

## CHAPTER 4

### METHOD AND RESULTS

In this chapter, I will present the design of the empirical investigation of the thesis, which aims to investigate the topic-comment structures of Chinese EFL learners' interlanguage. In section 4.1, the background information of the subjects will be reported. In section 4.2, the methodology of the present study will be described. In 4.3 and 4.4, the test materials and procedures will be shown. Then in 4.5, the results will be displayed. Finally, in section 4.6, the main points of this chapter will be summarized.

#### 4.1 Subjects

Ninety four second-grade junior high school students participated in this study. They were randomly selected from three junior high schools in Taipei.<sup>1</sup> Besides, 14 native speakers of English served as the control group. They all studied Chinese in the language center of NCCU. Before our investigation, the subjects were required to fill out their background information (see Appendix A). The personal information and language background of the subjects are shown in Table 5 and Table 6.

---

<sup>1</sup> Originally 112 samples were obtained, but 18 samples of them were considered invalid since these subjects responded nothing to the questionnaire.

Table 5: Personal Information of the Subjects

Subjects		Chinese (n=94)	English control (n=14)	
Gender	M	62	11	
	F	32	3	
Age	14		Mean	26.43
			SD	8.27
			Range	20-50
Academic Level	Junior high		College student	7
			College graduate	5
			Graduate school	2
Nationality	Chinese in Taiwan		British	5
			U.S.A.	7
			Australia	1
			Turkey	1
L2	English		Chinese	

Table 6: Language Background of the Subjects

Subjects	Chinese (n=94)			English control (n=14)		
	Mean	SD	Range	Mean	SD	Range
Background						
Years of L2 learning	<b>2.5188</b>	1.2163	1.25-6.25	<b>2.589</b>	2.311	0.8-8
Years of staying in an L2 environment	0.02	0.2064	0-2	1.944	2.098	0.4-8

As suggested in table 5 and table 6, the percentage of male and female subjects was uneven in both groups. Interestingly, the percentage of male subjects was higher than that of female ones in both groups. This difference in gender rate may not affect the results, for all the subjects were randomly selected. Moreover, despite their differences in age and academic level, the Chinese group and English control group were similar in their average years of L2 learning. Since many Chinese subjects had

received L2 education in cram schools before they went to junior high school, I also considered such additional L2 training. That is, if the subjects had learned English for 5 years in cram school, then their total years of L2 learning is 6.25 years (5 years plus 1.25 years in junior high school). However, most Chinese group learners never stayed in an L2 environment. In contrast, the majority of the English group had stayed in L2 setting for more than one year (mean=1.944). This discrepancy in the source of L2 input between the two groups was not considered important in this study, in that I only searched for native speakers who could read and write a little Chinese.

In addition, I divided the Chinese subjects into three L2 proficiency group based on their first-grade English scores of final-terms, as displayed in table 7 and figure 7.

Table 7: L2 proficiency of Chinese subjects

	Average English scores in the first year	Number of subjects
Level 1	80-100	60
Level 2	60-80	23
Level 3	0-60	11

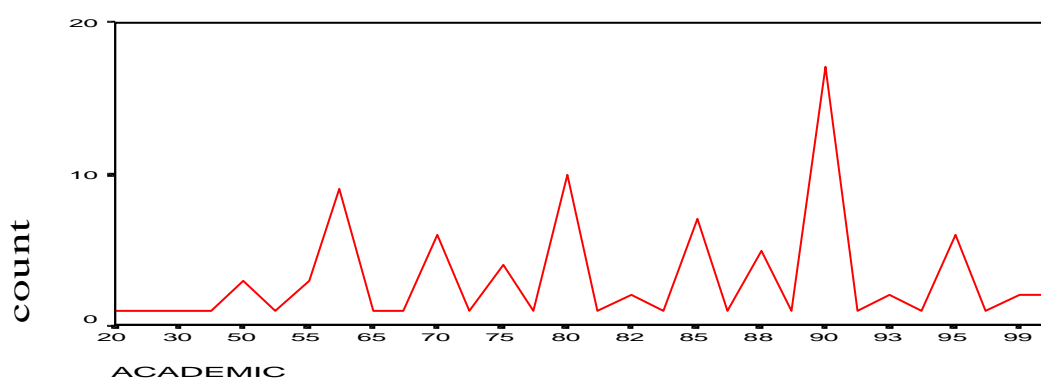


Figure 7: the distribution of academic scores

Note: The median is 84. (The median means “the centermost score in a distribution of scores arranged in sequence (Henning, 1987: 194)”).

## 4.2 Methodology

There are three broad types of methods to obtain data in SLA research: natural or elicited language use (in comprehension or production), metalingual judgment (e.g., grammaticality judgment), and self-report (Ellis, 1994). As reviewed in Chapter 2, in the studies of topic structures in L2 English learners' interlanguage, the most frequently used techniques are controlled language use (Rutherford, 1983; Green, 1996; Sasaki, 1990), spontaneous language use (Fuller & Gundel, 1987; Xiao, 1998;), and metalingual judgment tasks (Yuan, 1997; Yip, 1995). Most of these studies adopt only one kind of data, either comprehension or production one.

However, as noted by Corder (1981), production data should be obtained together with the intuition (i.e, comprehension) data to achieve descriptive adequacy in describing learners' interlanguage. In addition, Larsen- Freeman & Long (1991) state that subjects' performance varies from task to task, and different task requirements would yield different results. Therefore, judging learners' interlanguage elicited from single source is misleading and unconvincing.

In the present study, both comprehension tasks and production tasks were employed. The comprehension task was a grammaticality judgment task plus error correction (see appendix B). To avoid the problems of subjects' uncertainty and inconsistency of certain grammatical features, which are typical of grammaticality

judgment tasks (Ellis, 1991, 1994), I asked the subjects to circle the problematic part of a sentence if they judged it to be ungrammatical. In this way, their responses could be judged to be relevant or irrelevant to the tested features.

On the other hand, the production tasks were a guided writing task (see appendix C) and a translation task (see appendix D). The guided writing task was a revised version of Sasaki's (1990) controlled writing task, which elicits Japanese learners' existential constructions. This task required participants to write five sentences about the same topic based on the information given by a table. The other production task, translation task, asked the subjects to translate six Chinese sentences into English. In this way, all topic structures could be tested. The difference in guided writing task and translation task was that the former allowed relative free arrangement of the constituents but could not ensure that all structures could be elicited, while the latter was a more controlled elicitation task but is more likely to induce L1 transfer (Lococo, 1976). Thus, they could complement each other.

