and Year 2 students:

... carried a food basket. <*/ER_3PSF_S>She#*</3PSF_S> very happy..

....ready to go. And then <*/ER_3PSF_S>She#*</3PSF_S> get a food…

....the flowers so pretty. <*/ER_3PSF_S>She#*</3PSF_S> picked some..

..Knock! Knock! Knock! <*/ER_3PS_S>it#*</3PS_S> called loudly "Hi!.

Omissions in object position:

...basket. The girl took <*/ER_3PS_O>it#*</3PS_O> to grandma's house

...grandmother and bit <*/ER_3PSF_O>her#*</3PSF_O> fiercely. And..

.... the door and found <*/ER_3PS_O>it#*</3PS_O>. Hunter and dog ...

.... did not want to tell <*/ER_3PSF_O>her#*</3PSF_O>. Then.....

The different results of omissions among the students might imply first language transfer influenced students' writings in different degrees. In other words, the results get along with the hypothesis in the way that Year 1 was the most influenced by first language while Year 3 and Year 4 were almost not influenced by the language transfer.

It is interesting that pro-drop in Thai never occurs in a case of object of preposition; none of omissions were found when the students expressed pronouns in a case of object of preposition. For examples:

'...she </3PSF_S> talked to <3PS_O>it</3PS_O> and forgot ...'

'.. field there and looked at <3PP_O>them</3PP_O> and smiled..'

'.. the witch got angry with <3PS_O>it</3PS_O> very much...'

'..the house and married to <3PSF_O>her</3PSF_O> and then..'

'..so Cinderella cooked for <3PP_O>them</3PP_O> every day...'

'..sisters wanted marry to <3PSM_O>him</3PSM_O> and ...'

Objects of preposition seem not omissible for the students; they add pronouns right after the prepositions. This truly supports that similarity between two languages makes second language learning easy since the learners do not have to adjust any rules to learn the target language.

Overuse:

As Raimes (2008) proposed that Thai does not allow 'subject/object restatement' characteristic, first language transfer should not be the main cause for overuse of pronouns. However, the feature occasionally exists in Thai informal written and spoken language. Thai EFL learners in the present study might have been affected by first language transfer when writing in English. In other words, they sometimes over expressed a pronoun right after the noun phrase or the proper name which was its antecedent, for examples:

'..new queen <A/ER_3PSF_S>*#she</3PSF_S> was a witch and very..'

The pronoun errors reoccurring in students' writing might be affected by L1. Although Thai pronoun overuse is practically allowed in some contexts (Chaimongkol, 2007), the errors or pronoun overuse in English seemed decreased over time. It is remarkable that some pronoun overuse remaining in the students' narratives frequently occurred in relative clauses, which will be discussed in section 5.1.2.2 since first language transfer might not play important role in such errors.

Malformation: Case/gender/Number

Case errors occurred in the student writings were probably influenced by first language transfer in the same way as what happened with 'omissions' explained in the previous section. As reviewed in chapter 2, there is no case marking in Thai pronouns. First language transfer possibly had an effect on case errors produced by in Year 1 when compared to the rest groups. On the other hand, the transfer did not

really affect Year 3 and Year 4 students due to almost zero case error in their narratives on the average. For examples,

..wanted to fool <MF-C/ER_3PSF_O>her#she</3PSF_O> to spend..

....went to pick <MF-C/ER_3PP_O>them#they</3PP_O> in the gar..

... queen hated <MF-C/ER_3PSF_O>her#she</3PSF_O> so much..

..pushed and hit <MF-C/ER_3PSF_O>her#she</3PSF_O> before ..

....and told that <MF-C/ER_3PSF_S>she#her</3PSF_S> do work in

The other two malformation errors: gender and number errors were not really affected by Thai case system due to the errors occurring in student writings in the present study. Certainly, the transfer affected Year 1 the most; for examples:

(1) The girl walked to her grandmother's house and she#he though.

(2) The hunter tried to followed it for along time until he#she a dog stopped..

(3) Cinderella received letter too. She#He also want to go..

In sentence (1) 'he' was used instead of 'she' for replacing the mentioned noun phrase, 'The girl'. In the same way, 'she', in sentence (2), was used instead of 'he' for replacing the stated noun phrase, 'The hunter' which was clearly male according to the picture given. In sentence (3), 'He' was used for instituting 'Cinderella'. The errors occurring in the students' writing might have been affected from the first language; since for Thai language, it is optional to use unisex pronouns like 'เขา' /kwaw/. However, over time the type of error was hardly found in the students' writings.

Finally, the last pronoun malformation, number error, occurring in the students' essays was very few. First language transfer might not have influenced these English majors as expected

However, the errors were very small number on the average. The happening might be explained that gender and number systems exist in Thai; these systems (though not exactly the same as ones in English) might help Thai learners of English to acquire gender and number systems in English with ease.

Mis-coreference

Apart from omission, overuse, and malformation of pronouns in the students' writings, mis-coreference might have been influenced by the first language; since Thai language allows omission with unclear rules. Even Thai natives are sometimes confusing with Thai pronoun usage. When there is an ambiguous pronoun referent in Thai, they have to ask for more information or clarification from the speaker as well. However, in written mode they have no choices to make it clearer. Concerning with the reader understanding among Thais might be rather low; when expressing pronouns in English writing, their first language could be interference. For example, once a noun phrase 'Snow White', later in long distance repeated mention the Year 1 student (#65) used '*she*' though it might have been ambiguous for the readers.

'...**Snow White** was a very beautiful princess and has many friends birds, butterflies and insects. *The queen* hated her because she was very beautiful. She wanted to the most beautiful around the world. Call for a hunter and began dreadful plan. A hunter waited and took

to forest. He walked into deep forest. He did not want to kill

everybody and **she** cried a lot......'

The excerpt above confirms that the student writer did not realize how much

the pronoun '*she*' was far from its antecedent '*Snow White*'; and the student might

have thought that her readers would simply understand it. In addition, the student

might not have concerned that there was another candidate, the queen, possibly the

pronoun antecedent which could make ambiguity in the narrative. In this case, it

might be implied that the student might have considered *saliency* as her evaluation

principle of pronominal expression. In other words, the student tended to use

personal pronoun as pronominal when the referent was important or in focus.

Another example excerpted from Year 4 student (#19) writing. As mentioned

earlier that the students with longer exposure to English in the university decreasingly

used pronouns as a noun phrase substitution, the students used more choices of

pronominal to replace the noun phrase 'Snow White'; however, some mis-

coreferences occurred.

> '....**Snow White** grew up with light skin as snow and her
>
> lips were red as roses. The pretty lady was beloved but *queen*
>
> hated her. She called a hunter to get rid of the girl. She wanted to
>
> be the only beautiful lady in the world. Hunter came and took **her**
>
> to big forest. They walked all day long and got tired... '

The student once wrote 'Snow white' and he switch to other pronominals, e.g. *the*

pretty lady and *her*, after that. However, when he again replaced the proper noun with

'*her*', he might not have considered that the pronoun '*her*' was relatively away from

its antecedent, '*Snow White*'. The example confirms pronoun use of these Thai University English majors that the student writers might have thought of saliency the most among the others: distance, competition, and unity. The pronoun was not only in long distant position from its antecedent, but also there was the other competition in between that was '*queen*'. Using variety of pronominal forms in writing can be more effective if the writer considers more about the audience's comprehension.

Mis-coreference occurring among the students with Thai mother tongue might be affected by first language transfer since Thais tend to consider the topic as the most important in the narrative most when using personal pronouns and omissions; whenever Thai writers express personal pronouns without clues in the previous sentence, the writers tentatively refer to the noun phrase which is a topic of the writing. However, in English writing the writer cannot focus on topic only due to the evaluations proposed by Areil (1990) that others (*distance, competition,* and *unity*) also should be taken into considerations.

In short, the results in the present study support that first language transfer plays an important role in language learning (Corder, 1981). Though, in the present study, it really influenced students in Year 1 at the beginning stage of English major, it did not much influenced students in Year 2, 3 and 4. The errors of omissions, overuse, and malformation were evidently related to the students' mother tongue. However, some of the errors occurring in overuse are neither first language nor second language characteristics; they might be influenced by overgeneralization, ineffective teaching materials and methods, and others.

5.1.2.2 Other Factors Possibly Influenced the Pronoun Errors

Even though first language transfer could be the main factor related to pronoun use of Thai university English majors in narrative writing, other factors proposed in the research hypothesis in chapter 1 might be partially related to the pronoun use and errors. As mentioned earlier, overuse of pronouns in some sentences with relative clauses could be found in the students' writings; the errors might have been caused by overgeneralization as Thai language does not allow such overuse. Also ineffective pronoun teaching materials and methods could affect the students' pronoun use. Other factors proposed in the research hypothesis such as ignorance of rule restriction, incomplete application of rules, false concept hypothesizing (suggested by Richard 1973); and carelessness and translation (Norris 1983) might not be clearly seen in the present study.

Due to overgeneralization or learning new language data in learners' mind and then generating flawed rules for their language performance based on the evidence (Corder, 1981), some overuse of pronouns found in sentences with relative clauses. For examples,

'..old lady who <A/ER_3PSF_S>*#she</3PSF_S> was thin and poor..'

'..the girl who <A/ER_3PSF_S>*#she</3PSF_S> was living with..'

'...wolf which <A/ER_3PS_S>*#it</3PS_S> was very hungry and..'

'..grandma who <A/ER_3PSF_S>*#she</3PSF_S> tailored her dress..'

'..the lady who <A/ER_3PSF_S>*#she</3PSF_S> was a bewitch...'

'...seven dwarfs who <A/ER_3PP_S>*#they</3PP_S> were small..'

'..small house which <A/ER_3PS_S>*#it</3PS_S> was in deep forest..'

'...the lady who<A/ER_3PSF_S>*#she</3PSF_S> had two daughters.'

'... a prince who <A/ER_3PSM_S>*#he</3PSM_S> was very clever..'

'...find lady who <A/ER_3PSF_S>*#she</3PSF_S> can fit the shoe...'

As mentioned above that some overuses of pronouns in the study were found in relative clauses. Hoonchamlong (1991) and Jenks (2011) proposed that always relative clauses in Thai need a complementizer 'thîi'.

[NP nákriani [RC thîi THîI khruu khuan tii]] son máak

student teacher should hit naughty very

'The student that the teacher should hit is very naughty.

(Jenks, 2011: 139)

Considered with this rule, overuse of pronouns occurring might not be really affected by language transfer but overgeneralization.

The other factor might have been related to the students' pronoun use in this study is ineffective pronoun teaching materials and methods. For example, pronoun teaching materials possibly influenced mis-coreference in some ways.

Mis-coreference

Mis-coreference was the highest errors found in the student writings in the present study. Plus, the error seems stable over time due to unfailing rate from Year 1 to Year 4. Although Thai pronoun use is truly different from English in many ways and first language transfer might affect expressing English pronouns in narratives, mis-coreference might have been influenced by other factors. Based on my simple

survey of some commercial EFL textbooks possibly used as main texts or supplementary materials for Thai learners, the contents provided in pronoun chapter usually emphasize forms, not usage in contexts. The evidences are provided in the survey list below.

Table 5.1 Commercial Textbooks on Survey

	Book Title	Pronouns	
		Forms	Usage in contexts
1	Basic English Grammar: by Azar and Hagen (2005)	✓	Ø
2	Cambridge English Advanced Grammar in use (2nd ed) : by Howings (1999)	✓	Ø
3	Focus on Advanced English C.A.E.: Grammar Practice : by Walton (1999)	✓	Ø
4	Grammar Practice for Upper Intermediate Students : by Walker and Elsworth (2000)	✓	Ø
5	Oxford Practice Grammar: by Eastwood (2001)	✓	Ø
6	Test It, Fix It! Revised Intermediate: Grammar : by Bourke and May (2012)	✓	Ø
7	The Good Grammar Book : by Walter and Swan (2001)	✓	Ø

These textbooks fully provide information of pronoun forms and surface usage in a sentence level; however, none of them provide information of pronoun interpretation or pronoun resolution in contexts or at least across sentences. Pronoun

usage in contexts, therefore, should be pointed out as well as when and how native speakers of English express pronouns in writings.

However, this is only supposition to why mis-coreference occurred in the student writings; in reality the students have gained some knowledge of how pronouns are used in contexts through some courses such as Writing Expressions 1-4 and Introduction to Discourse Analysis (See Appendix A). It is remarkable that Discourse Analysis is provided as a selective course that English majors commonly register when they are in Year 2; the students of Year 3 and Year 4 had more choices of pronominal referents in contexts. Unfortunately, errors of pronoun mis-coreference still occur in the contexts. That means the students need more information of when to express pronouns for replacing local and distance noun phrases or proper names.

In brief, some pronoun errors (pronoun omissions, overuse, malformation, and mis-coreference) occurring in the narratives were possibly affected by first language transfer. However, degree of the effect were rather different among the groups that Year 1 was likely to be the most influenced while Year 4 was likely to be the least influenced. In addition, the other possible cause related to pronoun overuse in relative clauses might be overgeneralization. Finally, mis-coreference, which was the highest error type found in the student writings, was probably influenced by ineffective teaching materials due to the exits materials providing pronoun form and surface usage. Though referent expressions including pronouns in students writing seemed more variety, the pronoun mis-coreference was present among the four groups. In other words, students of the four groups might be equally influenced by ineffective teaching materials over time. However, it might not be concluded, since it is not clear-cut whether the factors really affect pronoun errors found in the present study or

not. More investigations such as in-depth interview to both students and instructors might be essential to gain the concrete answers.

5.1.3 Discussions of Pronoun Developmental Trends

In accordance with the pronoun development trends previously presented in chapter 4 as the respond to research question (3) that over time pronouns expressed by the English majors were quickly improved during Year 1 to Year 3, however, the improvement was rather slow down until almost stable during Year 3 to Year 4.

Undoubtedly, the students were English majors, the target language they have exposed to in university level would support them better communicate in English over time. Over time, the students were assumed to develop their pronoun use as mentioned in described below.

