Thematic structure analysis of Japanese advanced EFL college student writing

Hiroko Haenouchi
Tokyo Jogakkan College

Sonoko Ichinose
Keio University

Reference data:

This paper analyzed the thematic structure of Japanese advanced EFL college student writing to identify the textual features of higher-scoring and lower-scoring essays. As an analytical tool, the Theme-Rheme structure of Halliday's functional grammar was used. The thematic progression patterns were examined by focusing on what ideas were chosen for the sentence element that came first in each clause. The type of essays examined was a book report consisting of three sections, which explained why the author chose the book for the book report, what the story was about, and what impression the author gained after reading. The results showed that both of the groups used similar types of thematic patterns overall, but each essay group employed a different approach to developing themes in each of the three sections of the book report. The higher-scoring group showed more controlled use of different thematic progression patterns.

　この論文では、日本人大学生英語上級者の書いた英作文で、高い評価の作文とそうでない作文の特徴を明らかにするため主題構造の分析を行った。分析の手法として、ハリデイの選択機能文法の主題―題術構造を用いた。どのような考えが節の最初に置かれるかを調べることにより、作文の主題構造展開を調べた。分析した作文は３つの部分から構成される読書感想文で、なぜその本を選んだか、その本の内容、そして、読後感を述べるものであった。分析の結果、全体的には同じような主題構造パターンを使っていたが、作文の各構成部分では、異なる方法で主題を展開していることがわかった。評価の高い作文グループの方が様々なタイプの主題展開パターンを使いこなしていることがわかった。

Keiko Kawaguchi
Shibaura Institute of Technology

IN THE teaching practice of English writing in Japan, analytic scoring scales, the best-known one created by Jacobs, Zinkgraf, Wormuth, Hartfiel, and Hughey (1981), are widely used to assess written language. As a result, many EFL writing teachers still focus their feed-

back to students on errors of vocabulary and language use, which occur below the clause level. However, as Weigle (2002, p. 120) points out, it is often the case that if raters were to assign one combined score out of multiple analytic scores, experienced raters would adjust their final score based on the overall impression of the text, which involves the flow of information over a number of clauses.

This paper attempts to investigate what makes raters intuitively judge some texts better or more cohesive than others by focusing on the thematic structure of a text. Two groups of essays, higher-scoring and lower-scoring, written by advanced level Japanese college students were analyzed using the conceptual framework of Halliday's functional grammar (1994), specifically the Theme-Rheme structure, which contributes to the cohesive development of a text in which the flow of information follows certain patterns.

The Theme-Rheme structure is a grammatical resource that gives the clause its character as a message. "The Theme is the element which comes in first position in the clause" and "serves as the point of departure of the message" (pp. 37-38). The Theme is developed in the remaining part of the clause, which is called the Rheme. For instance, in the sentence *This book was written by a woman writer*, *this book* placed in the first position in the clause is the Theme, and the remaining part, *was written by a woman writer*, with which *the book* is concerned, is the Rheme. Thus, in Halliday's thematic structure, the Theme is identified by its position in the clause, not by the idea. There are three different types of elements that can come in the Theme position: textual Theme, interpersonal Theme, and topical Theme. The three Themes contextualize the clause to the three metafunctional aspects of the language: textual, interpersonal, and ideational. Textual Theme and interpersonal Theme precede topical Theme. For the current study, only topical Themes were investigated because these are most directly related to the flow of information.

A thematic progression pattern reveals how thematic elements succeed each other and how it makes a contribution to the cohesion and coherence of a text. Among the several types of thematic progression patterns, two of the main patterns are the Constant and the Linear Theme patterns (Bloor & Bloor, 2004) or in Eggins' (2004) terms, the Theme reiteration and the zigzag pattern. In the Constant Theme pattern, a common theme is shared by the very next clause as a Theme, whereas in the Linear Theme pattern, an idea of a Rheme is used as a Theme of the following clause. These two types of progression patterns create cohesion in a different way. It is generally known that the use of thematic progression is closely connected to the genre of the text.

There are not many ESL writing studies that have employed Halliday's Theme-Rheme structure. However, the following studies investigated thematic structure or idea development using similar concepts in an attempt to differentiate the qualities of student writing.