### **4.3 Materials**

Three tasks were conducted altogether via a questionnaire: grammaticality judgment, guided writing, and translation tasks. In these tasks, four topic structures were under investigation: double-subject construction (DS), left dislocation (LD), topic drop (DP), and topicalization (TOP). In this section the three tasks will be

described in detail. The comprehension task, or grammaticality judgment, is explained in 4.3.1, and then the production tasks--guided writing and translation--are introduced in 4.3.2 and 4.3.3 respectively.

### 4.3.1 The Grammaticality Judgment (GJ) Task

Each structure tested in the GJ task was exemplified by one or two ungrammatical sentences, and the corresponding grammatical sentence with respect to a certain topic-comment structure was also given. All these sentences were randomly ordered. There was a bracket beside each test sentence, with a line drawn below (see appendix B). If the subjects judged a test sentence to be right, they filled in the bracket with the symbol ?; if the sentence was wrong, the symbol was ?, and the subjects had to circle the problematic part, writing down the correct answer in the line drawn below the sentence. If they did not know the answer, they could just circle the wrong part and draw the symbol ?. The examples of the test sentences are shown in table 8.

Table 8: the test sentences of the Grammaticality Judgment Task

Tested features		Test sentences
1. DS	Grammatical	( ) In class, Bob's math is the best. _____
	Ungrammatical	DS(A) (×) My family, I am the youngest. <i>In my family...</i>
		DS(B) (×) His son, English is very good. _____ <i>son's</i>
2. LD(S)	Grammatical	( ) John is my friend. His father is a teacher. _____
	Ungrammatical	(×) My daddy, he plays basketball every day. <i>My daddy plays</i>
LD(O)	Grammatical	( ) The boy has big eyes. I think he is cute. _____
	Ungrammatical	(×) <del>The guy</del> , Mary saw <del>him</del> last week. <i>Mary saw the guy...</i>

Tested features		Test sentences
3.	Grammatical	( ) Susan saw her brother in the zoo. <b>She</b> was very happy. ___
	Ungrammatical	(× ) I have a girl friend. <b>Very</b> beautiful. <u>She is</u>
DP(S)	Grammatical	( ) The boy has big eyes. I think <b>he</b> is cute. ___
	Ungrammatical	(×) That restaurant is cheap and has good food, so we like <u>very</u> much. <u>it</u>
DP(O)	Grammatical	( ) The boy has big eyes. I think <b>he</b> is cute. ___
	Ungrammatical	(×) That restaurant is cheap and has good food, so we like <u>very</u> much. <u>it</u>
4. TOP	Ungrammatical	(×) <u>That book</u> I read last week. <u>I read that book last week.</u>

If the subjects judged a problematic sentence to be grammatical, they received a score of TP (topic prominent) and zero score of non-TP (non-topic-prominent). On the other hand, if they correctly rejected an ungrammatical sentence, they received zero score of TP and one score of non-TP. However, if the subjects judged an ungrammatical sentence to be ungrammatical but failed to locate the error or if they located the wrong position, they did not receive any TP or non-TP scores. Likewise, if the subjects judged a grammatical sentence to be incorrect and transformed it into an incorrect sentence to produce a topic-comment structure, they received a point of TP score of that structure. Nevertheless, if they transformed the correct sentences into the sentences other than topic structures, they did not receive any TP or non-TP scores. Thus, the total possible TP and non-TP score of each topic structure was one point.

#### 4.3.2 The Guided Writing (GW) Task

In guided writing task, all participants were required to write down five English sentences to describe a topic “the girl” based on the information provided by a table (see appendix C), as shown in table 9.

Table 9: the test sentences of the Guided Writing Task

Tested feature	Table information (topic: “the girl”)	
1. DS	1.eyes	big
	2.hair	long
	3.boy friend	handsome
2. TOP or DP or LD	4.many classmate see	go out with her boy friend
	5.her friends think	a happy girl

The rationale behind this task was to elicit the subjects’ responses under the pressure of two competing nominatives which could serve as the subject of a sentence—the topic “the girl” and the subject of a comment clause, being each test sentence in the table. The first three sentences were designed to elicit the DS construction, while the other two sentences were to induce TOP or DP or LS constructions. Some examples of DS structures were found in the Chinese learners’ data elicited by the first three sentences, as shown in (26)-(28), and those of “TOP or DOP or LD” structures elicited by the other two sentences, as displayed in (29)-(30).

(26) DS: \*The girl, eyes are big.

(Correct response<sup>2</sup>: The girl’ s eyes are big./ The girl has big eyes.)

(27) DS: \*The girl, hair is long.

(Correct response: The girl’ s hair is long./ The girl has long hair.)

(28) DS: \*The girl, boy friend handsome.

(Correct response: The girl’ s boy friend is handsome. / The girl has a handsome boy friend.)

(29) a. TOP: \*The girl, many classmates see go out with her boy friend.

---

<sup>2</sup> The possible correct responses were not limited to the examples listed here. As long as the subjects’ responses were considered grammatical in written English, these responses were judged to be correct.



b. DP: \*The girl go out with her boy friend and many classmates see.

c. LD: \*The girl, many classmates see her go out with her boy friend.

(Correct response: Many classmates see the girl go out with her boy friend.)

(30) a. TOP: \*The girl, her friends think a happy girl.

b. DP: \*The girl' s friends think a happy girl.

c. LD: \*Her friends think the girl, she is a happy girl.

(Correct response: Her friends think the girl is a happy girl.)

To compare the difficulty of the tested structures in each task, I also counted the TP scores and non-TP scores of each structure. The total TP and non-TP score of the first three sentences was 6 points, and so was that of the other two sentences. If the subjects correctly produced a sentence without the four topic-comment structures, they received 2 points of non-TP score for each DS sentence (1.2.3.) and 3 points of non-TP score for each TOP/DP/LD sentences (4.5.). On the other hand, if the subjects produced the four topic structures, they received the 2 points of TP score for each DS sentence (1.2.3) and 3 points of TP score for each TOP/DP/LD sentences (4.5.). I only focused on the four topic structures in this study, and ignored any irrelevant errors. Therefore, the irrelevant errors such as agreement or tense confusion, as well as the missing data (i.e., no response), received no scores. The total possible TP and non-TP scores of each topic structure produced by all subjects were 6 points. Thus, the scores needed to be divided by 6 in order to be comparable to the score of the other two tasks, both being one point.