Characteristics of Pronoun Development Over Time

Year 1	-Frequent pronouns used for taking the places of noun phrases/proper names -Some omissions, overuses and case errors -A few of gender and number errors -Mis-coreference = approximately 20% of overt pronouns

Year 2	-Occasional pronouns used for taking the places of noun phrases/proper names -A few omissions and overuses -Very few malformation - Mis-coreference = approximately 20% of overt pronouns

Year 3	-Optionally pronouns used for taking the places of noun phrases/proper names -A few omissions -Almost zero of overuses and malformation - Mis-coreference = approximately 20% of overt pronouns

Year 4	-Optionally pronouns used for taking the places of noun phrases/proper names -Almost zero of omissions, overuses, and malformation - Mis-coreference = approximately 20% of overt pronouns

The descriptions of pronoun used by students in each group imply better use of pronouns in terms of form (case, gender, and number) and some surface rules (non-pro drop and no overuse of pronouns). However, pronouns used in contexts need to be reviewed and redesigned for better effective teaching materials and methods to reduce the error of pronoun mis-coreference.

Before moving to the conclusions of the study, challenges for creating a tagged learner corpus will be presented so that the whole picture of the study can be seen.

5.1.4 Discussions of Challenges in the Computer-aided Error Analysis Study

Among existing corpora, very few error-tagged learner corpora have been investigated due to challenges in nature for creating learner corpus (Lea cock et al,

2010). The learner corpus-based studies employing Computer-aided Error Analysis (CEA) have been explored because of learner corpora projects restrictively accessible for specific users such as Cambridge Learner Corpus, one of the largest error-tagged learner corpora with more than 30 million words in size. The others two CLC projects with error-tagged learner corpora are HKUST with 30 million words in size and CLEC Corpus with 1 million words in size. Neither HKUST nor CLEC are available for outsiders. Development of CEA studies, therefore, is still on the way since each CEA project developed its own methods of detection and correction. The methods are improved in various directions; they are not truly comparable to the others. This means that the best performance of CEA have been under discussions.

The present study is described as a manually error-tagged learner corpus created under criterion suggested by Tono (2003) and Granger (2008) (See chapter 3). However, the study has got common difficulties in building learner corpus in terms of data collecting, transforming data into a corpus, and PRO/error tagging.

First, collecting learner data is certainly more challenging than one from native corpus which normally published. In addition, natural learner language is preferable in this field of study; however, to collect the data from learners is normally in class or during an examination. In such control, the learners producing language might be sick of writing. To write for three hours sounds not boring and it might affect learner corpus creation, in longitudinal research in particular. To continue students' motivation to write is essential for data collection methods.

The second difficulty encountered in the present study is transforming text data into a learner corpus since it means time-consuming and troublesome tasks. In general, learners in the present study were assigned to write on paper; then the texts

obtained must have been stored in electronic form involving typing every single word. A learner corpus was optionally included with the learner information such as age, gender, first language and university study level. The data in electronic format is manageable to store and access once they have been   electronically transformed.

Finally, PRO and PRO error tagging is not very simple since revision of the tagset designed is needed whenever an exception comes out. It was not very complicated to detect and tag pronouns occurring in the student writings in electrical format; however, PRO error tagging appeared in a different view. Detecting for pronoun omissions must have been manually managed; it came up with two issues: 1) what if omitted subjects/objects were found in fragments and 2) what if number and gender errors were overlapped with mis-coreference. Both issues were reconsidered and revisions of error descriptions were essential. The first issue was managed with redefining omissions as 'omitted subjects/objects are in account of errors when they are found in sentences or fragments with a predicate.' The fragment without a predicate for example, 'In deep forest far way from the village.' or 'On the way to go to grandmother's house.' were left aside in the present study. The later issue was dealt in the same way; mis-coreference was tagged as the guideline providing how to differentiate errors of number/gender from mis-coreference (See chapter 3, section 3.2.7). In addition, more than one tagger could be better promising for more reliability of the error annotation in the present study; remarkably, this is very time-consuming process.

At present the size of Thai learner corpus in the study is small with sub-corpus tagged and tagged errors included; the other problems for example POS annotation and copyright and mentioned in the previous studies (Nagata et al, 2011;

Rozovoskaya and Roth, 2010; Tetreault et al, 2010) were not really encountered. In the future, however, the issues will be taken into considerations when Thai learner corpus is developed.

To be brief, researchers have to be cautious when creating a learner corpus studies due to the challenging processes. First, collecting data should be in motivated settings to best eliciting the learner data with less or without learners' pressure. Second, storing and transforming learner data into electrical format takes time; however, once the data were kept they can be reuse in several ways. Lastly, parsing and error annotation schemes should be trained and more than one tagger is suggested for more reliability of tagging. Both time and team organization are very necessary for conducting learner corpus-based research.

5.2 Summary of the Chapter

This chapter provides explanations of pronouns and errors occurring in the student writings. Pronouns were produced most by Year 1 students followed by Years, 2, 3, and 4, respectively. While the students with longer time exposing to the target language, they produced less number of pronouns to substitute noun phrases, they produced more number and forms of pronominals, which could avoid repetitious pronouns in their writings. However, the Thai University English majors might not have realized how to engage the readers to pronominal comprehension by using the accessibility hierarchy principle proposed by Givon (1980) or by Ariel (1990). Although the principle suggests when expressing pronominals a writer/speaker should consider four main factors (distance, saliency, competition, and unity), the students in the present study tended to mainly consider saliency the most.

In terms of pronoun errors, the students tended to be influenced most by their first language; omissions, overuse, malformation, and mis-coreference were caused by Thai pronoun use by some ways. The students with less exposure to English tended to be influenced by L1 most; ones with more exposure to the target language were rather less affected by L1. Additionally, other factors might have been related to some pronoun errors. Overgeneralization might have influenced overuse of pronouns in sentences with relative clauses and ineffective pronoun teaching materials and methods probably affected the students' mis-coreference. It is remarkable that mis-coreference constantly exists over time. The factors related to the students' pronoun errors should be reconfirmed with further investigation to have more reliable conclusion.

Next, clarifications of pronoun development over time are included in order to give a picture of pronoun changing during four years. The students produced fewer errors of omission, overuse, and malformation; however, mis-coreference, which the most frequently occurred, was found at the same degree (20% of the total pronoun errors). The results possibly reflected effective pronoun teaching over time, but it might not be sufficient to make the students realize how English native express pronominals including pronouns. The supplementary information of pronoun use in contexts should be taken into instructors' considerations.

Lastly, challenges of computer learner corpus with computer-aided error analysis are reported before making conclusions of the study. The challenges were motivating students to write freely in three hours of the data collection; transforming text data into a learner corpus which means time-consuming and troublesome tasks;

and manual tagging requiring revising whenever the researcher found something different from the tagset designed.

In summary, the learner corpus-based study with tagged PRO/error approach is useful for language teaching and learning with its more reliable methods and better data management (store, annotation, and reusability). It can be extended for larger corpus size and with more language features in focus for further studies.

CHAPTER 6

PEDAGOGICAL IMPLICATIONS

AND CONCLUSIONS

The final chapter consists of three sections: pedagogical implications, suggestions for future studies and study conclusion. The pedagogical implications are proposed based on the previous information. Ideas for teaching methods and materials derived from the study are proposed to support pronoun learning for Thai EFL learners. Also, suggestions for further research are included.

Conclusions of the study are included to bear in mind what have been completed in the study, what the outcomes were, and why did they happen so. In addition, some difficulties of the study are additionally discussed.

6.1 Implications

Undoubtedly, the principal applications of corpus-derived learner data are in materials and syllabus design and classroom methodology (Granger, 2004). The information provided in this section will inform how to apply the results from learners into classroom application. Plus, it includes applicable ideas for the further studies.

6.1.1 Pedagogical Implications

As the discussed earlier that teaching pronouns should include both forms and usage. In addition, the usage is not only in sentence level but also across sentences and paragraphs. The results of pronouns used by the students both correct and

incorrect ones can be employed 'positive evidence' and 'negative evidence'. Since students cannot learn pronouns with only grammatical rules, they need to learn pronouns as cohesive device in communication as well.

For example, Areil's Accessibility Theory with the Accessibility Marker Scale and the four factors (distance, competition, saliency, and unity) can be taken into considerations so that the students have better choices of referent when writing. If the students are aware of 'which' and 'when' to express referents including pronouns by native speaker of English, they, hopefully, will more effectively use pronouns as noun phrase substitution when communicating. In this view, mis-coreference as the main problem in pronoun use for Thai EFL learners will be reduced.

Additionally, both positive and negative evidence are essential for the learners to know both 'Do' and 'Don't' about pronouns. Excerpts from student writings can be negative evidence to reconfirm the students what is correct and what is not. Moreover, some teaching and learning approaches go along well with materials for learner data, 'problem-based approach', for example.

Savery (2006) defined that problem-based learning (PBL) is an approach that challenges students to learn through engagement in *a real problem*. It is a format that simultaneously develops both problem solving strategies and disciplinary knowledge bases and skills by placing students in the active role of problem-solvers confronted with *an ill-structured situation* that simulates the kind of problems they are likely to face as future managers in complex organizations. In this perspective, the data-derived language learner seems perfect materials for the approach since it provides 'a real problem' with the authentic errors produced by students; the approach encourages learners to research and manage for accuracy. It is believed that the learners will

gained both knowledge and skills of language at the same time. An example of problem-based approach materials is below.

Directions: Discuss possible antecedents of the underlined pronouns in the story. Are all underlined pronouns correct or incorrect? If they are incorrect, explain the reasons and revise the incorrect ones.

"Once upon a time, Cinderella was her name. Cinderella she used to have mother but her died. Father married new wife. They had two daughters. After that she killed him and took all money and house to own. They gave her little food and ragged clothes. And she must to worked for all everything in house for them. They lived like millionaire but Cinderella she lived like slave.

One day everyone got invitation letter to a party from king. He announce if lady she come to a party and dace with the prince. He will marry to her. They were very excited about the party and they prepared the best dresses for the night. But she did not have beautiful dress to go. She sobbed and cried until fairy angel came. She said...."

(An excerpt from UBU student, Year 1)

The example above is a possible group work for pronoun learning. Groups of students will look through the story, discuss if the underlined pronoun correct or incorrect based on native use. Later, the students give reasons why the pronouns are correct or incorrect and the revised version of each will be proposed by the group. The instructor has to make sure that the students have enough resource of pronoun information; in this case, the instructor might have to prepare a pile of references for the class so that the students can direct themselves to learn how pronouns are used.

The learner writings could be great resource for other learners with the same mother tongue. The pronoun errors, how they occur, and how revise them in a manner will become in focus. From the self-directed learning activity, the students are expectedly successful when producing pronouns in contexts.

In addition, 'Accessibility Marking Scale' and factors should be supplemented in class materials. For examples,

---

Little Red Riding Hood

(1) Once upon a time, there was <u>a little girl, who lived in a village near the forest</u>. Whenever *she*_iwent out, *the little girl*_i wore a red riding cloak, so everyone in the village called *her*_i *Little Red Riding Hood*_i.

(2) One morning, *Little Red Riding Hood*_i asked <u>her mother</u>*_j* if *she*_i could go to visit <u>her grandmother</u>*_k* as *it*_o had been awhile since *the girl*_i and *her grandma*_k had seen each other.

(3) "That's a good idea," *her mother*_j said. So *she*_i packed <u>a nice basket</u>*_l* for *Little Red Riding Hood*_i to take to *her grandmother*_k.

(4) When *the basket*_l was ready, *the little girl*_i put on her red cloak and kissed *her mother*_j goodbye.

(5) "Remember, go straight to *Grandma's* house," *her mother*_j cautioned. "Don't dawdle along the way and please don't talk to strangers! The woods are dangerous."

Figure 6.1 Example of pronouns in contexts teaching material

According to the figure 6.1, instructors may let students analyze noun phrases and their referents in the narrative. Accessibility Marking Scale will be applied here; later, factors (distance, competition, saliency, and unity) influencing how the writer expresses each referent will help the students better understand when and how to use pronouns in contexts.

The narrative is separated into five parts and explanations will be attached.

(15) Pronoun

(1) Once upon a time, there was <u>a little girl, who lived in a village near the forest</u>. Whenever *she*_iwent out, *the little girl*_i wore a red riding cloak, so everyone in the village called *her*_i *Little Red Riding Hood*_i.

(2) Full name   *(3) Long difinite description*   *(15) Pronoun*

In the example part (1), a noun phrase 'a little girl' with a modifier 'who lived in a village near the forest' is initially mentioned. After that the noun phrase is replaced with variety of pronominal expressions. The pronominal referents consist of pronouns 'she' and 'her' with the highest accessibility marking degree; a long definite description 'the little girl' with low accessibility marking degree; and a full name 'Little Red Riding Hood' with low accessibility marking degree. The factor the most influences referent expressing might be 'distance' and 'saliency' since there is not any competitors in between. For example, the write uses 'she' right after the first sentence containing the noun phrase the pronoun refers to; the readers could find the antecedent of the pronoun 'she' easily in the previous sentence.

(2) Full name (15) Pronoun

(2) One morning, *Little Red Riding Hood*, asked her mother, if she could go to visit her grandmother, as it, had been awhile since *the girl*, and *her grandma*, had seen each other.

(4) Short definite description   Repeated noun phrase
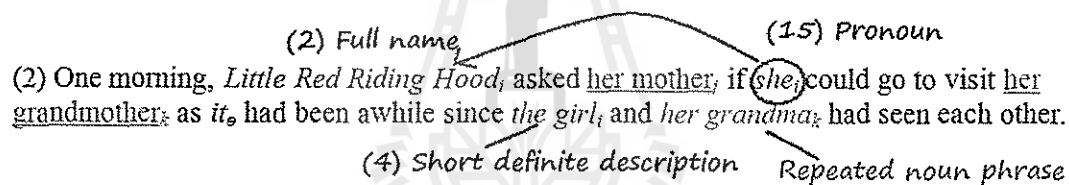
In part (2), a full name 'Little Red Riding Hood' is expressed again though it is the topic of the writing; the writer might consider 'distance' between the initial noun phrase 'a little girl' with a modifier in part (1). In the same sentence but different clauses, the pronoun 'she' is expressed to refer to 'Little Red Riding Hood', a topic of the sentence which is predictable for readers. Later, the writer mentions two new noun phrases 'her mother' and 'her grandmother' with the same number and gender as 'Little Red Riding Hood'. Instead of using the pronoun 'she', the writer express a short definite description 'the girl' due to the two competitors in between which might make the sentence ambiguous. In addition, the writer repeats the expression 'her grandma' to avoid ambiguity to the readers.