Connor and Farmer (1990) taught topical structure analysis as a revision tool for ESL students in intermediate- and advanced-level writing classes. As a result, they noticed great improvement in student writing, with regard to clearer focus and better development of subtopics by additional use of Constant and Linear Theme patterns. Sugiura (2000) examined four expressive essays written by one Japanese student over a period of four and a half years. He found that the student used the Constant Theme pattern extensively in the beginning; however, in her later work, she attempted to use the Linear Theme pattern. Sugiura concluded that the more diversified thematic progression reflected developmental changes of writing skills over a long period of time. Miyasako (2000) compared two groups of Japanese high school student essays and found that the lower group used more of the Constant Theme pattern. In contrast, the upper group used more of the Extended Constant pattern, in which a

Theme of a clause is used as the Theme of a distant clause in the same paragraph. However, the two groups showed no appreciable difference in the use of the Linear Theme pattern.

Study method

Subjects and materials

The texts examined for the present study were 16 book reports written by first-year Japanese students attending an advanced English class of the faculty of letters in a Japanese university.

The book reports were written for end-of-term papers. The length of the report was to be 400-600 words and its content to include the reason for choosing the book, a summary of the book, and the student's response to the reading. The books were chosen according to each student's major interest as side readers for the semester, which resulted in a variety of book genres: literature classics, contemporary works, fantasy stories, and biographies.

The number of book reports collected totaled 32 which were all rated by the three authors of this study on a holistic scoring system; a revised version of the TOEFL writing scoring guide (see Appendix 1 for the revised scoring guide). The scoring scale ranged from 0 to 6. Factors of the scoring guideline included addressing the writing task, organization and development, supporting details, consistent facility in language use, syntactic variety, and appropriate word choice. With regard to the inter-rater reliability of the three authors for the essays analyzed, the correlation coefficients between the raters were .772, .806, and .702 respectively.

The 32 reports ranged from levels 3 to 6 and were then divided into two groups; Group A consisting of higher-scoring papers marking levels 5 and 6 and Group B, lower-scoring papers marking levels 3 and 4. Not all 32 collected reports were used for this study, as some were not suitable for thematic progression analysis in their writing formats. Those that were discounted included reports on collections of short stories or non-fiction

books, and reports that did not meet the assignment's requirement to include the three sections of a book report: reason of choice, summary, and impression. Reports with no paragraph divisions and those heavily using quotations from the original book were also excluded. As a result, eight reports from each group were selected and the total number of reports used for the present analysis came to be 16.

Method of analysis

In order to investigate the types of Themes used in each text and their thematic progressions, Halliday's Theme-Rheme structure analysis was applied. First, each text was broken down into a list of clauses that were then divided into their Theme and Rheme parts. Themes were then categorized into three components: textual, interpersonal, and topical.

The topical Theme was further segmented into two types: unmarked and marked. Commonly, an unmarked topical Theme would be the subject of its clause, whereas a marked topical Theme would not, including the subject in its Rheme. A clause without a subject—a non-finite clause—is a Rheme Only clause. Table 1 is a sample segmentation format of data from one composition. There are two sentences, Sentence 13 with two clauses and Sentence 14 with four. Clauses 14.2 and 14.3 are Rheme Only clauses.

After segmenting the texts into clauses and distinguishing Theme from Rheme, patterns of idea development were categorized under seven types of thematic progression: Constant, Extended Constant, Linear, Extended Linear, Constant+Linear, Derived, and Independent.

The first type is the Constant Thematic Progression, in which a Theme of a clause is shared by the Theme of its consecutive clause. An example of this first type can be seen in Table 2, Clauses 3.2 and 4.1 where *this* and *this book*, both indicating the

Table 1. Text segmentation format

Clause No.	Theme				Rheme
	Textual	Inter-personal	Topical		
			Unmarked	Marked	
13.1			She		leaves in a hurry,
13.2	and		William		decides to forget her.
14.1				Over months that follow,	Anna shows up at his bookshop once again,
14.2					regretting the past,
14.3					hoping to resume their love affair,
14.4	but		William		turns her down.