### 4.3.3 The Translation (TR) Task

In translation task, all the four topic structures were tested with equal percentage, as they were in the other two tasks. Each test sentence was a Chinese sentence containing a tested topic structure. All participants were required to translate each test sentence into English. They were allowed to ask the spelling of words they did not know or leave a blank if they had problems with the vocabulary (see appendix D). The test sentences together with their tested features are shown in table 10.

Table 10: the test sentences of the Translation task

Tested features	Test sentences
1. DS	這個女孩，手上拿著一把雨傘。
2. LD(S)	我同學，他們都很喜歡去看電影。
LD(O)	那個男生，我很喜歡他。
3. DP(S)	他的爸爸是個老師。在我們學校教英文。
DP(O)	那枝筆很貴，你不要買。
4.TOP	晚餐我沒吃，我現在很餓。

The examples of the four structures found in our data are presented in (31)-(34).

(31) DS: \*The girl, hand takes an umbrella.

(Correct response: The girl is holding an umbrella in her hand.)

(32) LD(S): \*My classmate, they all like to go to the movies.

(Correct response: My classmates all like to go to the movies.)

LD(O): \*That boy, I like him very much.

(Correct response: I like that boy very much.)

(33) DP(S): \*His father is a teacher. In our school teach English.

(Correct response: His father is a teacher. He teaches English in our school.)

DP(O): \*That pen is expensive, you don' t buy.

(Correct response: That pen is expensive, you don' t buy it.)

(34) TOP: \*Dinner I don' t eat, I am not hungry now.

(Correct response: I didn' t eat dinner, I am hungry now.)

The scoring procedure of this task was similar to that of the grammaticality judgment task. If they correctly produced a sentence without topic structures, they received one point of non-TP score and zero point of TP score. In contrast, if they produced a sentence with a topic structure, they received one point of TP score and zero point of non-TP score for that structure. As in the previous tasks, I ignored all the irrelevant errors, which would receive no score and be considered equal to the missing data. The total possible TP and non-TP score of each topic structure was one point, as that in the other two tasks.

Since the total possible TP and non-TP score of each tested feature was the same in all the three tasks, the three tasks were comparable in terms of TP and non-TP mean scores. The grammaticality judgment task and the translation task were especially comparable, for their scoring procedure and tested features were exactly the same and they were both controlled in nature.

#### **4.4 Procedures**

The present study could be divided into three stages: the pilot study, the investigation, and the statistical treatment. These three stages will be described in

4.4.1, 4.4.2, and 4.4.3, respectively.

#### **4.4.1 Pilot Study**

To make the research more efficient, I conducted two pilot studies prior to the formal testing.

In the first pilot study, the questionnaire consisted of three parts: grammaticality judgment task, guided writing, and blank filling (see appendix E). The first part included 20 questions, 10 of which were all single sentences testing the four topic-comment structures (DS, LD, DP, TOP) and the other 10 are all dialogues containing two speakers, also testing the four structures. On the other hand, both the guided writing task and the blank filling task included 10 questions testing the four tested structures. I asked 30 second-grade junior high school students and one native speaker to finish this questionnaire, followed by an interview with the native speaker to discuss the test sentences.

The result suggested that in the grammaticality judgment task, the subjects showed no differences in their judgments of the tested features between the single-sentence questions and the dialogue questions. Besides, most non-native subjects (19 out of 30--63%) did not finish the third part: blank filling, partly due to time constraint and partly due to their limited ability to fill in the blanks with more than one word for each. Moreover, many students complained that the questionnaire

was too difficult for them and contained too many questions, so they could not finish it within an hour.

As for the native speaker, he finished the questionnaire in 20 minutes and judged all sentences with the four topic structures to be ungrammatical. However, the topicalization structure and left dislocation were elicited once in the blank filling task, as seen in (35).

(35) A: This lesson I' ve read (it) many times. But I still don' t understand it.

B: Me too. Let' s ask teachers our questions.

During the interview, this English subject said that topicalization and left dislocation are ungrammatical but acceptable in daily conversation. However, he did not personally prefer the topicalization structure since he thought this structure sounded archaic to him.

I made some improvements based on the drawbacks of the first pilot study, and then conducted a second one with a revised version of the questionnaire. This version was similar to that of the present one, with only 22 questions in total, but the second task did not have the number on the left of each question (see appendix F). I excluded the dialogue questions in the first task for the purpose of simplicity, since the first pilot study suggested no differences in the subjects' judgments of the same structure between the dialogue questions and single sentence questions. I asked one junior high school English teacher to proof-read this revised questionnaire to see if there were any

words students which would not know. Then, I asked two students to finish this questionnaire.

The second pilot study was more successful in that the two students finished this task within 15 minutes and they found no difficulty in finishing these questions, except for some confusion with the second task. For lack of numbers, the second task seemed to require them to arrange all the items in scrambled order. Therefore, I improved this problem by putting numbers in the second task and prepared to do the formal testing.

#### **4.4.2 Conducting the Investigation**

After the revision of this questionnaire, the formal test began from December 20, 2002 to January 6, 2003. I asked three teachers from different junior high schools in Taipei to gather the subjects from one or two classes. Besides, I found 14 native speakers of English from 7 classrooms in the language center of NCCU to finish this questionnaire.

The subjects were given a test sheet with a personal information survey in the back of each questionnaire. They were given 20 minutes to finish the tasks, but more time would be offered if they had not finished it. They could ask any questions if they were unclear of the instructions or if they had any vocabulary problem.

Moreover, to separate it from the structure cues given by the grammaticality

judgment task, the translation task was ordered after the guided writing task, which made students free of any influence from already given word orders.

#### **4.4.3 Statistical Treatment**

After the data collection process, all the samples' personal information and their responses to each task were transformed into SPSS files. The total TP and non-TP scores of each topic structures produced by all subjects in each task were counted. I compared the TP and non-TP mean scores of each structure in our subjects' responses. Then T-test, one-way ANOVA (Analysis of variance), and Scheffe post hoc multiple comparisons were conducted to test the significance of these values and group differences.

