Measuring Coherence in Chinese EFL Majors' Writing
through LSA (Latent Semantic Analysis)

Huili Wang and Danni Sui
Dalian University of Technology, China
hilarydut@126.com   suidanni114@sina.com

Abstract

In recent years, writing, as one of the most important communicative skills, has been drawing increasing attention in English language teaching in China. This paper presents an investigation into the use of decomposition in Latent Semantic Analysis (LSA) to analyze textual coherence. Based on LSA and Halliday and Hasan's (1976) concepts of cohesion and their categorization of cohesive devices, this study reports on the analysis of the writing of 70 Chinese students. The results are consistent and indicate that LSA is an objective and appropriate method in measuring textual coherence. Therefore, it is argued that close attention should be paid to the cultivation of textual coherence in the teaching of English writing.

Key words:

LSA (Latent Semantic Analysis), coherence, cohesion, English writing

1. Introduction

Writing is a process of forming a text as a communicative bridge between the reader and the writer. The aims of language teaching are usually termed through the four skills: listening, speaking, reading and writing (Seidlhofer & Widdowson, 1999). Learning to write is an indispensable part of language learning and can reinforce language learning. Without learning to do so, we cannot have effective acquisition of a language. With awareness of this necessity, writing is receiving more and more attention in English language teaching in the Chinese context.

It is not easy to write good compositions in Chinese, let alone in English. Writing is frequently the most difficult skill for both native and non-native speakers, Chinese EFL learners being no exception. In recent years research into student writing shows that one of the major problems is the lack of coherence in the flow of ideas through a composition (Guo & Wang, 2005; Mao, 2002). In order to address this problem, teachers need to place importance upon formulating measures to improve the textual coherence of their students' compositions.

Coherence and cohesion are two main features of a text. Analysis in these areas can be applied to the instruction of English writing because they play an important role in writing. Coherence, including both sentence and paragraph coherence, lies in the deep structure of a text, and refers to the complex non-linear notional-functional relations behind the linear sequences of words, sentences or paragraphs.

One way to raise awareness of these issues comes in the form of Latent Semantic Analysis (LSA), a fully automatic statistical technique for extracting and inferring relations of expected contextual usage of words in passages of discourse. It is an analysis which represents the words used, and also any set of these words, such as a sentence, paragraph, or essay. The use of LSA in investigating words, sentences and passages has led to insights into a variety of human cognitive phenomena, ranging from developmental acquisition of recognition vocabulary to word-categorization, sentence-word semantic priming, judgments of essay quality and discourse comprehension (Landauer, Foltz & Laham, 1998). The application of LSA in discourse comprehension plays an essential role in this research paper. According to the original research for the application of LSA, it is an effective method for text-based research, especially for measuring textual coherence in reading comprehension, because the comprehension of text depends heavily on its coherence (Foltz, 1996). The method is proved to be automatic and fast, permitting quick measurements of the semantic similarity between pieces of textual information (Foltz, Kintsch & Landauer, 1998).

The objective of this study is to measure the textual coherence by means of LSA in Chinese English majors' English writings. The paper begins with the problems in

Chinese students' English writings. It then turns to the theory of coherence, cohesion and LSA. After this, an investigation is carried out into the coherence of 50 Chinese college students' writing (junior English majors) measured by LSA according to local structure, global structure and the relevance to the subject. Finally, some implications are drawn concerning English language teaching, especially about the teaching of English writing.

2. Literature review: Problems in advanced Chinese students' English writing

Writing is a process of conveying the writer's thoughts during which he searches for ideas and language materials from his long-term memory and organizes them into a composition according to various purposes and rules. In this process, the writer's main concerns are to convey clearly what he encodes in writing, and the reader's main task is to decode it correctly. To achieve this, both parties have to abide by "Game rules", which mean certain rules the writer has used to organize the structure into a coherent whole regarding a particular topic (Mao, 2002). "Game rules" give both writer and reader a shared linguistic knowledge for communication. Coherence refers to the relationships that link the meanings of the sentences in a text and may be based on the speakers' shared knowledge (Altenberg, 1987). Generally, a paragraph has coherence if a series of sentences develop a main idea. To achieve the objective of conveying the writer's meaning, the writer has to avoid distracting the reader from his message by making the message understood (Adelstein & Pival, 1980), that is, to make sure there is continuity between one part of the text and another.

Generally speaking, writing coherence can be divided into two main concepts: conceptive coherence and semantic coherence. Conceptive coherence is the consistency of structure and standpoint meaning that the article should focus clearly on one theme. The whole structure and sentential arrangement should be arranged logically and correspond to expressive practices. Semantic coherence means how the article can be semantically coherent. Therefore, besides the appropriate application of grammar and syntax, the author should convey his real intention to the reader in order

to make the reader understand; meanwhile, unrelated information should be eliminated, resulting in text which conveys the semantic coherence close to the original ideas intended (De Beaugrande & Dressler, 1981).