same book, are used as subjects for each clause. The second type of thematic progression is the Extended Constant where a Theme of a clause is used as the Theme of a distant clause in the same paragraph, such as the *I* in Clauses 3.1 and 4.2 in Table 2. The third is the Linear Thematic Progression Pattern in which an idea of a Rheme becomes the Theme of the consecutive clause. For instance, *a book* in the Rheme position in Clause 1 is taken up as a Theme of Clause 2. The fourth pattern is the Extended Linear in which an idea of a Rheme becomes the Theme of a distant clause in the same paragraph. The fifth type is a combination, Constant & Linear Thematic Progression Pattern, where a Theme takes after a prior clause Theme and an idea of a prior Rheme. For instance, if the writer wrote *It changed my life* after Clause 1, *it* referring to the prior clause would be classified as Constant & Linear. The sixth pattern is the Derived Thematic Progression Pattern in which a Theme takes after an idea from a clause in a prior paragraph. The seventh, Independent Themes (Table 2, Clause 1&2) are new themes that appear in the text for the first time.

Table 2 is a thematic progression sample taken from a student composition showing some frequently used thematic patterns. Calculating the number of clauses and the Themes and their progression patterns by type, the present study compared the results between the higher-scoring Group A and the lower-scoring Group B. Comparisons were made with whole texts and also by separate parts of the texts: reason for book choice, summary of the book, and overall impression.

Results

Table 3 indicates the average number of clauses counted per essay for each group, the average numbers of unmarked and marked topical Themes and the average numbers of clauses without topical Themes classified as Rheme Only. The average number of clauses per essay was 66.8 for Group A, the higher-scoring group, and 64.3 for Group B, the lower-scoring group, with more varieties in clause length in the lower-scoring group's essays ranging from 47 to 93 compared to that in the higher-scoring group essays from 50 to 77 clauses. The ratios of the use of marked Themes, unmarked Themes and Rheme Only are given in parentheses. Both of the groups used unmarked and marked Themes at a similar rate with slight difference in the use of non-finite clauses (Rheme Only) such as the *to infinitive* or the *-ing* or *–ed particle* placed in the latter part of a clause complex.

Table 2. Thematic progression sample

Clause No.	Theme				Rheme
	Textual	Inter-personal	Topical		
			Unmarked	Marked	
1			I		read a book "Anything You Say Can and Will be Used Against You."
2			This book		was written by a woman writer, Laurie Lynn Drummond.
3.1			I		chose this book
3.2	because		this		seems very exciting and thrilling.
4.1			This book		was recommended in the bookstore [where I went]
4.2	so		I		noticed
4.3	and		(I)		interested in this book.

Table 3. Average numbers of clauses and topical Themes per essay

	# of clauses	SD	Topical Themes				Rheme Only	SD
			Unmarked	SD	Marked	SD		
Group A	66.8	9.43	56.9 (85%)	9.27	7.5 (11%)	2.67	2.4 (4%)	2.33
Group B	64.3	15.18	55.5 (86%)	15.32	7.3 (12%)	1.98	1.5 (2%)	1.69

Tables 4 and 5 show the average numbers of different types of topical Themes counted per essay for each group as a whole and in its three sections: why the writer chose a book for the book report (the Reason section), what the story was about (the Summary section) and what impression they obtained from the book (the Impression section).

Table 4. Average numbers of Themes used in the whole essay and its three sections (Group A)

	Constant		Extended Constant		Linear		Extended Linear		Constant + Linear		Derived		Independent	
	count	SD	count	SD	count	SD	count	SD	count	SD	count	SD	count	SD
Whole	16.0	6.55	12.6	4.50	8.8	1.16	4.9	1.96	3.0	1.93	6.0	3.21	13.0	3.89
Reason	2.4	1.41	1.5	0.76	1.4	1.51	0.8	0.89	0.6	0.74	0.1	0.35	3.0	1.77
Summary	8.3	4.53	7.4	3.74	5.1	1.55	3.0	1.41	1.5	1.20	2.1	1.13	5.8	2.66
Impression	5.4	2.83	3.6	2.26	2.4	1.06	1.5	0.93	0.9	0.83	3.8	2.49	4.3	2.60

Table 5. Average numbers of Themes used in the whole essay and its three sections (Group B)