#### **4.5 Results**

In this section I will compare the results of the three tasks in four parts: (1) the TP mean scores of each topic structures responded by the Chinese group and the English group (2) the non-TP mean scores of each topic structures responded by Chinese group and the English group, (3) the hierarchy of difficulty among the four topic-comment structures, and finally (4) the responses to subjects and objects in LD and DP. We will present the results of the four parts in section 4.5.1, 4.5.2, 4.5.2, and 4.5.4 separately, and in the final section a brief summary of the results will be made.

### 4.5.1 TP Mean Scores in the Chinese Group and the English Group

The results of the TP mean scores (i.e., incorrect responses) were displayed in figure 8-10 and table 11 respectively.

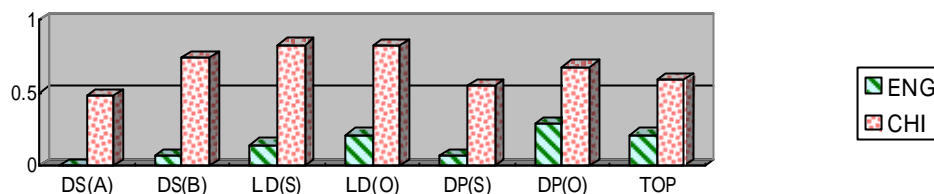


Figure 8-1: TP mean scores in the GJ task

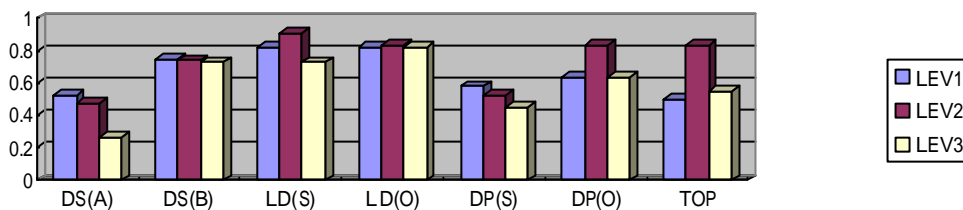


Figure 8-2: TP mean scores in the GJ task

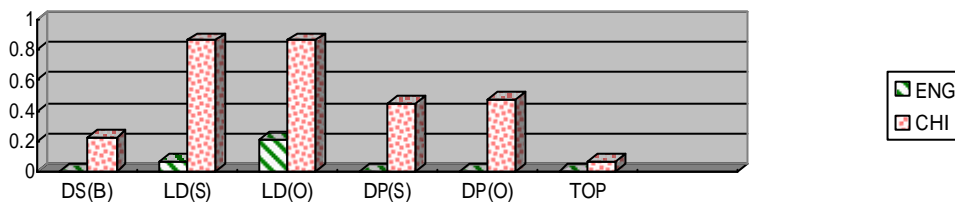


Figure 9-1: TP mean scores in the TR task

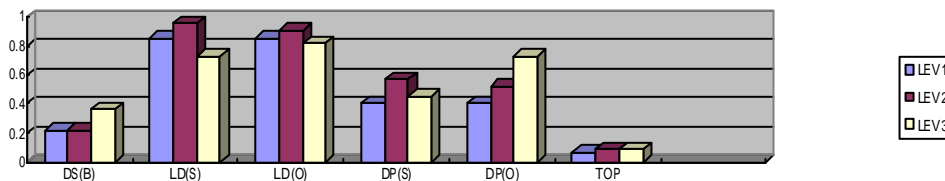


Figure 9-2: TP mean scores in the TR task



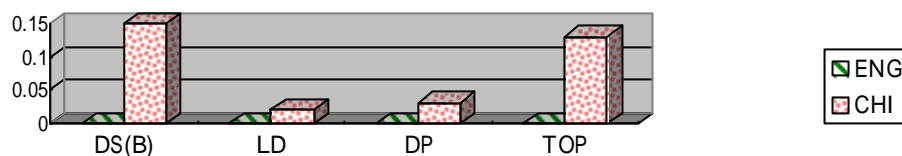


Figure 10-1: TP mean scores in the GW

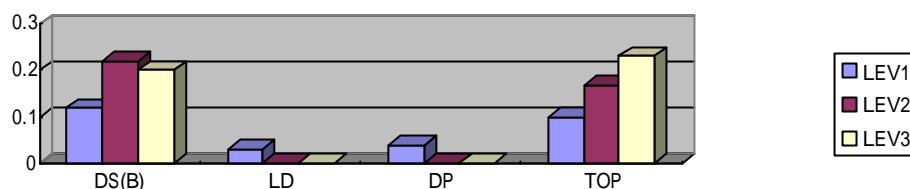


Figure 10-2: TP mean scores in the GW

Table 11: the hierarchy of mean TP scores in all groups and levels

Group	Task	TP mean score (Incorrect Responses)
English (n=14)	GJ	LD(O)>> DP(O)>>TOP >>LD(S) >> DP(S),DS(B)>> DS(A)
	TR	LD(O)>> LD(S)>> DS(B),DP(S), DP(O), TOP
	GW	All structures received no score
Chinese (n=94)	GJ	LD(S)>>LD(O)>>DS(B)>>DP(O)>>TOP>>DP(S)>>DS(A)
	TR	LD(S), LD(O)>>DP(S), DP(O)>>DS(B)>>TOP
	GW	DS(B), TOP>>LD, DP
Level 1 (n=60)	GJ	LD(S), LD(O)>>DS(B)>>DP(O)>>DP(S)>> DS(A)>>TOP
	TR	LD(S), LD(O)>>DP(S), DP(O)>>DS(B)>>TOP
	GW	DS(B)>>TOP>> DP>>LD
Level 2 (n=23)	GJ	LD(S)>>LD(O),DP(O),TOP >>DS(B)>>DP(S)>>DS(A)
	TR	LD(S)>>LD(O)>>DP(S)>>DP(O)>>DS(B)>>TOP
	GW	DS(B)>>TOP>> DP,LD
Level 3 (n=11)	GJ	LD(O)>>LD(S), DS(B)>>DP(O)>> TOP>>DP(S)>>DS(A)
	TR	LD(O)>> LD(S), DP(O)>>DP(S)>>DS(B)>>TOP
	GW	TOP>>DS(B)>> DP,LD

As can be observed easily, LD induced the most TP scores from the Chinese group in the first two tasks (GJ and TR), but the scores became the least in the third task (GW). Like LD, the TP scores of DP remained high in the first two tasks but

decreased sharply in the third task. It seemed that LD and DP incurred more incorrect responses in the controlled tasks but less in the relatively freer production task. On the other hand, DS(B) and TOP received more TP scores in the first task (GJ) than those in the other two tasks (TR and GW). It appeared that DS(B) and TOP were more likely to cause errors in the comprehension task than in the production tasks. Except for Chinese level 1, DS(A) elicited the least amounts of TP scores in the first task (GJ). It appeared that DS(A) was the easiest structure, at least in the comprehension task<sup>3</sup>. The T-test of significant differences between the Chinese group and English control was presented in table 12.