According to Halliday and Hasan (1976), cohesion is regarded as one of the most influential techniques in text analysis in terms of its current appeal in applied linguistics. De Beaugrande and Dressler (1981) point out that cohesion is created on the basis of presupposed coherence. The presence of cohesive devices, conjunctions, ellipsis, substitution, reference and lexical cohesion, all help a text to be an efficient means of expression and understanding, and also to be more stable or economic. We can see that the explicit cohesive devices in texts can not only help a writer to express ideas clearly, but also to help a reader understand a text more easily. If we strike out these cohesive ties in any text, the meaning of the text will not be changed, but the organization of the text will be less efficient, leading, in turn, to difficulties for the reader to achieve understanding.

For those advanced students of English, English majors, or those who have been trained in skills exclusively for preparation for TOEFL or IELTS tests, writing still remains difficult. These students, however, unlike other EFL learners, tend to make fewer mistakes in spelling and grammatical expressions because they develop a heightened awareness of discourse coherence in order to complete the tests successfully.

Specifically looking at Chinese students, it is often seen that they exhibit a lack of skills in use of vocabulary, grammar, rhetoric and punctuation, particularly making errors in the use of words and phrases, directly translating words from Chinese to English – interlanguage error – and organizing their written discourse poorly (Guo & Wang, 2005). Additionally, it is found that the viewpoints of Chinese students frequently shift. Due to different modes of thinking, a topic is not directly developed but rather explained with examples, suppositions, with conjunctions such as "if" and "whether" etc., many auxiliary verbs like "can", "must", and excessive use of "we" or "our". The passage has the appearance of a text but has no semantic connections between one part and another. That is the failure to create clear coherence among

4

themes. To address this, the use of appropriate cohesive devices can facilitate coherence in writing for all levels of learners.

In brief, it is necessary for a text to possess both coherence and cohesion, both of them representing the two basic considerations in developing writing.

3. Research Methodology

This study was conducted in December 2004 at Dalian University of Technology in China among 70 English majors (advanced English learners) in the School of Foreign Languages. They were juniors randomly sampled from 4 classes from the 2002 intake. They had studied English as their major for two years and had all passed the National English Test for English Majors (Grade Four), which is a requirement for all the English majors. The reason for choosing English majors as participants is that they make fewer grammatical mistakes in their writing and their main problems in writing are that of coherence and cohesion.

All the participants were doing classroom-writing assignments and were required to write an English composition of 300 words on a given topic from an IELTS written examination. A task of this nature was intended as a means to ascertain the students' real proficiency and was as follows:

Nowadays with the development of economy, existing cities are growing bigger and new cities are appearing. What do you think is ONE of the major problems that may result from this process of urbanization? Write an essay of about 300 words on the topic given below.

One Major Problem In The Process Of Urbanization

In the first part of your writing you should present your theses, and in the second part you should support the theses statement with appropriate details. In the last part you should bring what you have written to a natural conclusion or a summary.

There were actually two experiments in this paper. Firstly, the micro coherence

was measured by LSA; secondly, the macro coherence of the students' compositions was measured. The purpose was to prove that LSA is a feasible and objective method for teachers to accurately evaluate students' writing ability.

4. Data analysis and results of micro coherence by LSA

4.1 General analysis and results of micro coherence calculated by LSA

LSA can capture the semantic relatedness between words, sentences or passages because of the occurrences of keywords under the same topic, so there exists a similar semantic background for these compositions (Landauer & Dumais, 1997). In order to make the experimental data reliable, two groups of keywords were analyzed: word (0) and word (1). The former was gained from a small corpus of English majors' compositions and retrieved through Wordsmith; the latter was retrieved in the same way but taken from the newspaper, China Daily. So word (0) may mirror students' actual use while the other is perhaps more apt for native speakers.

The comparison of measurements of coherence between the students' actual scores and data carried out by LSA from word (0) and word (1) was made and the results show that the application of LSA in measuring textual coherence was probably an objective method in measuring sentence coherence. Moreover, the result was similar to the one achieved by our teachers (this includes some exceptions in that two compositions have the same score, though LSA's measurements show different data; the data belong to the same level, so they are effective.) Here are three samples with different levels shown in table 1.: lower level, the intermediate level and higher level.