	Constant		Extended Constant		Linear		Extended Linear		Constant +Linear		Derived		Independent	
	count	SD	count	SD	count	SD	count	SD	count	SD	count	SD	count	SD
Whole	20.3	9.42	10.8	4.30	9.0	3.25	4.8	3.45	2.5	1.93	3.1	2.72	10.4	2.50
Reason	3.0	2.20	2.0	1.60	2.3	1.67	0.1	0.35	0.1	0.35	0.0	0.00	1.6	0.74
Summary	10.4	10.18	4.3	3.20	4.5	2.62	2.8	2.55	1.3	1.51	1.3	1.11	4.4	1.77
Impression	6.9	2.95	4.5	1.69	2.3	2.25	1.9	1.73	1.8	1.17	3.9	2.64	4.4	2.56

The percentages of Themes in Tables 4 and 5 are graphically shown in Figures 1 and 2. As the figures show, the overall use of different Themes between the two groups' essays ("Whole") was not appreciably different, with the highest rate of Constant Themes (Group A 24.9%, Group B 32.3%), then Independent Themes (A 20.2%, B 16.6%), Extended Constant Themes (A 19.6%, B 17.2%), Linear Themes (A 13.6%, B 14.4%) and Extended Linear Themes (A 7.6%, B7.6%). Although the distributional patterns were similar, it should be noted that Group B used fairly more Constant Themes (32.3%) than Group A (24.9%).

Although the types of thematic progression patterns were not much different between the two groups overall, diversity could be noticed when focusing on the three separate sections of the essay. Major differences between the two groups lie in the use of Constant Themes and Independent Themes. While Group A used Constant Themes at a similar rate (about 25%) throughout the essay, Group B used Constant Themes with varying ratios ranging from 27% to 36%, with the highest ratio (36%) in the Summary section. The opposite trend was observed for the use of Independent Themes; Group A used Independent Themes most heavily in the Reason section (31%) compared to those of the Summary and Impression sections (17% and 20%) while Group B used Independent Themes with a similar ratio ranging from 15% to 18 % throughout the essay. Both groups used

the most Derived Themes in the Impression section. However, Group A used more Derived Themes in the Summary and Impression sections than Group B. For Linear and Extended Linear Themes, Group A used them more constantly throughout the essay than Group B. From these results, it was indicated that although not apparent in the analysis of topical Themes for whole essays, each group of essays adopted a somewhat different approach to developing themes.
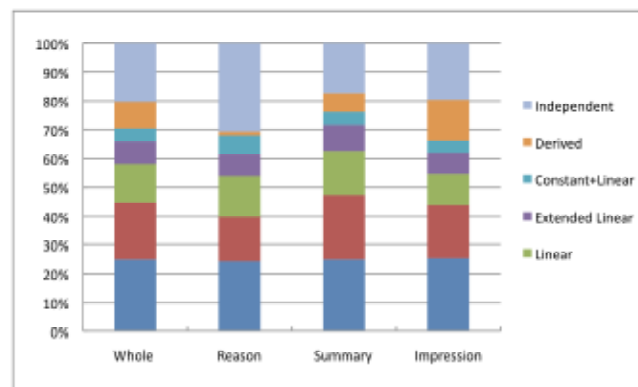

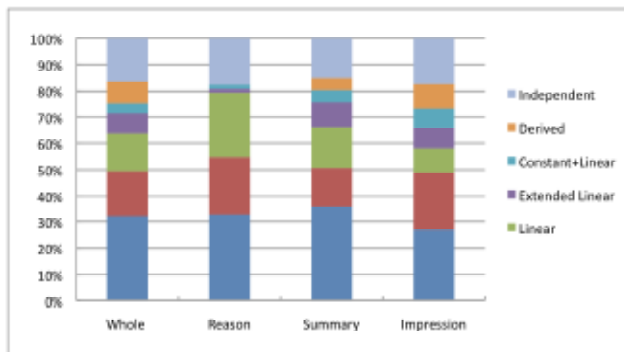Figure 1. Average ratios of different types of Themes (Group A)

Figure 2. Average ratios of different types of Themes (Group B)

Discussion

The results indicated that 1) Group B, the lower-scoring group, used more Constant Themes and 2) the two groups varied in the way that the different types of Themes were distributed across the different sections of the essay. The following section will discuss how differences in the use of Themes could be a factor that contributed to the evaluation of these essays.