Repeated noun phrase

(3) "That's a good idea," *her mother₁* said. So *she₁* packed <u>a nice basket₁</u> for *Little Red Riding Hood₁* to take to *her grandmother₂*.

(2) Full name

Repeated noun phrase

(4) Short definite description   (3) Long definite description

(4) When *the basket₁* was ready, *the little girl₁* put on her red cloak and kissed *her mother₁* goodbye.

Repeated noun phrase

Repeated noun phrase   Repeated noun phrase

(5) "Remember, go straight to *Grandma₂*'s house," *her mother₁* cautioned. "Don't dawdle along the way and please don't talk to strangers! The woods are dangerous."

In similar way, pronominal expressions in parts (3), (4), and (5) might be explained as follow. Since the noun phrases '*her mother*' and '*her grandma*' are not the topic of the writing, the writer repeats the noun phrases instead of using a pronoun '*she*'. This is to avoid ambiguity or to help the readers to better understand the story. Additionally, whenever the pronominals are distant the writer will choose short or long definite descriptions to make the sentences clearer for the readers.

The materials could be simply designed with an emphasis on how and what factors influencing natives pronominal expressions. The information, hopefully, encourages the learners to better understand when to use and not to use pronouns in contexts.

Finally, the long-term goal of learner corpus study is to adjust syllabus design. After presenting the results of the present study to English instructors, pronoun teaching should more emphasize pronoun and co-reference across sentences and paragraphs. Plus, in the course of Introduction to Discourse Analysis, pronouns as one of cohesive device should be clarified for when to express pronouns and when to express the other referents and how to make decision for the referent choices as natives do.

In summary, the results of the present study can be employed in classroom teaching and materials. Analysis of pronoun use and pronoun errors in contexts can be applied in classroom as exemplified above associated with the Accessibility Hierarchy and the factors proposed by Areil (1990). In addition, syllabus adjustment on pronouns is suggested for better communication in English, writing in particular.

6.1.2 Suggestions for Further Research

Although the present study is corpus-based, the size of corpus is relatively small. In addition, the features in focus are rather narrow. Extending further studies can be in many ways as follows:

(a) Corpus size:

The Thai corpus size can be larger by collecting more learner data. Directions of subjects could be extended to learners in different levels (high school students or graduates), learners in different fields (non-English majors), or teachers of English. Plus, comparisons among the subjects will be challenging to investigate.

To expand learner corpus size will make the results and conclusions more reliable. The suggestions from the studies with larger size of learner corpus will inform more about Thai EFL performance, their language needs, and language development. As a result, pedagogical implications will be more useful and applicable.

(b) Language:

Since the tagsets designed in the present study are extendable; other parts of speech (POS) can be tagged with error tagsets under each POS. Moreover, learner corpus can be explored in specific linguistic features such as syntax, phonology, lexis, or discourse; error descriptions must be clear-cut. For examples, some tagset can be

added some symbols at the beginning and at the end of the tagset: with <S> for a syntactic tagging and with <D> for a discoursal tagging.

To further investigate learner language, the puzzle picture of how particular learners produce the target language will be discovered and completed the jigsaw little by little with the same systematic tagset.

(c) Methodology:

To reduce one of the difficulties in CEA study, learner data could be collected in electrical form at first place. In other words, computerized internet technology should take parts here, for example, students type and send all writing via e-mail, or they type on a diary blog or a facebook note can be accessible choice in reality. Collecting data will be more authentic with less time-consuming of data transformation.

In addition, the researcher has to take the four characteristics of tagset suggested by Granger (2003 and 2008) that the tagset should be informative and manageable; reusable for variety of languages; flexible for addition or deletion of tags at annotation and post-annotation stages; consistent error tagging principles between annotators.

In brief, the further studies could support and expand value of learner corpus based studies in terms of investigation of larger corpus size, broader language features, and more applicable research designs in this field. The more data stored, the more reliable learner language conclusions can be accomplished.

6.2 Conclusions of the Study

The present study aims to investigate how pronouns are used on narrative writings by Thai university English majors. In addition, it explores what types of pronoun errors are commonly produced by the learners. Finally, developmental patterns of pronouns over time are in focused if any.

The fundamental approaches of the study are Contrastive Analysis (CA), Error Analysis (ER) and Interlanguage (IL). However, these approaches for studies in a field of Second Language Acquisitions have some limitations as mentioned in Chapter 2. Newer ways to manage with learner data associated with CA, EA, and IL concepts are applied in the present study. Learner corpus-based studies commonly occupied two approaches to examine the learner language: Computer Interlanguage Analysis (CIA) Computer-aided Error Analysis (CEA); the present study mainly makes use of CEA including tagging of the learner language after designing POS and error tagsets of the feature in focus in the study, pronouns.

The research instruments to elicit learner data were three narrative writings: *Little Red Riding Hood, Snow White and the Seven Dwarfs, and Cinderella.* After collecting all written texts, the data obtained were transformed into electrical format in order to store and tag with pronoun tagset. Later, manually error tagged was processed by two taggers trained beforehand. WordSmith Tools was employed here as a supportive tool for concordancing contexts pronouns and pronoun errors occurred.

The results showed that Year 1 produced the highest pronouns and errors; Year 4 produced high pronouns but the least errors. In addition, pronouns were commonly used by students in Year 1 and 2 for substitute mentioned noun phrases/proper names; they were used interchangeably with other referents such as

definite noun phrases and proper names by students in Year 3 and Year 4. Some errors of omission, overuse, case, gender and number were found from writings of Year 1 and Year 2; the errors produced by Year 3 and 4 almost entirely dropped. However, errors of mis-coreference equally existed in the student writings.

The results can be partially explained with first language transfer. Omission, overuse, and malformation were probably affected by first language transfer. In addition, overuse in relative clauses might have been influenced by overgeneralization since such overuse occurs neither in Thai nor in English. Lastly, due to the present teaching materials for EFL learners, in general pronoun usage in contexts (across sentences or paragraphs) is not included in textbooks.

With some challenging issues of creating a tagged Thai learner corpus: data collecting, transforming, and tagging, learner corpus-based study with CEA had to be cautiously processed. Moreover, other challenging issues might be found if the corpus study is extended in the future. Learner corpus can be collected; later POS tagsets and other error tagsets can be designed and employed to investigate in the way of second language acquisition but with more effective approaches in terms of storing, managing, and reusing the learner data. These advantages of study could be useful for future pronoun language teaching and learning and research.

Finally, pedagogical implications as classroom teaching and materials applications with some examples are presented. Also, the results of pronoun used in contexts by Thai students in the present study could be additional information to support improving English course syllabus. In addition, some further ideas for future studies in this field are given. The researchers interested in this type of study might

expand their investigations in terms of corpus size, language features, and learner corpus based approach (both CEA and CIA).

6.3 Summary of the Chapter

The final chapter provides explanations of pronouns and errors occurring in the student writings. Also clarifications of pronoun development over time are included in order to give a picture of pronoun changing during four year of the Thai university English majors.

Finally, pedagogical implications as classroom teaching and materials applications with an example are presented for improving syllabus after getting the results of pronouns used by Thai students in writings. In addition, some further ideas for future studies in this field are given.

In summary, the learner corpus-based study with tagged PRO/error approach is applicable and useful for language teaching and learning with its more reliable methods and better data management (store, annotation, and reusability). It can be extended for larger corpus size and with more features of language investigated in the future.

REFERENCES
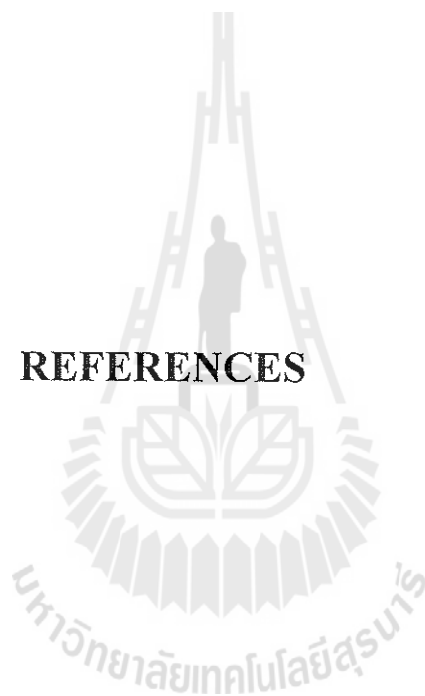
REFERENCES

Aart, J. and S. Granger, (1998). *Tag sequences in learner corpora: a key to interlanguage grammar and discourse.* In Granger (ed.) 1998a, pp. 132-141.

Abe M. (2003) A Corpus-based Contrastive Analysis of Spoken and Written Learner Corpora: The Case of Japanese-speaking Learners of English. In Archer et al. (eds.) (2003), 1-9.

Adamson, H. D. (1989). Does the bioprogram affect second language acquisition? In M. Eisenstein (Ed.), The dynamic interlanguage: Empirical studies in second language variation (pp. 33-47). New York: Plenum Press.

Adjémian, Ch. (1976). On the nature of interlanguage systems. *Language Learning.* 26, 279-320.

Agerström J. (2000) Hedges in Argumentative Writing: A Comparison of Native and Non-Native Speakers of English. In Virtanen T. and Agerström J. (eds) *Three Studies of Learner Discourse. Evidence from the International Corpus of Learner English.* Rapporter Från Växjö Universitet, Humaniora 10, Växjö Universitet, 5-42.

Aikawa, T. (1991). Bound variable interpretation and overt pronouns in Japanese. Alexander, M. & M. Dressler (eds.), *Papers from the Second Annual Meeting of the Forum Linguistics Society of MidAmerica.* University of Mchigan, Ann Arbor, pp. 196-216.

Altman, R. (1997). Oral production of vocabulary. In J. Coady & T. Huckin (Eds.), Second language vocabulary acquisition (pp. 67-97). Cambridge: Cambridge University Press.

Andersen, R. W. (1983). *Pidginization and creolization as language acquisition.* Rowley: Newbury House.

Ariel, M. 1988. Referring and accessibility. *Journal of Linguistics*, 24.65-87.

Ariel, M. 1990. *Accessing noun-phrase Antecedents.* London: Routledge.

Ariel, M. 1994. Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of Linguistics*, 30.3-42.

Aroonmanakun, W. (1999). *Extending Focusing for Zero Pronoun Resolution in Thai.* PhD, Georgetown University.

Asao K. (1997) Learner's Corpus and Bilingual Dictionaries. *Paper presented at JACET (Japan Association for College English Teachers) 36th Annual Convention, 5, September 1997.*

Aston, G. (2000), Corpora and language teaching, *in* L. Burnard and T. McEnery(eds), *Rethinking language pedagogy from a corpus perspective: Papersfrom the Third International Conference on Teaching and Language Corpora,* Frankfurt am Main: Peter Lang, pp. 7-17.

Baker, C. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry*, 10(4), 533-581.

Beck, M., Schwartz, B. D., & Eubank, L. (1995). Data, evidence and rules. In L. Eubank, L. Selinker, & M. Sharwood Smith (Eds.), *The current state of interlanguage: Studies in honor of William E Rutherford* (pp. 177-195). Amsterdam: John Benjamins.

Biber, D. & Conrad, S. (2009). *Register, Genre, and Style.* Cambridge: Cambridge University Press.

Bickerton, D. (1983). Comments on Valdman's 'Creolization and second language acquisition'. In R. Andersen (Ed.), *Pidginization and creolization as language acquisition* (pp. 235-240). Rowley: Newbury House.

Blagoeva, R. (2002). Language and Computers, Advances in Corpus Linguistics. *Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23).* 297-307.

Borjars, K., and Burridge, K. (2001). *Introducing English Grammar.* London: Arnold.

Bowker, Lynne/Pearson, Jennifer (2002), *Working with Specialized Language. A Practical Guide to Using Corpora.* London: Routledge.

Brezolin, A. (2008) Disposable corpus in translator training: translating medical abstracts into L2. In: 8th Teaching and Language Corpora Conference, 2008, Lisboa. *Proceedings of the 8th Teaching and Language Corpora Conference.* Lisboa : Associação de Estudos e de Investigação do ISLA - Lisboa, 45-50.

Cambridge University Press. (2006). *Cambridge Learner Corpus.* [available at http//:www.cambridge.org/elt/corpus/learner corpus.htm]

Campbell, R. (1969). *Noun substitutes in Modern Thai: A study in pronominality.* The Hague: Mouton.

Campbell, S. and Shaweevongse, M.S. (1957). *The Fundamentals of the Thai Language.* US: Thai Australia Co.Ltd.

Carroll, M. (1984). *Cyclical learning processes in second language production.* Frankfurt: Peter Lang.

Cazden, C. (1968). The acquisition of noun and verb inflections. *Child Development,* 39(2), 433-448.

Chafe, W. 1994. *Discourse, consciousness, and time.* Chicago: Chicago University Press.

Chen, H-J.H. (1998). *Underuse, overuse and misuse in Taiwanese EFL learner corpus.* In S. Granger and J. Hung (eds) 1998, pp. 25-28.

Chi A.M., Pui-yiu K.W. and Chau-ping M.W. (1994), *Collocational problems amongst ESL learners: a corpus-based study,* in Flowerdew & Tong (eds.): 157-165.

Chomsky, N., 1981. *Lectures on Government and Binding.* Foris, Dordrecht.

Chomsky, N. (1995). *The Minimalist Program.* Cambridge, MA: The MIT Press.

Clark. H. (1996). *Using language.* Cambridge University Press.

Clark, H. & Bangerter, A. (2004) Changing Ideas about Reference. In Noveck, I. & Sperber, D. (eds.) *Experimental Pragmatics.* Basingstoke: Palgrave Macmillan. 25-49.

Clahsen, H. (1995). German plurals in adult second language development: Evidence for a dualmechanism model of inflection. In L. Eubank, L. Selinker, & M. Sharwood Smith (Eds.), *The current state of interlanguage: Studies in honor of William E Rutherford* (pp. 123-137). Amsterdam: John Benjamins.

Cobb T. (2003) Analyzing late interlanguage with learner corpora: Québec replications of three European studies. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes* 59(3): 393-423.