	Higher level	intermediate level	lower level
$r_{(0)}$	0.57	0.46	0.24
$r_{(1)}$	0.44	0.26	0.12
actual score	18	13	6

Table 1. The comparison between r(**1**), r(**2**) and actual score. r: relatedness

LSA firstly calculated the coherence between sentences and then the mean of the

sentence coherence shows the micro coherence. From table 1, we can see that a clear distance between $r_{(1)}$ and $r_{(2)}$ existed in that that $r_{(1)}$ was closer to the ideal data, because 0.57 was close to 1.00 meaning that the coherence of this composition was higher. But here the $r_{(2)}$ is taken as the analytic standard because it was inclined to the use of native speakers. That is to say that, though $r_{(2)}$ is lower than $r_{(1)}$, its validity was much higher. This phenomenon can be explained in such a way that the words our students used were quite different from native speakers. Different thinking modes and cultural backgrounds are the basic factors exerting great influence upon the choices of words that our students use (Liang & Jia, 2004).

Chinese students place emphasis on rational analysis, logical thinking and always pay attention to the conclusion. Native speakers' writings tend to be quite different. Their English writing is often more direct. The topic is explicitly stated, things are clearly explained, and details or facts are required to support the topic. English discourse emphasizes logical reasoning and clear organization of ideas, with the main idea clearly stated, facts provided to support the main idea and references cited (Su, 2000). However, that is not to say the two kinds of choices are quite different. To some extent they are the more or less the same.

Here are some examples in table 2:

	A	E	H	P
Word (0)	all, also, and are, areas	economic		people
Word (1)	all, also, and are, areas	earth, economy environment	human	problem

Table2. The words input from word (0) and word (1) groups

Teachers read all the compositions and found there were fewer grammatical mistakes. Seen from the semantic perspective, the information they convey is the description of the results brought by urbanization but fewer cited examples to prove their points of view. They often wrote that the reason for urbanization was economic

changes but the disadvantages caused by urbanization were environmental pollution. Considering "urbanization', Chinese may associate this issue with the larger issue of population explosion, while native speakers may write something about the country planning. In a word, different cultures, mindsets and situations of a country can influence the way in which people think about a topic. Therefore, we cannot make such a conclusion that the word choices made by native speakers are better than the choices made by Chinese, because what they are thinking and have experienced are quite different. Though the word choices may not be the exact standard to estimate the compositions, the data from word (1) may be seen as more normal because it is closer to pure English from both language and thinking aspects.

4.2 Detailed description of LSA process of calculating a sample

Below is the detailed analysis made by LSA to calculate the semantic relatedness or coherence of the composition. Here a low level sample was selected randomly based on the China Daily corpus, in order to give a detailed description for analysis in figure 1, figure 2 and figure 3.

Firstly, a word-document matrix was established in terms of original matrix. A word by context matrix was formed from the titles of five articles about human-computer interaction and four about graph theory. Cell entries are the number of times that a word (rows) appeared in a title (columns) for words that appeared.

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}
areas	0	0	0	1	0		0	1	0	0	0	0…
china	0	0	0	0	0	0	0	0	0	0	0	1…
development	1	1	0	0	0	0	0	0	0	0	0	0…
economic	0	1	0	2	0	1	0	0	0	0	0	0…
industrial	0	0	1	0	1	0	0	0	0	0		0…
investment	1	0	0	0	0	0	0	0	0	0	1	0…
rural	0	0	1	0	0	0	0	0	0	0	0	0…
social	0	0	0	0	1	0	0	0	0	0	0	0…
work	0	0	0	0	0	0	0	1	1	0	0	1…

....		....	....	....	....	....	....	....	....	....	....	....	...

Figure 1. A word-document matrix.

However, a matrix containing too much information is much larger, so the mathematical technique of singular value decomposition (SVD) was needed to reduce the semantic dimension. Finally, we reconstructed the matrix in a manner similar to the original one but much smaller. It kept the essence of the semantic relationships in the texts but discarded incidental and irrelevant details.

Singular Value Decomposition of Term-Sentence
Right Vectors-Sentences

-0.64	0.40,	-0.15	0.15	0.00,	-0.30	-0.14	-0.07,	-0.36	-0.38...
-0.56,	0.41	0.27	0.27	-0.00	0.54	0.23	-0.03	-0.14	-0.08…
0.19	0.46	-0.09	-0.09	-0.00,	-0.19	-0.06	0.37,	-0.74	0.08…
0.44	0.05	0.08	0.08	0.00	0.16	0.04	-0.25	-0.13	-0.82…
0.21,	-0.56	0.18	0.18	-0.00	0.36	0.01	-0.25	-0.52	0.34…
0.00,	-0.00	0.55	0.59	0.14	-0.57	-0.00	-0.00,	0.00	-0.00…
0.00	0.00	-0.21	0.35	-0.91	0.07	0.00	-0.00	0.00	-0.00….
0.00	-0.00	0.70	-0.60	-0.39	-0.05	-0.00	0.00	0.00	-0.00…
-0.03	0.32	0.14	-0.14	-0.00,	-0.28	0.21	-0.83	-0.11	0.21…
0.04	-0.20	-0.07	-0.07,	-0.00	-0.14	0.94	0.22,	-0.03	-0.04…