The results of thematic structure analysis can suggest several characteristics of each group's essays. The use of Independent Themes indicates introduction of discrete ideas. Marked Themes are likely to be counted as Independent Themes, and the percentages of the overall use of marked Themes were similar between the two groups' essays (Group A 11%, Group B 12%) (Table 3). Nevertheless, the percentages of Independent Themes used for the whole essay, 20% for Group A and 17% for Group B, suggest that slightly more discrete ideas were introduced in Group A essays than Group B essays. Moreover, the two groups

used Independent Themes differently across the three sections of the essay. Group B used Independent Themes more or less at a similar rate in each section of the essay. In contrast, Group A used Independent Themes most heavily for the Reason section (31%). This is because Group A used more marked Themes mentioning when, where, or how the writer came to know the book while Group B typically used unmarked topical Themes, mostly *I* and occasionally *this book* or *this story* as shown in the sample compositions (see Appendix 2).

Derived Themes show ideas discussed beyond the boundaries of paragraphs. The number of Derived Themes naturally becomes larger towards the end of the essay if the writer discusses the same ideas throughout the essay. This makes the text more coherent. Group A essays using more Derived Themes especially in the Impression section (14.4%) compared to those of Group B (9.3%) indicates more ideas were discussed beyond the boundaries of the paragraphs in the higher-scoring group. For instance, the student writer of the sample composition from Group A (Appendix 2) introduced the key elements or themes dealt in the book in the Reason section and explained or expanded them later, thus resulting in more Derived Themes in the Impression section. In contrast, the student writer from Group B briefly explained why s/he chose a certain book in the Reason section and the Themes were never taken up again in the later sections.

The Constant thematic pattern, where a common Theme is shared by each clause, is an effective way to create cohesion. However, too much repetition of the same Theme would be boring to read and sometimes give an impression that the text is going nowhere (Eggins 2004, p. 324). On the other hand, the Linear thematic pattern, where the idea in a Rheme of one clause is taken up as the Theme of the subsequent clause, achieves cohesion in the text by building on newly introduced information. This gives the text a sense of cumulative development.

Considering such general effects of Constant and Linear thematic patterns, the extensive use of Constant Themes (32.3%) by Group B compared to the controlled use of Constant Themes (24.9%) by Group A across the essay could partly explain why Group B essays received lower evaluation. For instance, in the Summary section, Group B used Constant Themes more (35.9%) than Group A (24.9%) focusing on the actions of the protagonist. In contrast, the writers of Group A tended to include other key factors of the story in addition to the actions of the protagonist, thus breaking the pattern of successive use of Constant Themes by adding other types of Themes. A similar tendency was observed for the use of Linear Themes. Group A used Linear Themes at a similar rate throughout the essay, while Group B used Linear Themes heavily in the Reason section. The controlled use of Linear Themes across the essay might have contributed to the overall cohesion and gave the readers a sense of development of the text. Such controlled use of different types of thematic progression patterns are seen in Figure 1. Therefore, it could be inferred that the extensive use of Constant Themes, fewer Derived Themes, and less controlled use of Linear Themes by Group B might have led to lower evaluation of the essays because the text became monotonous and lacking in coherence.

Conclusion

This paper analyzed the thematic structure of two groups of essays, higher-scoring and lower-scoring, to identify what thematic structural features contributed to the evaluation of the essays. The findings showed that the two groups developed themes differently; the higher-scoring group used different types of thematic patterns in a balanced proportion throughout the essay, while the lower-scoring group extensively used Constant Themes, with other topical Themes used at a varying rate in different sections of the essay. It was concluded that these differences in the use of thematic progression patterns may have led to the qualitative differences of the two groups, especially in terms of coherence and content.

It seems that writers usually adopt different types of thematic progression patterns unconsciously to suit the different purpose of texts and make the text coherent. However, it is important for writing teachers to be aware of what type of thematic progression is preferred for a certain type of text and how such preferred thematic structure is realized by grammatical or structural resources. What is equally important is to provide the students with the opportunities to practice different types of thematic progression to have them recognize that the way they arrange ideas in a clause and the subsequent clause directly affects the development of themes, makes the text more coherent, and improves quality of their writing.

The present study analyzed just one type of text, a book report. Further thematic structure analyses of different types of text are desirable to reveal the types of thematic structure that make texts more coherent and suitable to their purposes.

Bio data

Keiko Kawaguchi has been with Shibaura Institute of Technology since 2005. She lives in Tokyo, and her research interests are testing, L2 writing assessment, and text analysis.