Cohen,J.A.(1998).Practice parameters for the assessment and treatment of children and adolescents with posttraumatic stress disorder. *Journal of the American Academy of Child & Adolescent Psychiatry,37*(10),4S-26s.

Cooreman, A., & Kilborn, K. (1991). Functionalist linguistics: Discourse structure and language processing in second language acquisition. In T. Huebner & C. A.Ferguson (Eds.), *Cross Currents in Second Language Acquisition and LinguisticTheory* (pp. 195-224). Amsterdam: Benjamins.

Corder, S.P. (1967). The significance of learners' errors. *International Review of Applied Linguistics* 5 (4). pp. 161-170.

Corder, S. P. (1977). 'Simple codes' and the source of the second language learner's initial heuristic hypothesis. *Studies in Second Language Acquisition*, 1(1), 1-10.

Corder, S.P. (1981). *Error Analysis and Interlanguage*. Oxford. Oxford University Press

Corder, S.P. (1981). *Error analysis and interlanguage*. Oxford: Oxford University Press.

Corder, S. P. (1992). A role for the mother tongue. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language learning* (rev. ed.), (pp. 18-31). Amsterdam: John Benjamins.

Cook, V.J. (1993), *Linguistics and Second Language Acquisition*, Basingstoke: Macmillan

Cooke, J. (1968). *Pronominal Reference in Thai, Burmese, and Vietnamese*. University of California Publications in Linguistics 52. Berkeley and Los Angeles: University of California Press.

Cresswell, J.W. (2005). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. New Jersey: Pearson Education Inc.

Dagneaux, E., S. Denness, S. Granger, and F. Meunier. (1996). *Error Tagging Manual Version 1.1*. Centre for English Corpus Linguistics, Université catholique de Louvain, Louvain-la-Neuve.

Debrock, M. & Flament-Boistrancourt, D. (1996) "Le corpus LANCOM: bilan et perspectives", *ITL – Review of Applied Linguistics* 111-112: 1-36.

Degand, L. and Perrez, J. (2004). *Causale connectieven in het leerdercorpus Nederlands*. N/F 4: 129-144.

De Beaugrande, R. and Dressler, W.V. (1981). *Introduction to Text Linguistics.* Longman, New York: NY.

De Cock, S. (2003) *Recurrent sequences of words in native speaker and advanced learner spoken and written English.* Unpublished PhD thesis, Université catholique de Louvain.

De Haan, P. (1997). How native-like are advanced learners of English? In Renouf, A. (ed.) *Explorations in Corpus Linguistics*, Amsterdam: Rodopi, pp. 55-65.

Díaz-Negrillo, A., & Valera, S. (2009). What results give which strategies in error annotation. CALICO Pre-Conference Workshop, Tempe, AZ. Retrieved from http://www.sfs.uni-tuebingen.de/~dm/events/aall09/diaz-negrillo-valera.pdf

Diez-Bedmar, M. B. (2009). Written learner corpora by Spanish students of English: an overview. In Pascual Cantos Gómez and Aquilino Sánchez Pérez (eds.) *A Survey on Corpus-based Research.* Proceedings of the AELINCO Conference Murcia: Asociación Española de Lingüística del Corpus, pp. 920-933.

Dulay, H. and Burt, M. (1974). A new perspective on the creative construction process in child second language acquisition. *Language Learning* 24(2):253–278.

Dulay, H., Burt, M. and Krashen, S. (1982). *Language Two.* New York: Oxford University Press.

Edge. J. (1989). *Mistake and Correction.* London: Longman.

Ellis, R. (1985a). Sources of variability in interlanguage. *Applied Linguistics*, 6(2), 118-131.

Ellis, R. (1985b). *Understanding second language acquisition*. Oxford: Oxford University Press.

Ellis, R. (1992). Learning to communicate in the classroom: A study of two language learners' requests. *Studies in Second Language Acquisition*, 14(1), 1-23.

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Ellis, R. (1997). SLA and language pedagogy: An educational perspective. *Studies in Second Language Acquisition*, 19: 69-92.

Faerch, C., Haastrup, K., & Phillipson, R. (1984). *Learner Language and Language Learning*. Clevdon: Multilingual Matters.

Felix, S.W. (1980). *Second Language Developmental Trends and Issues*.Gunter Narr Verlag Tübingen.

Flowerdew, L. (1998). "The exploitation of small learner corpora in EAP materials design". *Small Corpus Studies and ELT. Theory and practice*. Eds. Mohsen

Flowerdew, (2000) Using a genre-based framework to teach organizational structure in academic writing. *ELTJournal* 54, 4, 369-78.

Flowerdew, Lynne (2001), The Exploitation of Small Learner Corpora in EAP Materials Design. In: Ghadessy/Henry/Roseberry 2001, 363_380.

Flynn, S. (1984). A universal in L2 acquisition based on a PBD typology. In F. Eckman,L. Bell, & D. Nelson (Eds.), *Universals of second language acquisition*). Rowley, MA: Newbury House.

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

Fox, C. (1993). *Plurals and Mass Terms in Property Theory*. Ph.D. thesis, University of Essex, Colchester, UK.

Gabrielatos C. and McEnery T. (2005) Epistemic modality in MA dissertations. In Fuertes Olivera P.A. (ed.) (2005) *Lengua y Sociedad: Investigaciones recientes en lingüística aplicada. Lingüística y Filología* no. 61. Valladolid: Universidad de Valladolid, 311-331.

http://eprints.lancs.ac.uk/102/01/Epistemic_modality_in_MA_dissertations.pdf

Gilquin, G., S. Granger and M. Paquot (2007) Writing sections, in M. Rundell (Editor in chief) *Macmillan English Dictionary for Advanced Learners* (second edition), pp. IW1-IW29. Oxford: Macmillan Education.

Gass, S. M. and Selinker, L. (2001). *Second Language Acquisition: An Introductory Course*. (2nd eds.). US: Lawrence Erlbaum Associates.

Giacobbe, J. (1992). A cognitive view of the role of L1 in the L2 acquisition process. *Second Language Research*, 8 (3), 232–250.

Gilquin G. (2007) To err is not all. What corpus and elicitation can reveal about the use of collocations by learners. *Zeitschrift für Anglistik und Amerikanistik* 55(3): 273-291.

Gilquin G. and Paquot M. (2007) Spoken features in learner academic writing: identification, explanation and solution. In *Proceedings of the Fourth Corpus Linguistics Conference*, University of Birmingham , 27-30 July 2007.

http://ucrel.lancs.ac.uk/publications/CL2007/paper/204_Paper.pdf

Givón, T. (1979): *On Understanding Grammar*. New York: Academic Press.

Givón, T. 1980. The binding hierarchy and the typology of complements. *Studies in Language* 4: 333-377.

Givón, T. et al. (1983). *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. John Benjamin, Amsterdam.

Granger, S. (1998). *Learner English on Computer*. London & New York: Addison Wesley Longman.

Granger, S. (2002). A bird's-eye view of learner corpus research, in Granger, S., Hung, J., and Petch-Tyson, S. (eds), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Atlanta & Amsterdam: John Benjamins: 3-33.

Granger, S. (2003). The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research. *TESOL Quarterly*, 37/3: 465-80.

Granger, S. (2008). Learner Corpora in Foreign Language Education. In Van Deusen-Scholl N. and Hornberger N.H. (eds) *Encyclopedia of Language and Education. Volume 4. Second and Foreign Language Education*. Springer, 337-351.

Granger, S., Hung, J., & Petch-Tyson, S. (eds.) (2002), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.

Granger S. and Rayson P. (1998) Automatic Lexical Profiling of Learner Texts. In Granger S. (ed.) *Learner English on Computer*. London & New York: Addison Wesley Longman, 119-131.

Granger, S. and Tyson, s. (1996), *Connector usage in the English essay writing of native and non-native EFL speakers of English. In: World Englishes* 15(1), 17-27.

Grosz, B. Anavind, K, & Scott, W. (1995). Centering a Framework for modeling local coherence of discourse. *Computational Linguistics*, 21 (2).

Gundel, J. K., Hedberg, N. & Zacharaski, R. 1993. *Cognitive status and the form of referring expressions*. Language, 692.274-307.

Guo, X. (2006). Verbs in the written English of Chinese learners: A Corpus-based comparison between non-native speakers and native speakers. Doctoral dissertation. University of Birmingham, Birmingham. Retrieved from http://www.linguistics-journal.com/thesis_Guo.pdf

Haley, B. (1980). Interlanguage and their relations. *Interlanguage Bulletin* 5, 3-30.

Hatton, H. (1978). *First Person Pronominal Realization in Thai Autographical Narrative: A Sociolinguistic Description*. Ph.D. Dissertation, University of Pensylvania.

Hasselgren A. (1997). The EVA Corpus of Norwegian School English. *ICAME Journal* 21: 123-124.

Hewings M. and A. Hewings (2002) It is interesting to not that …. A comparative study of anticipatory "it" in student and published writing. *English for Specific Purposes* 21(4): 367-383.

Hoonchamlong, Y. 1991. *Some Issues in Thai Anaphora: A Government and Binding Approach*. PhD, University of Wisconsin-Madison.

Howard, C.D. (2009) Part of speech tagged asynchronous CMC: Comparing native, non-native, and newspaper English. *Studies in the Linguistic Sciences*, 34(1), 110-126.

Huang, J. and Hatch, E. (1978). A Chinese child's acquisition of English. In E. Hatch (Ed.), *Second Language Acquisition: A book of readings* (pp. 118-31). Rowley, MA: Newbury House.

Huang, Y. (1994). *The syntax and pragmatics of anaphora: A study with special reference to Chinese.* Cambridge, England: Cambridge University Press.

Huang, Y. (2000). *Anaphora: a cross-linguistic study.* New York: Oxford University Press.

Hyland, K.(2002). *Options of Identity in academic writing.* ELT Journal, 56(4), 351-358.

Hyltenstam, K. (1977). Implicational patterns in interlanguage syntax variation. *Language Learning,* 27 (2), 383-410.

Ide, N. (2000). The XML framework and its implications for corpus access and use. *Proceedings of Data Architectures and Software Support for Large Corpora.* Paris: European Language Resources Association.

Ife A. (2004). The L2 learner corpus: reviewing its potential for the early stages of learning. Baynham M., A. Deignan and G. White (eds) *Applied Linguistics at the Interface.* London/Oakville: British Association for Applied Linguistics in association with Equinox. 91-103.

Izumi et al. (2003). "The development of the spoken corpus of Japanese learner English and the applications in collaboration with NLP techniques". *Proceedings of the Corpus Linguistics 2003 Conference.* Lancaster: UK, 28-31.

Izumei, E. and Isahara, H. (2004). Investigation into language learners' acquisition order based on the error analysis of the learner corpus. *In Proceeding of Pacific-Asia Conference on Language, Information and Computation (PACLIC) 18 Satellite Workshop on E-Learning,* Tokyo, Japan.

Izumi et al. (2005). "Error annotation for corpus of Japanese Learner English". *Proceedings of the Sixth International Workshop on Linguistically Interpreted Corpora (LINC 2005).* Jesu Island, Korea: 71-80.

Ivanič, R & Camps, D. 2001. *I am how I sound. Voice as self representation in L2 writing*. Journal of Second Language Writing, 10, 3-33.

Iwasaki S. and Ingkaphirom, P. (2005). *A Reference Grammar of Thai*. England: Cambridge University Press.

Jacobs, R. A. (1995). *English Syntax*. Oxford: Oxford University Press.

James, C. (1998). *Errors in language learning and use: Exploring error analysis*. London & New York: Longman.

Johnson, K. (1988). *Cognitive Skill Acquisition and Second-Language Acquisition*. Department of Linguistic Science, University of Reading, Whiteknights, POBox218, Reading RG62AA.

Johnstone, B. (2002). *Discourse Analysis*. Blackwell Publishing: UK.

Kang, J. (2004). Telling coherent story in a foreign language: analysis of EFL learners' referential strategies in oral narrative discourse. *Language Learning*, 44, 101-122.

Kellernam, E. and Sharwood Smith, M. (1986). *Cross-linguistic influence in Second Language Acquisition*. New York: Pergamon.

Kim, Y.J. (1997). The acquisition of Korean. In: Slobin, D. (Ed.), *The Crosslinguistic Study of Language Acquisition*, 4: 335-444.

Klein, W. (1986). *Second Language Acquisition*. Cambridge: Cambridge University Press.

Klein, W. & Perdue, C. (eds.) (1992). *Utterance structure. Developing grammars again*. Amsterdam: John Benjamins.

Kolln, M. (1991). *Understanding English Grammar*. Fours Edition, New York: Macmillan.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

Lappin, Shalom and Herbert Leass. 1994. *An algorithm for pronominal anaphora resolution*. Computational Linguistics, 20(4): 535-561.

Larsen-Freeman, D. & M. Long. (1991). *An Introduction to Second Language Acquisition Research*. London: Longman. Chap. 6: 153-219.

Leech, G. 1998. Learner Corpora: what they are and what can be done with them. In Granger, S. (ed). *Learner English on Computer*. Addison Wessley Longman: London, xiv–xx.

Leech, G. 2001. *English Grammar and Usage*. London: Longman.

Levinson, S. (1991). Pragmatic reduction of the binding conditions revisited. *Journal of Linguistics*, 27: 107-61.

Levy, E. 1982. *Toward an objective definition of "Discourse Topic"*. Papers from the eighteenth regional meeting of the Chicago Linguistic Society, ed. by K. Tuite, R. Schneider, & R. Chametzky, 259-304. Chicago: CLS.

Long, M.H. and Sato, C.J. (1983). Classroom foreigner talk discourse: forms and functions of teachers' questions. In H.W. Seliger and M.H. Long (Eds.) *Classroom Oriented Research in Second Language*. Rowley, MA: Newbury House.

Lyons, J. 1968. *Introduction to theoretical linguistics*. London, Cambridge University Press.

Lüdeling, A., Adolphs, P., Kroymann E. and M. Walter (2005) Multi-level error annotation in learner corpora. *Proceedings from The Corpus Linguistics Conference Series*, vol 1 n°1. Available from http://www.corpus.bham.ac.uk/PCLC/#corpora.

Marcus, G.F. (1993). Negative evidence in language acquisition. *Cognition*, 46, 53–85.

Martínez, I. 2005. *Native and non-native writers' use of first personal pronouns in the different sections of biology research articles in English.* Journal of Second Language Writing, 14, 174-190.