Table 12: T-test of TP scores between the Chinese group and English control

GR task	DS(A)	DS(B)	LD(S)	LD(O)	DP(S)	DP(O)	TOP
CHI-ENG	3.552**	5.580**	6.376**	5.387**	3.522**	2.943**	2.656**
LEV1-ENG	3.816**	5.559**	5.890**	5.114**	3.716**	2.481*	<u>1.961</u>
LEV2-ENG	3.484**	5.032**	7.144**	4.488**	3.044**	3.799**	<u>4.488**</u>
LEV3-ENG	2.198*	4.426**	3.524**	3.597**	2.385*	1.779	<u>1.748</u>
TR task		DS(B)	LD(S)	LD(O)	DP(S)	DP(O)	TOP
CHI-ENG		2.049*	8.155**	6.319**	3.331**	3.477**	1.051
LEV1-ENG		1.941	7.599**	5.745**	3.013**	3.013**	.986
LEV2-ENG		1.918	11.251**	5.963**	4.149**	3.801**	1.123
LEV3-ENG		2.713*	4.426**	3.597**	3.276**	5.861**	1.135
GW task		DS(B)	LD		DP		TOP
CHI-ENG		2.004*	2.033*		.968		1.735
LEV1-ENG		1.749	1.762		1.113		1.550
LEV2-ENG		2.362*	1.000		1.000		1.999
LEV3-ENG		2.503*	NA		NA		2.086*

\*p<0.05

\*\*p<0.01

<sup>3</sup> I did not expect the two types of the structure DS--DS(A) and DS(B)--would yield different results, so I only differentiated them in the GJ task.

As seen in table 12, it was clear that the comprehension task (GJ) induced more TP scores than did the production tasks, since there were more significant differences of TP scores between Chinese and English in the comprehension task than in the production tasks. Likewise, GW elicited the least amounts of TP scores since it showed the least significant differences between Chinese and English TP scores.

The earlier finding that LD and DP incurred significantly more TP responses in the controlled tasks (GJ and TR) than in the free writing task (GW) was also confirmed in table 12, in that all learners' responses to the two structures were significantly different from those of the English control in the former tasks (GJ and TR), but achieved far less significant differences in the latter task (GW), especially in level 1 and level 2.

On the other hand, the observation that TOP and DS(B) were found easier in the production tasks—TR and GW—than the comprehension task (GJ) was further confirmed in table 12, which indicates that, though students in level 2 showed significant differences of TP scores from the English control, all Chinese learners' TP responses to TOP in the TR task were not significantly different from those of the English control, and the same applied to the GW task, except for level 3; besides, TP responses of DS(B) in the TR task were not significantly different between Chinese level 1 and the English control, and between level 2 and the English control, and the

same applied to level 1 in the GW task.

Moreover, in table 12, I found a U-shaped pattern in TOP of the GJ task. That is, error rate was initially low in the beginner level, then rose in the intermediate level, and finally fell again in the advanced level (Kellerman, 1983, cited from Larsen-Freeman and Long, 1991: 106). This U-shaped pattern was suggested in the presence of significant differences in TOP error scores between Chinese level 2 and English control and the absence of them between Chinese level 1 and English control, and between level 3 and English control. In this task, LD(S) and DP(O) also showed this tendency of U-shaped pattern—students in level 2 produced higher TP scores than did those in level 1 and level 3. However, only TOP suggested significant differences between level 2 and the other level, as shown in table 13.

Table 13: one-way ANOVA of TP mean scores across Chinese levels (1, 2, 3)

GJ task	DS(A)	DS(B)	LD(S)	LD(O)	DP(S)	DP(O)	<b>TOP</b>
	F=1.099	F=.015	F=.999	F=.005	F=.364	F=1.477	<b>F=3.867*</b>
TR task	NA	DS(B)	LD(S)	LD(O)	DP(S)	DP(O)	TOP
		F=.572	F=1.745	F=.367	F=.908	F=2.208	F=.072
GW task	NA	DS(B)	LD		DP		TOP
		F=1.003	F=.278		F=.634		F=1.242

\*p<0.05

\*\*p<0.001

Note: GJ--TOP: Sheffe? (1,2) p=\*.025 2>>1

Table 13 also shows that all Chinese levels' responses in the three tasks were not significantly different except for TOP in the GJ task, in which students in level 2 produced significantly more TP scores than did those in level 1.

In sum, there were task effects in the results of TP scores. That is, different tasks

elicited different topic structures. LD and DP were more frequent in the two controlled tasks (GJ and TR) than in the free writing task (GW), LD being the most frequent in both tasks; DS(B) and TOP were more frequent in the comprehension task (GJ) than in the production tasks (TR and GW); DS(A) was the most infrequent TP structure in the comprehension task (GJ). On the other hand, there is an obvious U-shaped pattern found in the TOP in the comprehension task, where students in level 2 scored higher than did those in level 1 and 3. However, except for TOP in the comprehension task, there were no significant differences of TP scores across Chinese levels.

#### 4.5.2 Non-TP Mean Scores in Chinese Group and the English Group

The results of the non-TP mean scores (i.e., correct responses) were displayed in figure 11-13 and table 14 respectively.