Hiroko Haenouchi has been with Tokyo Jogakkan College since 2004. Her research interests are L2 writing assessment and discourse analysis.

Sonoko Ichinose has been with Keio University since 2003. She lives in Tokyo, and her research interests are text analysis and reading comprehension.

References

Bloor, T., & Bloor, M. (2004). *The functional analysis of English* (2nd ed.). London: Arnold.

Connor, U., & Farmer, M. (1990). The teaching of topical structure analysis as a revision strategy for ESL writers. In B. Kroll (Ed.), *Second language writing* (pp. 126-139). Cambridge: Cambridge University Press.

Eggins, S. (2004). *An introduction to systemic functional linguistics* (2nd ed.). London: Continuum International Publishing Group.

Halliday, M. A. K. (1994). *An introduction to functional grammar* (2nd ed.). London: Arnold.

Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V., & Hughey, J. (1981). *Testing ESL composition: A practical approach.* Rowley, MA: Newbury House.

Miyasako, N. (2000). A topical structure analysis of high school student compositions. *JACET Bulletin, 31,* 49-63.

Sugiura, M. (2000). A thematic analysis of journal writing: A case study of a Japanese learner of English. *JACET Bulletin, 31,* 79-91.

Weigle, S. C. (2002). *Assessing writing.* Cambridge: Cambridge University Press.

Appendix 1

Essay scoring guideline (Levels 6-3)

An essay at…

- Level 6 effectively, Level 5 somewhat, Level 4 adequately, Level 3 inadequately addresses the writing task
    - » appropriate length (400-600 words)
- Level 6 is well, Level 5 is generally well, Level 4 is adequately, Level 3 is inadequately organized and well developed
    - » includes three factors: reason of book choice, summary, impression
    - » writing format – written in multiple paragraph form or divided sections (e.g. Introduction, Summary, Impression) and not in one extremely long paragraph or a list of sentences
    - » coherent ideas among paragraphs
- Level 6 uses clearly appropriate details, Level 5 uses details, Level 4 uses some details, Level 3 uses inappropriate or insufficient details to support a thesis or illustrate ideas
    - » written for any reader so that s/he can understand the report without having read the book (should include basic details of main characters or content when referring to specific points of the book)
    - » includes clear examples from the book and from personal experiences to explain an impression or idea
    - » should not include unnecessary information such as introducing of all the characters if not important in discussing an idea or impression
    - » should be written in the student's own words; unnecessarily long quotes from the book should be avoided
- Level 6 displays consistent facility, Level 5 displays facility, Level 4 demonstrates adequate but possibly inconsistent facility with syntax and usage, Level 3 demonstrates a noticeably inappropriate choice of words or word forms in use of language
    - » written for a writing task, not a speech script; should not be too colloquial (e.g. I'm going to tell you about…, it's very, very, very good!!! etc.)
    - » has good control of the target language (mechanics)
- Level 6 demonstrates syntactic variety and appropriate word choice though it may have occasional errors, Level 5 demonstrates some syntactic variety and range of vocabulary, though it will probably have occasional errors, Level 4 contains some errors that occasionally obscure meaning, Level 3 demonstrates an accumulation of errors in sentence structure and/or usage
    - » vocabulary use and variety

Appendix 2

Sample composition from the Reason section

Sample from Group A essay

Clause No.	Theme				Rheme
	Textual	Inter-personal	Topical		
			Unmarked	Marked	
1				As a side reader,	I chose "Momo," by the author of Michael Ende.
2.1				In this report	I tell you three things
2.2		why	why		I chose this book,
2.3		what	what		this book is about, and my impression of this book.
3			There		are two reasons I decided to read "Momo."
4				During a Japanese class in my high school,	I came across an essay on [how modern people's life is becoming dull and insipid because they like to save their time for efficiency.]
5.1	And			in that essay,	"Momo" was quoted
5.2					to explain the relationship between time and people.

Sample from Group B essay

Clause No.	Theme				Rheme
	Textual	Inter-personal	Topical		
			Unmarked	Marked	
1.1			I		chose this book in English.
1.2	because		I		have interested in this story for a long time.
2			This story		became a movie a few years ago.
3.1			I		wanted to see it
3.2	but		I		couldn't.
4			I		have wanted to see it someday.
5.1			I		found this book in a book store;
5.2	and		(I)		thought to try reading this story in English.