McEnery, T., Xiao, R., and Tono, Y. (2006). *Corpus-Based Language Studies: an advanced research book.* US: Routledge.

Milton J. (1998) Exploiting L1 and interlanguage corpora in the design of an electronic language learning and production environment, in S. Granger (ed.), *Learner English on Computer,* 186-198.

Milton, J. and Tsang, E. (1993). A corpus-based study of logical connectors in EFL students' writing. In Pemberton, R. and Tsang, E. (eds) **Studies in Lexis.** Language Centre, The Hong Kong University of Science and Technology, 215-46.

Mitchell, R., & Myles, F. (1998). *Second language learning theories.* London: Arnold.

Muansuwan, N. (2001). Directional serial verb constructions in Thai. *Proceedings of the 7th International HPSG Conference, UC Berkeley (22-23 July, 2000).* Dan Flickinger and Andreas Kathol, (eds). Stanford: CSLI Publications.

Myhill, J. (1991). Typological text analysis: Tense and aspect in creoles and second languages. In T. Huebner & C. A. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theories* (pp. 93-121). Amsterdam: John Benjamins.

Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning,* 48(3), 323-363.

Myles, F., Mitchell, R., & Hooper, J. (1999). Interrogative chunks in French L2: A basis for creative construction? *Studies in Second Language Acquisition*, 21(1), 49-80.

Neff J., F. Ballesteros, E. Dafouz, F. Martinez, J.-P. Rica, M. Diez and R. Prieto (2004) Formulating Writer Stance: A Contrastive Study of EFL Learner Corpora. In Connor U. and T.A. Upton (eds) *Applied Corpus Linguistics. A Multidimensional Perspective*. Amsterdam / New York: Rodopi. 73-89.

Negrillo, A. and Dominguez, J. 2006. Error Tagging System for Learner Corpora. *Reseach in Second Language Acquisition*, 19. 83-102.

Nesselhauf, N. (2004). "Learner corpora and their potential for language teaching". *How to Use Corpora in Language Teaching*. Sinclair, J.M. (ed.). Amsterdam: John Benjamins, 125-52.

Neumer, J. L. (1987). Cohesive ties and chains in good and poor freshman essays. *Research in the Teaching of English*, 21(1), 93-105.

Nicholls D. (2003) The Cambridge Learner Corpus – error coding and analysis for lexicography and ELT. In Archer et al. (eds.), 572-581.

Noguchi, T. (1997). Two types of pronouns and variable binding. *Language*, 73, pp. 770-797.

Norrish, J. (1983). *Language learners and their errors*. London: The Macmillan Press.

Noss, R.B. (1964). Thai Reference Grammar. US: Foreign Service Institute.Office of the Nation Education Commission (ONEC), Office of the Priminister, Thailand. (1999). *National Education Act of B.E. 2542 (1990)*. Retrieved June 8, 2007, from http://www.moe.go.th/English/edu-act.htm

Nunan, D. (1996). *Atlas*. Boston, MA: Heinle and Heinle.

Odlin, T. (1989). *Language transfer*. Cambridge, UK: Cambridge University Press.

Palakornkul, A. (1972). *A Socio-Linguistic Study of Pronominal Strategy in Spoken Bangkok Thai*. PhD dissertation. University of Texas.

Perdue, C. ed. (1993). *Adult language acquisition: Cross-linguistic perspectives. Volume 2. The results*. Cambridge: Cambridge University Press.

Pica, T. (1998). Second language learning through interaction: Multiple perspectives. In V. Regan (Ed.), *Contemporary approaches to second language acquisition in social context* (pp. 9-31). Dublin: University College Dublin Press.

Pingkarawat, N. (1989). *Empty Noun Phrases and the Theory of Control, with Special Reference to Thai*. Ph.D. Dissertation, University of Illinois at Urbana-Champaign.

Pravec, N. (2002). Survey of learner corpora. *ICAME Journal*, 26: 81-114.

Pu, M. M. (1995). Anaphoric patterning in English and Mandarin narrative production. *Discourse Processes, 19*, 278–300.

Ragan, P. 2001. 'Classroom use of a systemic-functional small learner corpus' in M. Ghadessy, A. Henry & R. Roseberry (eds.): *Small Corpus Studies and ELT*. Amsterdam: John Benjamins.

Reder, P., Duncan, S. & Lucey, C. (2003) *Studies in the Assessment of Parenting*.London: Routledge.

Richards, J.C. (1974). *Error Analysis*. London: Longman.

Robert, J.M. (1989). Intralanguage, agrammatism and the conceptual system. International Review of Applied Linguistics, 27(3), 217-221.

Römer, U. (2004). Comparing real and ideal language learner input: the use of an EFL textbook corpus in corpus linguistics and language teaching, in G. Aston, S. Bernardini and D. Stewart (eds) *Corpora and Language Learners*, pp. 151–68. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Sato, C. J. (1988). Origins of complex syntax in interlanguage development. *Studies in Second Language Acquisition*, 10(3), 371-395.

Savery, J. R. (2006). Overview of Problem-based Learning: Definitions and Distinctions. *Interdisciplinary Journal of Problem-based Learning*, 1(1).

Schachter, J. (1974). *An error in error analysis*. Language Learning, 24: 205-215.

Schachter, J. (1992). A new account of language transfer. In S. M. Gass & L. Selinker (Eds.), *Language transfer in language learning* (rev. ed.), (pp. 32-46). Amsterdam: John Benjamins.

Schachter, J. & Celce-Murcia, M. (1977). *Some reservations concerning error analysis*. TESOL Quarterly, 11: 441-451.

Schofield, A. (ed.) (1998). *Benchmarking in HigherEducation: An International Review*. CHEMS: London and UNESCO: Paris.

Schumann, J. (1978). Second language acquisition: The pidginization hypothesis. In E. Hatch (Ed.), *Second Language Acquisition: A book of readings* (pp. 256-71). Rowley, MA: Newbury House.

Schwartz, B. D. (1993). On explicit and negative data affecting competence and linguistic behavior. *Studies in Second Language Acquisition*, 15(2), 147-163.

Scott, M. (1998) WordSmith Tools Version 3. Oxford University Press, Oxford, England. Siewierska, A. (1993). Syntactic weight vs. information structure and word order variation in Polish. *Journal of Linguistics*, 29.233-265.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10: 209-31.

Selinker, L. (1983). Language transfer. In S. Gass & L. Selinker (Eds.), *Language transfer in language learning* (pp. 33-68). Rowley, MA: Newbury House.

Selinker, L. (1992). *Recovering interlanguage*. London: Longman.

Selinker, L., & Douglas, D. (1985). Wrestling with 'context' in interlanguage theory. *Applied Linguistics*, 6(2), 190-204.

Shaughnessy, Mina P. 1977. *Errors and expectations*. New York: Oxford University Press.

Silva-Corvalán, Carmen. 1991. "Spanish Language Attrition in a Contact Situation with English." *First Language Attrition*, ed. by Herbert Seliger & Robert Vago,151-171. Cambridge, UK: Cambridge University Press.

Skiba, R., & Dittmar, N. (1992). Pragmatic, semantic, and syntactic constraints and grammaticalization. *Studies in Second Language Acquisition*, 14(3), 323-349.

Spence R. (1998) A corpus of student L1-L2 translation. In Granger S. and Hung J. (eds) *Proceedings of the First International Symposium on Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching (14-16 December 1998)*. The Chinese University of Hong Kong, 110-112.

Spillner, B. (1991). *Error Analysis: A Comprehensive Bibliography*. John Benjamins, Amsterdam.

Takami, K. (1987). Anaphora in Japanese: Some semantic and pragmatic considerations. *Journal of pragmatics*, 11: 169-191.

Tan, M (2005) *Authentic Language or Language Errors? Lessons from a learner corpus*. English Language Teaching, 59 (2). pp. 126-134.

Tagnin, S. (2003), COMET - A Multilingual Corpus for Teaching and Translation, *PALC 2001: Practical Applications in Language Corpora, 2003*, Lodz, Poland, Frankfurt am Main: Peter Lang - Europäischer Verlag der Wissenschaften, 535-540.

Tarone, E. (1982). Systematicity and attention in interlanguage. *Language Learning*, 32(1), 69-84.

Tarone, E. (1983). On the variability of interlanguage systems. *Applied Linguistics*, 4(2), 143-163.

Tarone, E. (1989). On chameleons and monitors. In M. Eisenstein (Ed.), *The dynamic interlanguage: Empirical studies in second language variation* (pp. 3-15). New York: Plenum Press.

Tarone, E. (1994). Research approaches in studying SLA; or, If the shoe fits, wear it ... in E. Tarone, S. Gass & A. Cohen (Eds.), *Issues in Research Methodology in Second-Language Acquisition*. (323-336). Hillsdale, NJ: Lawrence Erlbaum Assoc.

Tenfjord K., P. Meurer and K. Hofland (2004) The ASK corpus - a language learner corpus of Norwegian as a second language. *Paper presented at the TALC 2004 conference, Granada - Spain, 6-9 July 2004.*

Teresa, T. (1977). The *Least You Should Know About English Basic Writing Skill.* Macomb Illinois : Holt, Rineheart and Winston.

Tono, Y. 1998 *A computer learner corpus-based analysis of the acquisition order of English grammatical morphemes*. In TALC (Teaching and Language Corpora) 98 – Conference Proceedings, Keble College Oxford, 24-27 July 1998, pp. 183-187.

Tono, Y. (2000a). A computer learner corpus-based analysis of the acquisition order of English grammatical morphemes. In Burnard, L. and T. McEnery (eds.) *Rethinking Language Pedagogy from a Corpus Perspective*. Frankfurt: Peter Lang, pp. 123-132.

Tono, Y. (2000b). *A corpus-based analysis of interlanguage development: Analysing POS tag sequences of EFL learner corpora*. In Lewandowska-Tomaszczyk, B. and J.P.Melia 2000, pp. 323-343.

Tono, Y. (2001). *Research on Dictionary Use in the Context of Foreign Language Learning*. Tübingen: Max Niemeyer Verlag.

Tono, Y. (2003). *Learner corpora: design, development and applications*. Paper resented at the Corpus Linguistics 2003 Conference (CL 2003), Lancaster.

Tono, Y. (2004). *Standard Speaking Test (SST) Corpus*.

[http://leo.meikai.ac.jp/~tono]

Tono, Y. (2006) Learner language data for pedagogical dictionaries: an analysis of EFL learners' use of the verb *make*. In Y. Tono, *Exploring L2 Learner's Dictionary Use: Issues and Methodologies. (LEXICOGRAPHICA* Series Maior) Tübingen: Max Niemeyer.

Tono, Y. and M. Aoki . (1998). *Developing the optimal learning list of irregular verbs based on the native and learner corpora*. In S. Granger and J. Hung (eds) 1998, pp. 113-118.

Towell, R. and Hawkins, R.D. (1994). *Approaches to Second Language Acquisition*. USA: Multilingual Matters.

Unger, C.(2006). *Genre, relevance and global coherence. The pragmatics of discoursetype*. Basingstoke & New York: Palgrave Macmillan.

Uzar R. (2002) A Corpus Methodology for Analysing Translation. In *Cadernos de Tradução* 2002 Universidade Federal de Santa Catarina. 237-265.

Uzar R. (2003a) Different Ways of Analysing Translation. In *Conference Proceedings of Practical Applications of Language Corpora 2001*: Peter Lang. 541-556.

Uzar R. (2003b) Computers in the Analysis of Translation. In *Research and Scholarship in Integration Processes*: Lódz University Press. 365-378.

Valin, V.R. Jr. (1990). Semantic parameters of split intransitivity. *Language*, 66:221-260.

van Els, T. et al (1984). *Applied linguistics and the learning and teaching foreign languages.* London: Edward Arnold.

van Hoek, K. (1995). Conceptual reference points: A cognitive grammar account of pronominal anaphora constraints. *Language* 71/2: 310–340.

Van Moere, A. & Howard, C. (2002) Introducing Email Exchange in a Freshman Curriculum. *The Journal of the Research Institute of Language Studies and Language Education, 12*(1), 99-117.

Virtanen, T. (1998). *Direct questions in argumentative student writing.* In Granger, S. (ed.), 94-118.

Walker, M.A. (1998). Centering, anaphora resolution, and discourse structure. In Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors, *Centering in Discourse*. Oxford University Press, 1998

Wales, K. 1996. *Personal Pronouns in Present-Day English.* Cambridge University Press: Cambridge.

Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly*, 4 (2), 123-30.

Wendan, L. (2004). Topic Chain in Chinese Discourse. *DISCOURSE PROCESSES*, 37(1), 25–45.

White, L. (1992). Subjacency violations and empty categories in L2 acquisition. In H. Goodluck & M. Rochemont (Eds.), *Island Constraints* (pp. 445-464). Dordrecht: Kluwer.

White, L. (1998). Introduction: The implications of divergent outcomes in L2 acquisition. *Second Language Research*, 14(4), 321-323.

Widdowson, H.G. 2000. 'On the limitations of linguistics applied.' *Applied Linguistics* 21/1: 3-5

Wikborg, Eleanor. 1985. Types of Coherence Breaks in University Student Writing, in N. E. Enkvist (ed.), *Coherence and Composition: A Symposium*, Abo Akademis Kopieringscentral, pp. 93-133.

Wikborg, Eleanor. 1990. Types of coherence breaks in Swedish student writing: misleading paragraph divisions, in U. Connor and A. M. Johns (eds.), *Coherence in Writing: Research and Pedagogical Perspectives*, TESOL, Alexandria, Virginia, pp. 131-150.

William Ancker, .*Errors and Corrective Feedback; Up date Theory and Classroom Practice.*, English Teaching Forum, XXXVIII, (Oktober, 2000), P. 21

Zaytseva, E. (2011). Register, genre, rhetorical functions: Variation in English nativespeaker and learner writing. In H. Hedeland, T. Schmidt & K. Wörner (eds) *Multilingual Resources and Multilingual Applications (Hamburg Working Papers in Multilingualism B 96)*, 239-242.

Zobl, H. (1982). A direction for contrastive analysis: the comparative study of developmental sequences. *TESOL Quarterly*, 16: 169-83.

Zobl, H. (1984). *Cross-language generalization and hypothesis* in Davies et al (Eds.) 1984.