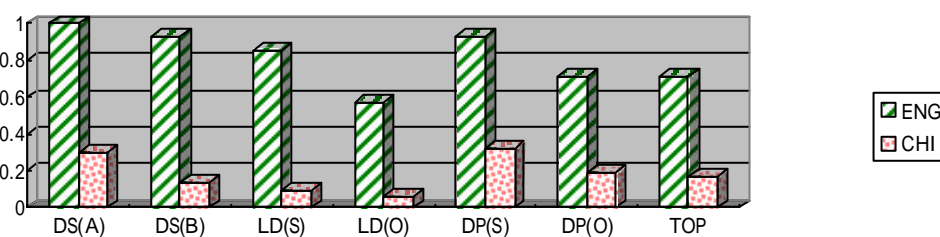


Figure 11-1: non-TP mean scores in the GJ task

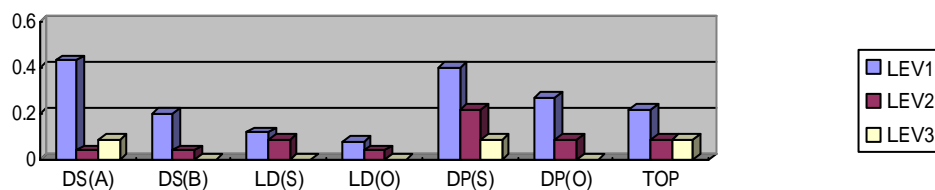


Figure 11-2: non-TP mean scores in the GJ task

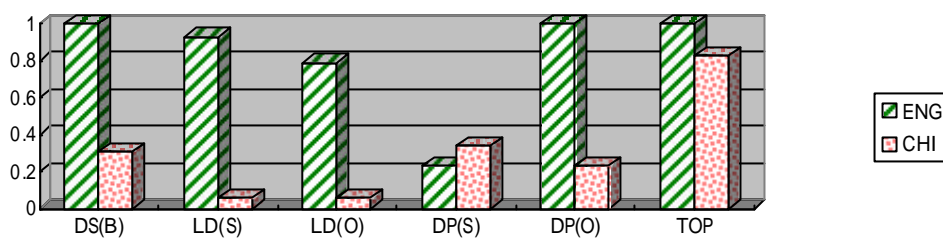


Figure 12-1: non-TP mean scores in the TR task

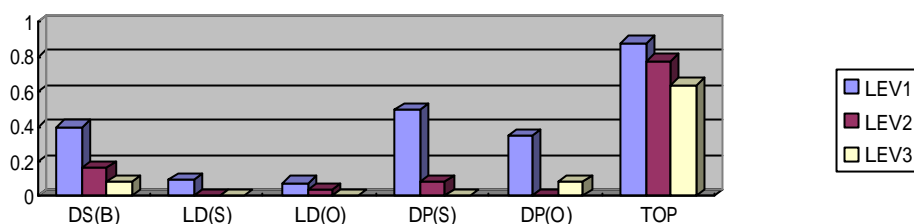


Figure 12-2: non-TP mean scores in the TR task

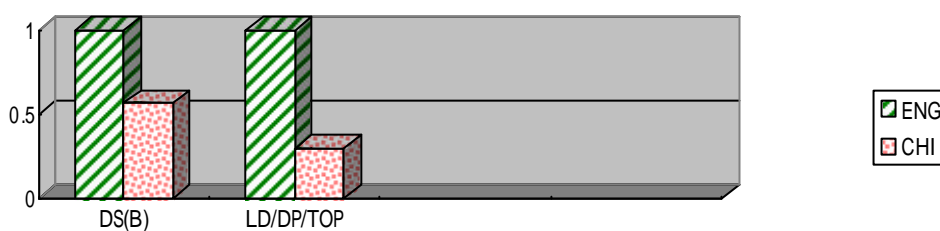


Figure 13-1: non-TP mean scores in the GW task



Figure 13-2: non-TP mean scores in the GW task

Table 14: the hierarchy of mean non-TP scores in all groups and levels

Group	Task	Non-TP mean score (Correct Responses)
English	GJ	DS(A)>> DS(B),DP(S)>>LD(S) >>DP(O) , TOP >>LD(O)
	TR	DS(B), DP(S), DP(O), TOP >>LD(S) >> LD(O)
	GW	All structures received the full scores: 2 points
Chinese	GJ	DP(S)>>DS(A)>>TOP>>DP(O)>>DS(B)>>LD(S)>>LD(O)
	TR	TOP>>DS(B)>>DP(S)>>DP(O)>>LD(S),LD(O)
	GW	DS(B)>>LD/DS/TOP

Group	Task	Non-TP mean score (Correct Responses)
Level 1	GJ	DS(A)>>DP(S)>>DS(B)>>DP(O)>>TOP>>LD(S)>>LD(O)
	TR	TOP>>DP(S)>>DS(B)>>DP(O)>>LD(S)>>LD(O)
	GW	DS(B)>> LD/ DP/TOP
Level 2	GJ	DP(S)>>DP(O)>>LD(S),TOP>>DS(A),DS(B),LD(O)
	TR	TOP>>DS(B)>>DP(S)>>LD(O)>>LD(S), DP(O)
	GW	DS(B)>>LD/ DP/TOP
Level 3	GJ	DS(A), DP(S) >>TOP>> DP(O)>>LD(S), DS(B),LD(O)
	TR	TOP>>DS(B), DP(O)>>LD(S),LD(O), DP(S)
	GW	LD/ DP/TOP>> DS(B)

As seen in figure 11-13, LD received the least non-TP scores from the Chinese group in all the tasks, except for level 3 in the GW task. It was likely that LD was the most difficult structure across tasks. On the other hand, non-TP scores of DP were similar across tasks. In contrast, like the results of TP scores, DS(B) and TOP appeared more difficult in the comprehension task (GJ) than in the production tasks (TR and GW), in that non-TP scores of them were lower in the former task than in the latter tasks. However, TOP seemed easier in the controlled production task while DS(B) easier in the free production task. Finally, I also found a U-shaped pattern in the responses to DS(A) in the comprehension task (GJ). That is, accuracy was initially high in the beginner level, then fell in the intermediate level, and finally rose again in the advanced level. This U-shaped pattern, as seen in figure 10, was suggested in the relatively higher non-TOP error scores in Chinese level 2 than those of level 1 and level 3.

The T-test of significant differences in non-TP scores between the Chinese group

and the English control is shown in table 15.