APPENDICES

APPENDIX A

Curriculum Plan of English Majors,

Ubon Ratchathani University (UBU)

Curriculum of UBU English Majors

หลักสูตรศิลปศาสตรบัณฑิต
สาขาวิชาภาษาอังกฤษและการสื่อสาร
โครงสร้างหลักสูตร

จำนวนหน่วยกิตรวม ไม่น้อยกว่า 143 หน่วยกิต ประกอบด้วยวิชาต่าง ๆ ดังนี้

หมวดวิชาศึกษาทั่วไป ไม่น้อยกว่า	38 หน่วยกิต
ก. กลุ่มวิชาวิทยาศาสตร์ คณิตศาสตร์	7 หน่วยกิต
ข. กลุ่มวิชาภาษา	9 หน่วยกิต
ค. กลุ่มวิชาสังคมศาสตร์	6 หน่วยกิต
ง. กลุ่มวิชามนุษยศาสตร์	15 หน่วยกิต
จ. กลุ่มวิชาพลศึกษา	1 หน่วยกิต
หมวดวิชาเฉพาะ ไม่น้อยกว่า	102 หน่วยกิต
ก. กลุ่มวิชาเอก	78 หน่วยกิต
กลุ่มวิชาเอกบังคับ	63 หน่วยกิต
กลุ่มวิชาเอกเลือก	15 หน่วยกิต
ข. กลุ่มวิชาโท	24 หน่วยกิต

นักศึกษาเลือกศึกษาสาขาวิชาในสาขาวิชาหนึ่งที่เปิดสอนในมหาวิทยาลัยอุบลราชธานีเป็นวิชาโทโดยศึกษาตาม
ข้อกำหนดและเงื่อนไขของหลักสูตรวิชาโทสาขาวิชานั้น ๆ ให้ครบจำนวนหน่วยกิตที่กำหนด

หมวดวิชาเลือกเสรี ไม่น้อยกว่า 3 หน่วยกิต

นักศึกษาเลือกศึกษาวิชาที่เปิดสอนในมหาวิทยาลัยอุบลราชธานีเป็นวิชาเลือกเสรี

รายวิชา

หมวดวิชาศึกษาทั่วไป	38 หน่วยกิต

นักศึกษาสาขาภาษาอังกฤษและการสื่อสารลงทะเบียนใน รายวิชาต่าง ๆ โดยมีเนื้อหาสาระครอบคลุมกลุ่มกรรมในกลุ่ม
วิชาวิทยาศาสตร์และคณิตศาสตร์ มนุษยศาสตร์ สังคมศาสตร์ ภาษาและพลศึกษา ตามที่กำหนดไว้และสอดคล้องกับปรัชญา
การศึกษาทั่วไป จำนวนไม่น้อยกว่า 38 หน่วยกิต ประกอบด้วยกลุ่มวิชาและรายวิชาต่าง ๆ ดังนี้

ก. กลุ่มวิชาวิทยาศาสตร์ คณิตศาสตร์	7 หน่วยกิต
1101 140 มนุษย์กับสิ่งแวดล้อม (Man and Environment)	2(2-0-0)
1101 141 วิทยาศาสตร์ในชีวิตประจำวัน (Science in Everyday Life)	3(3-0-0)
ให้เลือกเรียนไม่น้อยกว่า 2 หน่วยกิต ในรายวิชา	
1101 142 โลกของพืช (Plant World)	2(2-0-2)
1101 143 การจัดการสิ่งแวดล้อม (Environmental Management)	3(3-0-3)
1101 144 จุลินทรีย์กับชีวิต (Microbiology and Life)	2(2-0-2)
1102 108 วัสดุศาสตร์ในชีวิตประจำวัน (Material Science in Everyday Life)	2(2-0-2)
1103 128 การจัดการเครื่องใช้ไฟฟ้าในชีวิตประจำวัน (The Management of Daily Appliances)	3(3-0-3)
1103 129 การถ่ายภาพด้วยกล้องดิจิทอล (Digital Photography)	3(3-0-3)
1103 130 โลกวิทยา (Earth Science)	3(3-0-3)
1103 126 ดาราศาสตร์เบื้องต้น (Introduction to Astronomy)	3(3-0-3)
1104 160 คอมพิวเตอร์ทั่วไป (General Computer)	2(2-0-2)
1104 161 การใช้ซอฟต์แวร์ชุดสำนักงาน (Office Software Usage)	2(1-2-2)
1104 162 การสร้างเวบขั้นพื้นฐาน (Basic Web Development)	2(1-2-2)

ข. กลุ่มวิชาภาษา   9 หน่วยกิต

1411 101   ภาษาไทยกับการสื่อสาร (Thai Language and Communication)   3(3-0-6)

1421 100   ภาษาอังกฤษพื้นฐานสำหรับ นักศึกษาศิลปศาสตร์ (Foundation English for Liberal Arts Students)   6(5-2-11)

ค. กลุ่มวิชาสังคมศาสตร์   6 หน่วยกิต

1441 100   มนุษย์กับสังคม (Man and Society)   3(3-0-6)

1708 100   เศรษฐศาสตร์ทั่วไป (General Economics)   3(3-0-1)

ง. กลุ่มวิชามนุษยศาสตร์   15 หน่วยกิต

1432 100   มนุษย์กับอารยธรรม (Man and Civilization)   3(3-0-6)

1432 101   วัฒนธรรมไทยในอดีตและปัจจุบัน (Thai Culture Past and Present)   3(3-0-6)

1447 200   มนุษย์กับการสื่อสาร (Man and Communication)   3(3-0-6)

1431 100   ภูมิปัญญาและพฤติกรรมมนุษย์ (Human Wisdom and Behaviour)   3(3-0-6)

1431 101   มนุษย์กับสุนทรียภาพ (Man and Aesthetics)   3(3-0-6)

จ. กลุ่มวิชาพลศึกษา   1 หน่วยกิต

1439 100   กิจกรรมพลศึกษา 1 (Physical Education Activity I)   1(0-2-1)

1439 101   กิจกรรมพลศึกษา 2 (Physical Education Activity II)   1(0-2-1)

หมวดวิชาเฉพาะ จำนวนไม่น้อยกว่า   102 หน่วยกิต

ประกอบด้วย 2 กลุ่มวิชา ดังนี้

ก. กลุ่มวิชาเอก   78 หน่วยกิต

    วิชาเอกบังคับ   63 หน่วยกิต

1411 240   ภาษากับวรรณกรรมไทย (Thai Language and Literature)   3(3-0-6)

1421 101   ภาษาอังกฤษสำหรับนักศึกษาศิลปศาสตร์ (English for Liberal Arts Students)   6(5-2-11)

1421 210   การพูด 1 (Oral Communication I)   3(3-0-6)

1421 211   การพูด 2 (Oral Communication II)   3(3-0-6)

1421 220   การเขียน 1 (Written Expression I)   3(3-0-6)

1421 230   ภาษาเบื้องต้น (Introduction to Language)   3(3-0-6)

1421 231   สัทศาสตร์ภาษาอังกฤษ (English Phonetics)   3(2-2-5)

1421 240   วรรณคดีเบื้องต้น (Introduction to Literature)   3(3-0-6)

1421 241   การอ่านวรรณกรรม (Reading Literature)   3(3-0-6)

1421 330   ภาษาศาสตร์กับการสื่อสาร (Linguistics and Language Use)   3(3-0-6)

1421 221   การเขียน 2 (Written Expression II)   3(3-0-6)

1421 320   การเขียน 3 (Written Expression III)   3(3-0-6)

1421 321   การแปล 1 (Translation I)   3(3-0-6)

1421 322   การแปล 2 (Translation II)   3(3-0-6)

1421 331   การสื่อสารในกลุ่มชนข้ามวัฒนธรรม (Cross Cultural Communication)   3(3-0-6)

1421 340   แนวคิดในวรรณกรรม (Themes in Literature)   3(3-0-6)

1421 341   วรรณกรรมกับศิลปะแขนงอื่น (Literature in Relation to Other Artistic Forms)   3(3-0-6)

1421 430   ลีลาภาษาและการสื่อสาร (Language Style and Communication)   3(3-0-6)

1421 440   แนวคิดโลกร่วมสมัย (Contemporary World Views)   3(3-0-6)

1421 490   การศึกษาอิสระ (Independent Study)   3(3-0-6)

    วิชาเอกเลือก ไม่น้อยกว่า   15 หน่วยกิต

1421 400   ภาษาอังกฤษธุรกิจ (English in Business)   3(3-0-6)

1421 401   ภาษาอังกฤษในหนังสือพิมพ์ (English in Newspapers)   3(3-0-6)

1421 420	การแปล 3 (Translation III)	3(3-0-6)
1421 332	ประวัติภาษาอังกฤษ (The Story of English)	3(3-0-6)
1421 410	ศิลปะการใช้เสียงและการสื่อความ (Speech Communication)	3(2-2-5)
1421 441	นิยายวิทยาศาสตร์ (Science Fiction)	3(3-0-6)
1421 342	วรรณกรรมและภาพยนตร์ (Literature and the Screen)	3(2-2-5)
1421 334	ภาษาศาสตร์เชิงจิตวิทยา (Psycholinguistics)	3(3-0-6)
1421 336	การวิเคราะห์วาทกรรมเบื้องต้น (Introduction to Discourse Analysis)	3(3-0-6)
1421 337	วรรณกรรมสำหรับเด็กและเยาวชน(Children's and Young Adults'Literature)	3(3-0-6)
1421 301	ภาษาอังกฤษเพื่อการท่องเที่ยว (English for Tourism)	3(3-0-6)

ข. กลุ่มวิชาโท   **24 หน่วยกิต**

ให้เลือกเรียนกลุ่มวิชาโทที่เปิดสอนในมหาวิทยาลัยอุบลราชธานี จำนวนไม่น้อยกว่า 24 หน่วยกิต หมวดวิชาเลือกเสรี
ไม่น้อยกว่า 3 หน่วยกิต นักศึกษาสามารถเลือกเรียนรายวิชาที่มีการเปิดสอนโดยคณะต่างๆ ภายในมหาวิทยาลัยอุบลราชธานี จำนวน
ไม่น้อยกว่า 3 หน่วยกิต
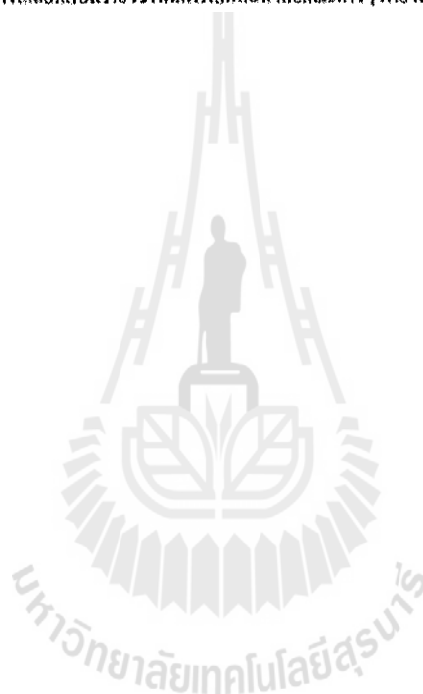
แผนการศึกษา

ชั้นปีที่ 1 (First Year)

ภาคการศึกษาต้น (First semester)

1101 141	วิทยาศาสตร์ในชีวิตประจำวัน (Science in Everyday Life)	3(3-0-0)
1421 100	ภาษาอังกฤษพื้นฐานสำหรับนักศึกษาศิลปศาสตร์ (Foundation English for Liberal Arts Students)	6(5-2-11)
1411 101	ภาษาไทยกับการสื่อสาร (Thai Language and Communication)	3(3-0-6)
1441 100	มนุษย์กับสังคม (Man and Society)	3(3-0-6)
1432 100	มนุษย์กับอารยธรรม (Man and Civilization)	3(3-0-6)

รวม (Total) 18 หน่วยกิต (Credits)

ภาคการศึกษาปลาย (Second Semester)

1101 140	มนุษย์กับสิ่งแวดล้อม (Man and Environment)	2(2-0-0)
1421 101	ภาษาอังกฤษสำหรับนักศึกษาศิลปศาสตร์ (English for Liberal Arts Students)	6(5-2-11)
1447 200	มนุษย์กับการสื่อสาร (Man and Communication)	3(3-0-6)
1431 100	มนุษย์กับสุนทรียภาพ (Man and Aesthetics)	3(3-0-6)
1439 xxx	กิจกรรมพลศึกษา (Physical Education Activity)	1(0-2-1)
	ให้ถือหน่วยเรียนไม่น้อยกว่า 2 หน่วยกิต ในรายวิชา	
1103 126	ดาราศาสตร์เบื้องต้น (Introduction to Astronomy)	3(3-0-3)
1104 160	คอมพิวเตอร์ทั่วไป (General Computer)	2(2-0-2)
1104 161	การใช้ซอฟต์แวร์ชุดสำนักงาน (Office Software Usage)	2(1-2-2)
1104 162	การสร้างเวบขั้นพื้นฐาน (Basic Web Development)	2(1-2-2)

รวม (Total) 17 หน่วยกิต (Credits)

ชั้นปีที่ 2 (Second Year)

ภาคการศึกษาต้น (First Semester)

1411 240	ภาษากับวรรณกรรมไทย (Thai Language and Literature)	3(3-0-0)
1431 100	ภูมิปัญญาและพฤติกรรมมนุษย์ (Human Wisdom and Behaviour)	3(3-0-0)
1421 210	การพูด 1 (Oral Communication I)	3(3-0-0)
1421 230	ภาษาเบื้องต้น (Introduction to Language)	3(3-0-0)
1421 231	สัทศาสตร์ภาษาอังกฤษ (English Phonetics)	3(2-2-1)
1421 240	วรรณคดีเบื้องต้น (Introduction to Literature)	3(3-0-0)

รวม (Total) 18 หน่วยกิต (Credits)

ภาคการศึกษาปลาย (Second Semester)

1708 100	เศรษฐศาสตร์ทั่วไป (General Economics)	3(3-0-1)
1432 101	วัฒนธรรมไทยในอดีตและปัจจุบัน (Thai Culture Past and Present)	3(3-0-0)
1421 211	การพูด 2 (Oral Communication II)	3(3-0-0)
1421 220	การเขียน 1 (Written Expression I)	3(3-0-0)
1421 241	การอ่านวรรณกรรม (Reading Literature)	3(3-0-0)
XXXX XXX	วิชาโท	3(3-0-0)

รวม (Total) 18 หน่วยกิต (Credits)

ชั้นปีที่ 3 (Third Year)

ภาคการศึกษาต้น (First Semester)

1421 221	การเขียน 2 (Written Expression II)	3(3-0-0)
1421 321	การแปล 1 (Translation I)	3(3-0-0)
1421 330	ภาษาศาสตร์กับการสื่อภาษา (Linguistics and Language Use)	3(3-0-0)
1421 340	แนวคิดในวรรณกรรม (Themes in Literature)	3(3-0-0)
XXXX XXX	วิชาโท	3(3-0-0)
XXXX XXX	วิชาโท	3(3-0-0)

รวม (Total) 18 หน่วยกิต (Credits)

ภาคการศึกษาปลาย (Second Semester)

1421 320	การเขียน 3 (Written Expression III)	3(3-0-0)
1421 322	การแปล 2 (Translation II)	3(3-0-0)
1421 331	การสื่อสารในกลุ่มชนข้ามวัฒนธรรม (Cross Cultural Communication)	3(3-0-0)
1421 341	วรรณกรรมกับศิลปะแขนงอื่น (Literature in Relation to Other Artistic Forms)	3(3-0-0)
XXXX XXX	วิชาโท	3(0-0-0)
1421 xxx	วิชาเอกเลือก1 (Major Elective I)	3(0-0-0)

รวม (Total) 18 หน่วยกิต (Credits)

ชั้นปีที่ 4 (Fourth Year)

ภาคการศึกษาต้น (First Semester)

1421 430	ลีลาภาษาและการสื่อสาร (Language Style and Communication)	3(3-0-6)
1421 440	แนวคิดโลกร่วมสมัย (Contemporary World Views)	3(3-0-6)
1421 xxx	วิชาเอกเลือก 2 (Major Elective II)	3(x-x-x)
1421 xxx	วิชาเอกเลือก 3 (Major Elective III)	3(x-x-x)
xxxx xxx	วิชาโท	3(x-x-x)
xxxx xxx	วิชาโท	3(x-x-x)

รวม (Total) 18 หน่วยกิต (Credits)

ภาคการศึกษาปลาย (Second Semester)

1421 490	การศึกษาอิสระ (Independent Study)	3(3-0-6)
1421 xxx	วิชาเอกเลือก 4 (Major Elective IV)	3(x-x-x)
1421 xxx	วิชาเอกเลือก 5 (Major Elective V)	3(x-x-x)
xxxx xxx	วิชาโท	3(x-x-x)
xxxx xxx	วิชาโท	3(x-x-x)
xxxx xxx	วิชาเลือกเสรี (Free Elective)	3(x-x-x)

รวม (Total) 18 หน่วยกิต (Credits)
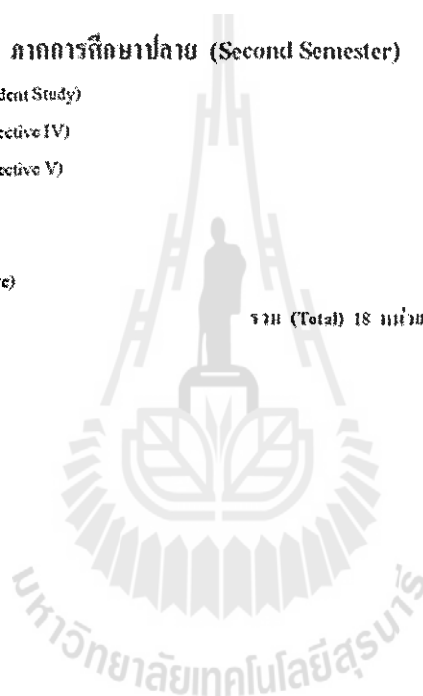
APPENDIX B

Picture Story Telling Tasks

Picture-based narrative writing tasks

The main instrument of this study is picture-based narrative writing tasks consisting of three popular fables: Little Red Riding Hood, Little Three Pigs, and Cinderella. Each task begins with directions, in Thai, including pictures and number in order to help the subjects writing with ease.

Example:

Directions: Write a story based on the given pictures with 250 words or a page long. Dictionaries are allowed when writing. (Time duration: 1 hour)

In addition, at the end of each story is vocabulary which might be unknown so that the subject will not have any obstacles about word selections.

Subject's bio-data

At the beginning of the data collection the subjects have to fill in their own bio-data in the same format as following:

Student's information   Code#_____

Name _____ Year ☐ 1 ☐ 2 ☐ 3 ☐ 4

Gender ☐ Male ☐ Female

Nationality   ☐ Thai ☐ Other _____

Mother's nationality ☐ Thai ☐ Other _____

Father's nationality ☐ Thai ☐ Other _____

Mother tongue   ☐ Thai ☐ Other _____

Experience in English native speaking countries: ☐ Yes ☐ No

The First Picture Story Telling Task:

Little Red Riding Hood

คำสั่ง จงเขียนเรื่องจากรูปภาพที่กำหนดให้ความยาวอย่างน้อย 250 คำมีคำศัพท์ช่วยในการเขียน
อยู่ด้านหลังและนักศึกษาสามารถใช้พจนานุกรมช่วยในการเขียนเรื่องได้ (เวลา 1 ชั่วโมง)

(1)

(2)

(3)

(4)

(5)

(6)

(7)

(8)

(9)

(10)

(11)

(12)

(13)

(14)

(15)

(16)

(17)

(18)

(19)
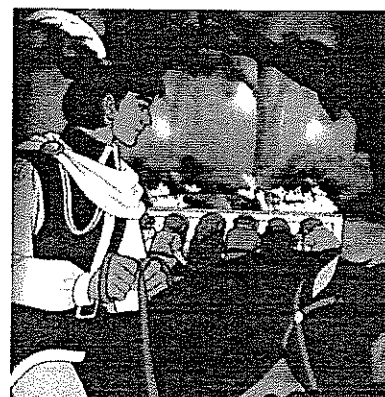
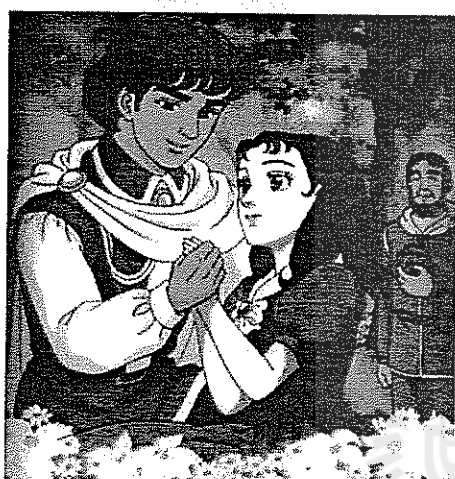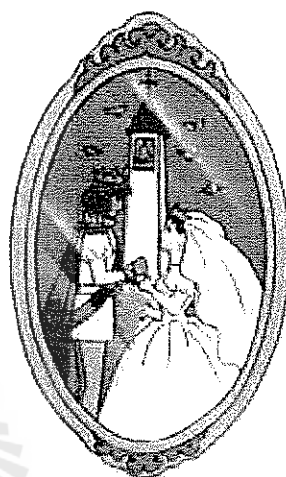(20)

(21)
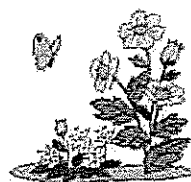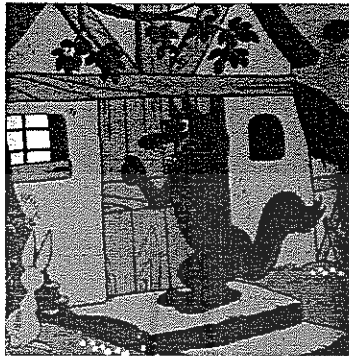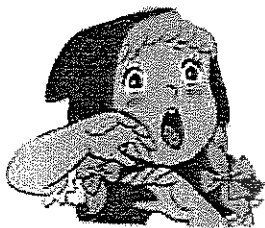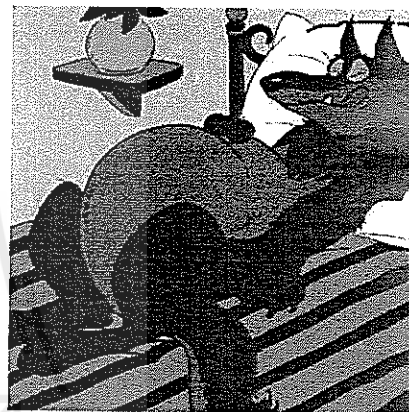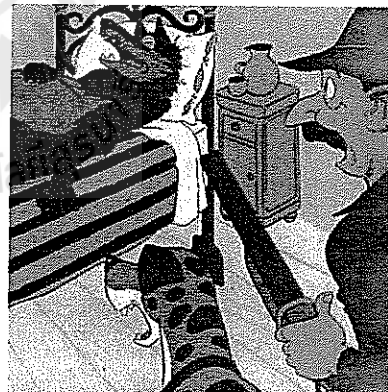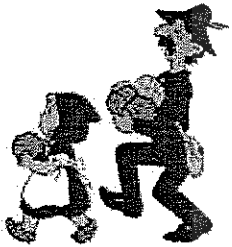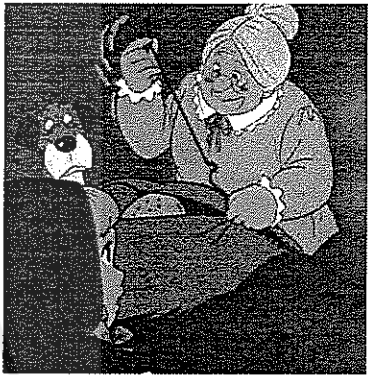
(22)

(23)

(24)
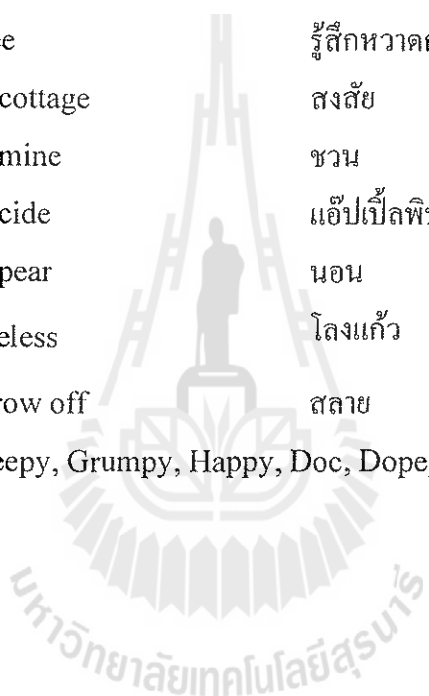

(25)

(26)
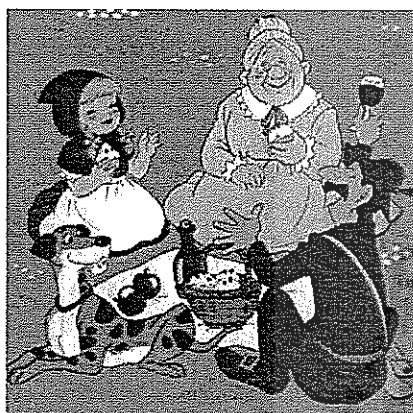
(27)

(28)

(29)

(pictures from http://sukumal.brinkster.net/i_story/akasukin/akasukin01.html literally
permitted to be used for academic purposes)

คำศัพท์ที่อาจต้องใช้ในการเขียน

หนูน้อยหมวกแดง	: Little Red Riding Hood	ตะกร้าอาหาร	: a food basket
เก็บดอกไม้	: to pick flowers	หมาป่า	: a wolf
ปรากฏ	: to appear	เคาะ	: to knock
กระท่อม	: a cottage	ดัดเสียง	: to imitate one's voice
หมวกคลุมผม	: a nightcap	เขี้ยว	: frank
ผ้าห่ม	: a blanket	กลืน	: to gobble
สงสัย	: to wonder	อิ่ม	: to be full
ลิ้น	: tongue	กรรไกร	: scissors
นายพราน	: a hunter	เข็ม	: needle
ตัด	: cut	ด้าย	: thread
ก้อนหิน	: rock(s)	ท้อง	: stomach, belly
ตื่น	: to wake up	เย็บ	: to saw
หัวคะมำ	: to tumble into	กระหายน้ำ	: to be thirsty
บ่อน้ำ	: a pond	แอบซุ่มดูต้นไม้ใหญ่	: to hide behind a big tree
เค้ก	: cake	หล่น	: to fall down
ตะกร้าผลไม้	: a fruit basket	ฉลอง	: to celebrate
น้ำผลไม้	: fruit juice		

The Second Picture Story Telling Task:
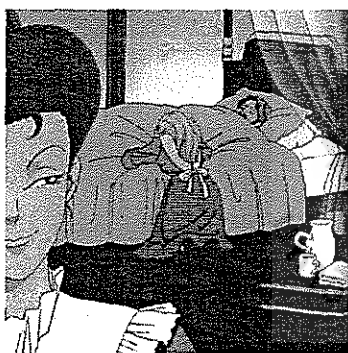
Snow White and the Seven Dwarfs

คำสั่ง จงเขียนเรื่องจากรูปภาพที่กำหนดให้ความยาวอย่างน้อย 250 คำ มีคำศัพท์ช่วยในการเขียน
อยู่ด้านหลังและนักศึกษาสามารถใช้พจนานุกรมช่วยในการเขียนเรื่องได้ (เวลา 1 ชั่วโมง)

(1)

(2)

(3)

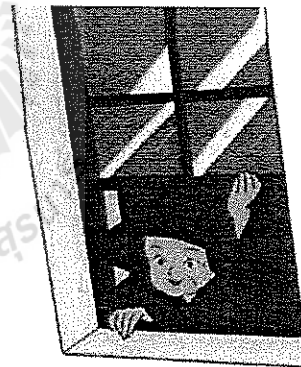(4)

(5)

(6)

(7)

(8)

(9)

(10)