Table 15: T-test of non-TP scores between the Chinese group and English control

GJ task	DS(A)	DS(B)	LD(S)	LD(O)	DP(S)	DP(O)	TOP
CHI-ENG	-5.691**	-8.155**	-8.718**	-6.066**	-4.740**	-4.503**	-4.868**
LEV1-ENG	-4.221**	-6.419**	-7.533**	-4.930**	-3.860**	-3.350**	-3.939**
LEV2-ENG	-17.069**	-11.251**	-7.144**	-4.400**	-5.641**	-5.059**	-5.059**
LEV3-ENG	-11.349**	-11.470**	-7.792**	-3.673**	-7.355**	-5.030**	-3.823**
TR task		DS(B)	LD(S)	LD(O)	DP(S)	DP(O)	TOP
CHI-ENG		-5.550**	-12.147**	-9.187**	-5.160**	-6.706**	-1.679
LEV1-ENG		-4.520**	-9.416**	-7.622**	-3.691**	-5.030**	-1.341
LEV2-ENG		-7.931**	-16.818**	-7.116**	-11.792**	NA	-1.918
LEV3-ENG		-11.349**	-11.470**	-6.092**	NA	-11.349**	-2.713*
GW task		DS(B)	LD		DP		TOP
CHI-ENG		-3.892**			-6.667**		
LEV1-ENG		-3.104**			-5.064**		
LEV2-ENG		-5.471**			-17.532**		
LEV3-ENG		-8.920**			-23.833**		

\*p<0.05

\*\*p<0.01

Table 15 shows that all the non-TP scores of the English group were significantly higher than those of the Chinese group, and only TOP in the TR task revealed no significant differences between the Chinese group (except for level 3) and English control.

On the other hand, unlike the results of TP scores in which all levels showed no significant differences in most cases, table 16 indicated substantial significant differences of non-TP scores across levels.



Table 16: one-way ANOVA of non-TP mean scores across Chinese levels (1, 2, 3)

GJ task	DS(A)	DS(B)	LD(S)	LD(O)	DP(S)	DP(O)	TOP
	<b>F=8.390**</b>	F=2.783	F=.732	F=.634	F=2.849	<b>F=3.335*</b>	F=1.261
TR task		DS(B)	LD(S)	LD(O)	DP(S)	DP(O)	TOP
		<b>F=3.51*</b>	F=1.829	F=.634	<b>F=11.574**</b>	<b>F=7.163**</b>	F=2.285
GW task		DS(B)	LD		DP		TOP
		<b>F=13.70**</b>	<b>F=11.378**</b>				

\*p<0.05    \*\*p<0.01

Note: GJ--DS(A): Sheffe? (1.2) p=\*.002 **1>>2**

TR--DP(S): Sheffe ? (1.2) p=\*.001; (1.3) p=\*.003 **1>>2, 3**

DP(O): Sheffe ? (1.2) p=\*.003 **1>>2**

GW--DS(B): Sheffe ? (1.2) p=\*.001; (1.3) p=\*.000 **1>>2, 3**

LD/DP/TOP: Sheffe? (1.2) p=\*.001; (1.3) p=\*.005 **1>>2, 3**

Students in level 1 obtained significantly higher non-TP scores in DS(A) in the GJ task than did those in level 2, in DP(S) than did those in level 2 and level 3, in DP(O) than did those in level 2 in the TR task, and in all structures than did those in level 2 and level 3 in the GW task. Production tasks (TR and GW) tended more to elicit correct responses from level 1 than did the comprehension task (GJ), and the free production task (GW) induced more correct responses than did the controlled one (TR). This was consistent with our previous finding that TP scores were more prevalent in the comprehension task (GJ) than in the production tasks (TR and GW), with the controlled one (TR) eliciting more TP features than the free one (GW). I thus inferred that, among the three tasks, the comprehension task (GJ) induced the most incorrect responses and the least correct responses, and the free production task (GW) incurred the most correct responses and the least incorrect responses; the two responses of the translation task (TR) were in-between.

In sum, as in results of the TP scores, there were also task effects in the findings of non-TP scores. Except for LD, which was the most difficult structure across tasks, different tasks elicited correct responses in different structures. Like the results of TP scores, non-TP scores of DP remained constant in controlled tasks (GJ and TR). In contrast, DS(B) and TOP received more correct responses in the production task than in the comprehension task (GJ). Moreover, there was a U-shaped pattern found in DS(A) in the comprehension task (GJ), where students in level 2 scored lower than did those in level 1 and 3. Furthermore, there were significantly higher non-TP scores in level 1 than the other two levels in all tasks. However, production tasks (TR and GW), especially the free one (GW), induced more non-TP scores than did the comprehension task (GJ).

#### 4.5.3 The Hierarchy of Difficulty Among the Topic Structures

In order to determine the hierarchy of difficulty among the topic structures, the results of TP scores and non-TP scores were combined, as shown in table 17.

Table 17: outline of the results of TP and non-TP scores

Structure	LD(S)/LD(O)	DP(S)/DP(O)	DS(B)	TOP (U-shaped in TP)	DS(A)
TP	High in controlled tasks		More in comprehension task		Least in GJ
Non-TP	<b>Least in all</b>	Consistent	More in production tasks		U-shaped

Since LD induced the most TP scores in controlled tasks (GJ and TR) and the least non-TP scores in all tasks, it was considered the most difficult structure of all. In contrast, DS(A) incurred the least TP scores in the GJ task, so that it was considered

the easiest structure.

On the other hand, DS(B) and TOP incurred more TP scores in the comprehension task (GJ) and more non-TP features in production tasks (TR and GW), suggesting that the two structures were only difficult in the comprehension task and became easier in the production tasks. In contrast, TP scores in the DP remained high in the two controlled tasks, including comprehension one (GJ) and production one (TR). Moreover, non-TP scores in DP were also consistent across tasks. Therefore, DP was more difficult than DS(B) and TOP. In addition, as noted earlier, DS(B) was the easiest structure in the free production task (GW), and TOP the easiest in the controlled one (TR). It was difficult to differentiate the relative difficulty of these two structures, and thus I put them in the equal status of difficulty.

The hierarchy of difficulty is then presented in (36).

(36) the Hierarchy of Difficulty Among the TP Features:

LD(S), LD(O)>>DP(S), DP(O)>>DS(B),TOP>>DS(A)

#### **4.5.4 Responses to Subjects and Objects in LD and DP**

In the two controlled tasks—grammaticality judgment and translation task—I separated both LD and DP into two types, the structure with the coreferential subject or with the coreferential object. However, up until now I have not discussed whether the subjects responded differently with respect to the syntactic positions. Now we turn to see if there were asymmetric responses to the subject/object counterparts of the two

structures respectively. The percentage of the subject and object in the two structures

is presented in table 18, and the result of T-test of subject/object mean score

differences in table 19.

Table 18: the TP and non-TP mean scores of the subject and object in LD and DP

GJ task	LD		Non-LD		DP		Non-DP	
	S	O	S	O	S	O	S	O
English	.14	.21	<b>.86</b>	<b>.57</b>	.07	.29	.93	.71
Chinese	.83	.81	.10	.64	.55	.68	<b>.32</b>	<b>.10</b>
Level 1	.82	.82	.12	.83	.58	.63	.40	.27
Level 2	.91	.83	.40	.87	<b>.52</b>	<b>.83</b>	.22	.35
Level 3	.73	.82	.00	.00	.45	.64	.00	.09
TR task	LD		Non-LD		DP		Non-DP	
	S	O	S	O	S	O	S	O
English	.07	.21	.93	.79	.00	.00	1.00	1.00
Chinese	.86	.86	.06	.06	.44	.46	.34	.23
Level 1	.85	.85	.10	.08	.40	.40	.83	.50
Level 2	.96	.91	.00	.04	.57	.52	.09	.00
Level 3	.73	.82	.00	.00	.45	.73	.00	.09

Table 19: T-test of subject/object TP and non-TP mean scores in LD and DP

GJ task	LD(S)-LD(O)	n-LD(S)- n-LD(O)	DP(S)-DP(O)	n-DP(S)- n-DP(O)
English	t(13)=-.563 p=.583	t(13)=2.280 p=*. <b>.040</b>	t(13)=-1.385 p=.189	t(13)=1.385 p=.189
Chinese	t(93)=.228 p=.820	t(93)=1.000 p=.320	t(93)=-1.997 p=.051	t(93)=2.413 p=*. <b>.018</b>
Level 1	t(59)=.000 p=1.000	t(59)=.704 p=.484	t(59)=.597 p=.553	t(59)=1.734 p=.088
Level 2	t(22)=1.000 p=.328	t(22)=1.000 p=.328	t(22)=-2.612 p=*. <b>.016</b>	t(22)=1.817 p=.083
Level 3	t(10)=-.559 p=.588	NA	t(10)=-1.000 p=.341	t(10)=1.000 p=.341
TR task	LD(S)-LD(O)	n-LD(S)- n-LD(O)	DP(S)-DP(O)	n-DP(S)- n-DP(O)
English	t(13)=-1.000 p=.336	t(13)=1.000 p=.336	NA	NA
Chinese	t(93)=.000 p=1.000	t(93)=.000 p=1.000	t(93)=-.341 p=.734	t(93)=1.917 p=.058
Level 1	t(59)=.000 p=1.000	t(59)=.444 p=.659	t(59)=.000 p=1.000	t(59)=1.835 p=.072
Level 2	t(22)=1.000 p=.328	t(22)=-1.000 p=.328	t(22)=.430 p=.665	t(22)=1.447 p=.162
Level 3	t(10)=-.559 p=.588	NA	t(10)=-1.936 p=.051	t(10)=-1.000 p=.341

As shown in table 18, three cases of subject-object asymmetry were found with

significant differences in the GJ task—the Chinese group's overall response to non-DP (S>O), level 2's response to DP (O> S), and the English group's response to non-LD (S>O). However, in the TR task, there were no significant differences in the responses to subject and to the object of the same structure. Again the task effect was responsible for the subject-object asymmetry.

The findings of this study are summarized in 1-4:

1. TP scores: LD incurred the highest TP scores in controlled tasks, but the same scores decreased in the free production task. On the other hand, TP scores in DS(B) and TOP were more in comprehension task but less in production tasks. However, those of DP remained high in controlled tasks but, like LD, these scores decreased in the free production task. DS(A) induced the least TP scores in the comprehension task (GJ). Across levels, except for TOP in the comprehension task, the subjects' responses to TP structures were not significantly different.
2. Non-TP scores: LD received the least non-TP scores in all tasks. DS(B) and TOP received less non-TP scores in the comprehension task (GJ) than the production tasks (TR and GW). However, non-TP scores in DP remained constant across tasks. On the other hand, there was a U-shape pattern in non-TP scores of DS(A) in the comprehension task (GJ). In all the three tasks, students in level 1 received significantly higher non-TP scores than those in level 1 or level 2 or both.

3. Hierarchy of difficulty: LD was the most difficult structure since it was the most frequent TP structure in controlled tasks and the least frequent non-TP structure in all tasks. In contrast, DS(A) was the easiest structure since it was the least frequent TP structure in the comprehension task. DP was the second most difficult structure since its TP scores, though lower than LD, remained high in controlled tasks. In contrast, TP scores of DS(B) and TOP were more in comprehension task but became less in production tasks, and their non-TP scores were less in comprehension task but became more in production tasks. Therefore, DS(B) and TOP were easier than DP. Thus, the hierarchy of difficulty is--LD>>DP>>DS(B), TOP>>DS(A)
4. Responses to subjects and objects in LD and DP: there were three cases of subject-object asymmetry found in this study-- the Chinese group' s overall response to non-DP (S>O), level 2' s response to DP (O> S), and the English group' s response to non-LD (S>O). However, this asymmetry was only found in the comprehension task.

#### **4.6 Summary of Chapter 4**

In this chapter I have described the subjects, methodology, materials, procedures and the results of the study. It was found that, except for TOP in the comprehension task, TP scores were not significantly different across levels, but non-TP scores were,

with students in level 1 scoring the highest in all tasks. It was also found that a hierarchy of difficulty among the TP structures responded by the Chinese group-- LD(S), LD(O) >>DP(O), DP(S)>>DS(B), TOP>>DS(A). In addition, three cases of subject-object asymmetry existed in the comprehension task—Chinese students' responses' to non-DP(S>>O), and to DP(O>>S), English subjects' responses to LD (O>>S).