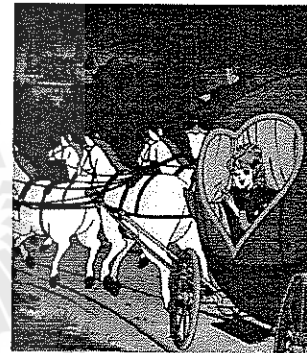

(11)

(12)

(13)

(14)


(15)

(16)

(17)

(18)

(19)

(20)

(21)

(22)

(23)

(24)

(25)

(26)

(27)

(28)

(29)

(30)
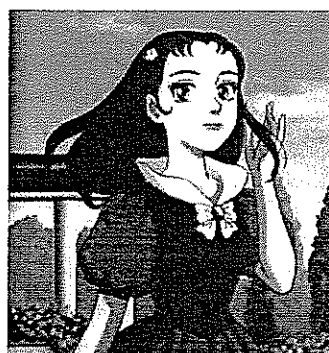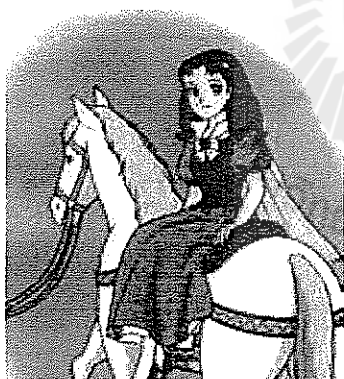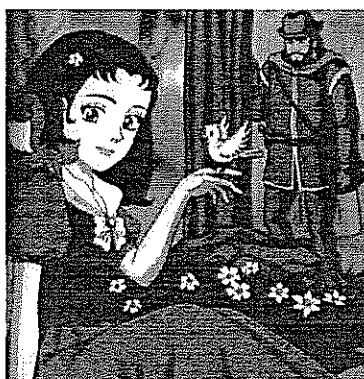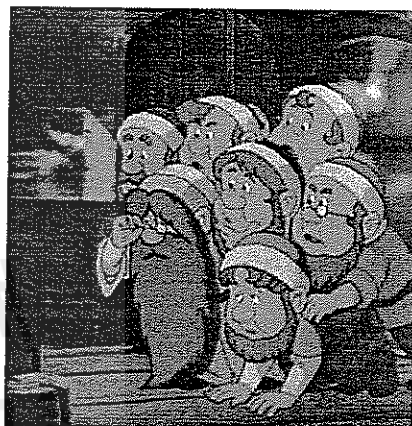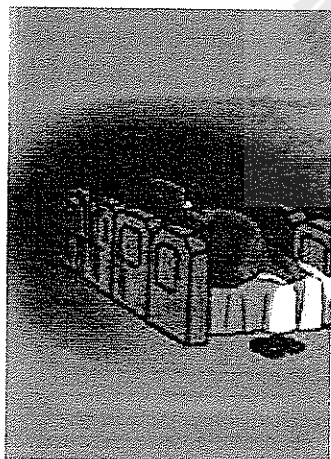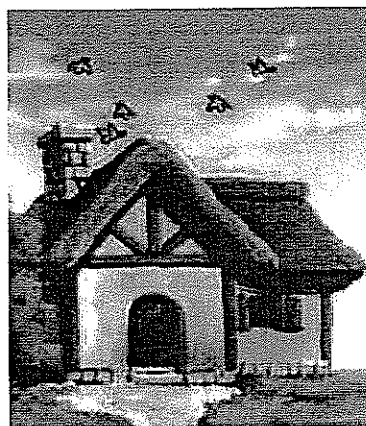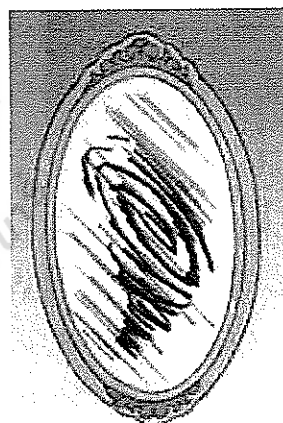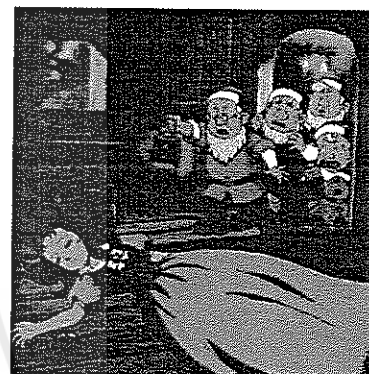
(31)

(32)

(33)

(34)

(35)

(36)
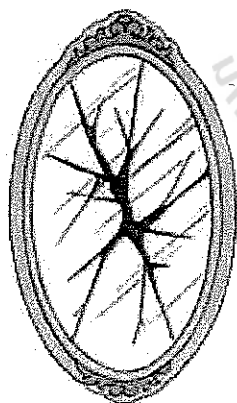
(37)

(38)
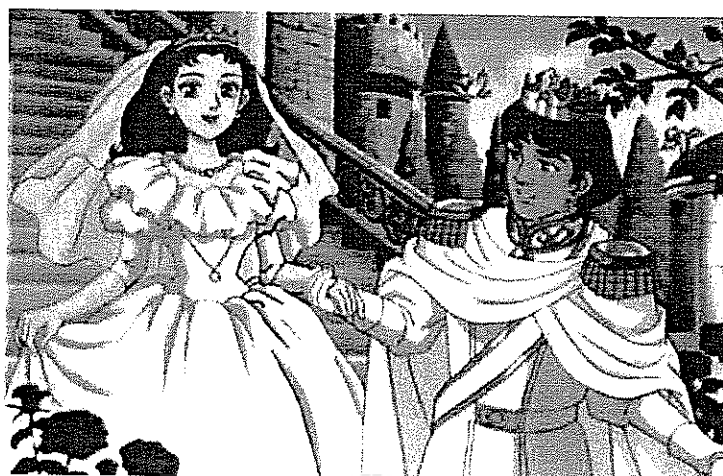
(39)

(pictures from http://sukumal.net/i_story/shirayukihime/yukihime01.html literally
permitted to be used for academic purposes)
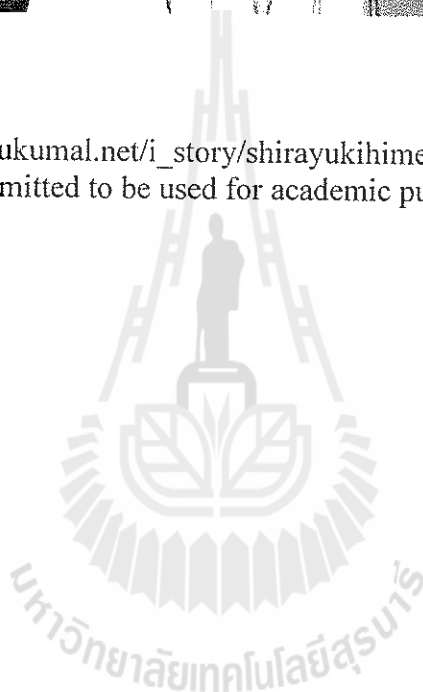
คำศัพท์ที่อาจต้องใช้ในการเขียนเล่าเรื่อง

สโนว์ไวท์	: Snow White	คนแคระ	: dwarf
ความปรารถนา	: desire	แดงชาด/แดงสด	: red as rose
ดำขลับ	: black as coal	แม่มดร้าย	: evil witch
ปลอมตัว	: disguise	กระจกวิเศษ	: a magic mirror
หลงใหล	: infatuated with	ความงาม	: beauty
พอใจ	: content	อิจฉา	: jealous
สั่ง	: command	เพลิดเพลิน	: enjoy
หนี	: flee	รู้สึกหวาดกลัว	: frighten
กระท่อม	: a cottage	สงสัย	: wonder
เหมือง	: a mine	ชวน	: invite
ตัดสินใจ	: decide	แอ๊ปเปิ้ลพิษ	: poisoned apple (s)
ปรากฏตัว	: appear	นอน	: lye
ไร้ลมหายใจ	: lifeless	โลงแก้ว	: a gold and glass coffin
กระเด็น	: throw off	สลาย	: decay/decompose

*ชื่อของคนแคระได้แก่ Sleepy, Grumpy, Happy, Doc, Dopey, Sneezy, and Bashful

The Third Picture Story Telling Task:

Cinderella

คำสั่ง จงเขียนเรื่องจากรูปภาพที่กำหนดให้ความยาวอย่างน้อย 250 คำ มีคำศัพท์ช่วยในการเขียน
อยู่ด้านหลังและนักศึกษาสามารถใช้พจนานุกรมช่วยในการเขียนเรื่องได้ (เวลา 1 ชั่วโมง)

(1)

(2)

(3)

(4)

(5)

(6)

(7)

(8)

(9)

(10)

(11)

(12)

(13)

(14)

(15)

(16)

(17)

(18)

(19)

(20)

(21)

(22)

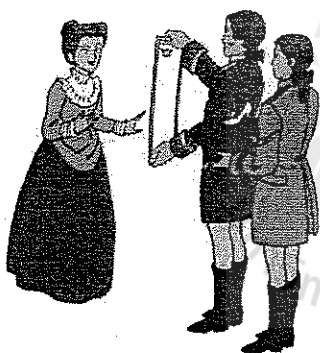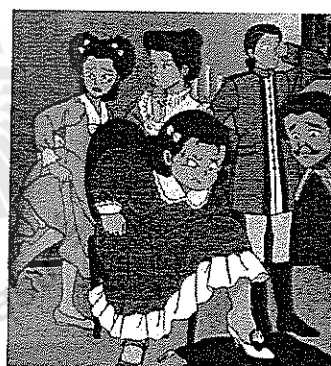(23)

(24)

(25)

(26)

(27)

(28)

(29)

(30)
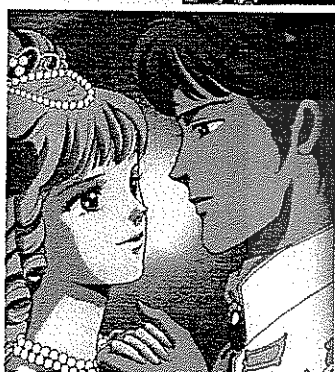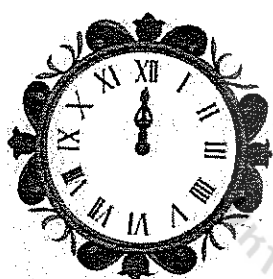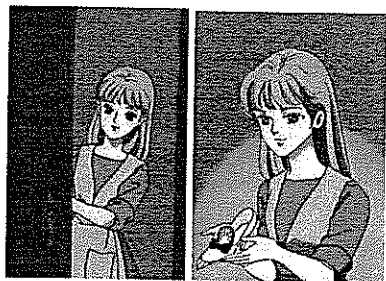

(31)
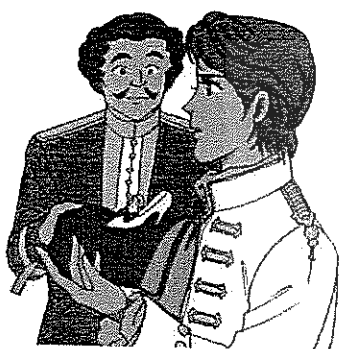
(32)

(33)

(34)

(35)

(36)

(37)

(38)

(pictures from http://www.sukumal.com/forum/view.php?qID=87 literally permitted to be used for academic purposes)
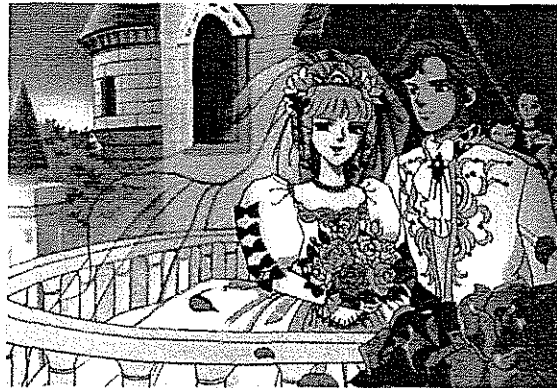
คำศัพท์ที่อาจต้องใช้ในการเขียนเล่าเรื่อง

ซินเดอเรลล่า	: Cinderella	เสียชีวิต	: passed away
แม่เลี้ยง	: stepmother	พี่เลี้ยงที่เป็นหญิง	: stepsister
แจ่มจรัสและสง่างาม	: splendid and elegant	ซุ่มซ่าม	: clumsy
เสแสร้ง	: pretend	ริษยา	: envy
ปฏิบัติต่อ	: treat	คนรับใช้	: servant
จดหมายเชิญ	: invitation letter	ชุดราตรียาว	: an evening dress
สวมเสื้อผ้าขาดวิ่น	: dress in rags	ตะปุ่มตะป่ำ	: lumpy
ร่ำไห้	: weep/sob	ปรากฏกาย	: appear
นางฟ้า/เทพธิดา	: a fairy	ส่องแสง	: a burst of light
รถม้า	: a coach	รองเท้าไม้	: clog
ตะหวัด	: a flick	เสก	: pronounce an incantation over
กลายเป็น	: become/turn into	ไม้กายสิทธิ์	: the magic wand
ผลฟักทอง	: a pumpkin	รถม้า	: carriage/equipage
รองเท้าแก้ว	: glass slippers	พระราชวัง	: the royal palace
ตกตะลึงอ้าปากค้าง	: stare with open mouth	ตกหลุมรัก	: fall in love
ฉงน	: doubt/wonder	ก่อนเที่ยงคืน	: before midnight
วิ่งหนี	: run away	ประกาศ	: announce
พระราชสาส์น	: Royal Letter	ลอง (สวม)	: try on
พอดี	: fit	อภิเษกสมรส	: royal wedding

CURRICULUM VITAE

Jarunee Anupan, an inborn Thai, received a Bachelor of Arts in English from Ubon Ratchathani Ratjabhat Universityn (UBRU) in 1999, and Master of Arts in TEFL (Teaching English as a Foreign Language) from Ubon Ratchathani University (UBU). She obtains a Degree of Doctor of Philosophy in English Language Studies in academic year 2013 from Suranaree University of Technology (SUT), Nakhon Ratchasima, Thailand.

She is currently a lecturer at the Faculty of Liberal Arts, Ubon Ratchathani University (UBU), Thailand. Her academic areas of interest mainly include second language acquisition, learner-corpora, and sociolinguistics.