Left Vectors-Sentence

-0.03,	0.09,	-0.06	0.05,	0.18	0.59	0.35	-0.60	-0.22,	-0.27…
0.68	0.66	-0.11	0.07	0.01	-0.00	0.00	-0.00	0.13	0.25…
-0.50	-0.53	-0.59,	-0.14	-0.27	0.00	0.00	0.00	-0.10,	-0.13…
-0.08,	-0.03	0.05	-0.55	0.33	0.00	-0.00,	-0.00	0.33,	-0.17…
-0.22	-0.06	0.38	0.33	-0.34,	0.00	0.00	-0.00,	0.46,	-0.61…
-0.06	0.19	-0.11	0.10	0.35	-0.57	-0.07	-0.05,	-0.44	-0.54…
-0.22	-0.23	0.16	-0.25	0.53	-0.00	-0.00	-0.00	0.29	-0.02…
-0.32	-0.10	0.64	-0.38,	-0.25	-0.00	-0.00	0.00	-0.52	0.06…

-0.27	-0.40	0.22,	0.59	0.40	0.00	0.00	-0.00	-0.09	0.30
0.00		0.00	0.00,	0.00	0.00,	0.00		0.00,	0.00…
	0.00,						0.00		

Figure 2. Complete SVD of Matrix in Figure 1.

1.00	0.48	0.23	0.23	0.00	0.25	0.30	0.35	0.64	0.65…
0.00	1.00	0.68	0.67	-0.01,	0.74,	0.84,	0.56	0.37	0.54…
0.00	0.00	1.00	1.00	0.45	0.74	0.88	-0.03,	0.36,	0.36…
0.00	0.00	0.00	1.00	0.48	0.72	0.86	-0.02	0.35,	0.36…
0.00	0.00	0.00	0.00	1.00	-0.26	0.00	-0.01	0.01	0.00…
0.00	0.00	0.00,	0.00	0.00	1.00,	0.95	-0.03,	0.39	0.39…
0.00	0.00	0.00,	0.00	0.00	0.00,	1.00	0.08	0.47	0.46…
0.00	0.00	0.00,	0.00,	0.00	0.00	0.00,	1.00	-0.01	0.57…
0.00	0.00	0.00	0.00	0.00,	0.00,	0.00,	0.00,	1.00	0.60…
0.00	0.00	0.00,	0.00	0.00	0.00	0.00	0.00	0.00,	1.00…

Figure 3. Two dimensional reconstructions of original matrix shown in Fig.1 based on shaded columns and rows from SVD as shown in Fig. 2.

The comparison between the shaded and boxed rows and cells of Figs.1 and 3 illustrates how LSA induces similarity relations by changing estimated entries up or down to accommodate mutual constraints in the data. In this way, we could see the semantic similarity between words, sentences or paragraphs through the cosine of their vectors. Through computational calculation, the coherence of adjoining sentences could be calculated as follows.

-0.02

0.41

-0.01

0

0

1

0.5

0.34

0

0

…

……
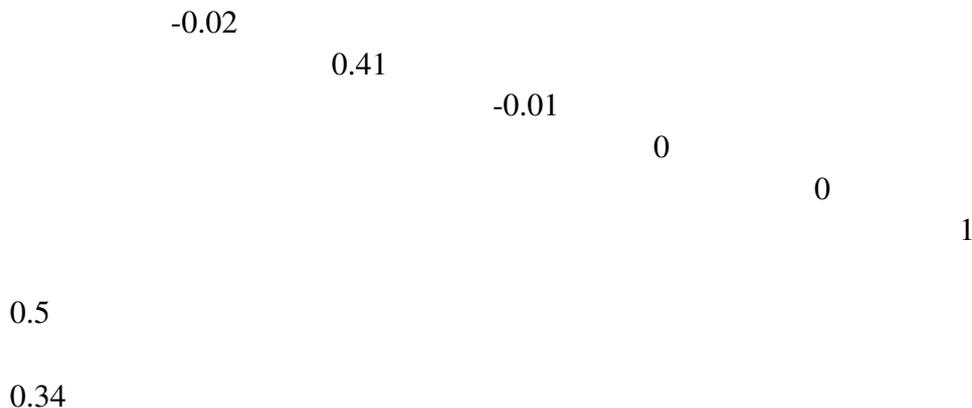
Figure 4. The final results of the coherence of adjoining sentences

An overall coherence measure was then calculated for the text by averaging the cosines between the vectors for all pairs of adjoining sentences. So the overall coherence of this sample was 0.12, which was not qualified data according to the standard data measured by LSA in that it should get close to 1. So this composition resulted in a much lower coherence.

4.3 A cohesive coding analysis VS deep structure analysis based on a sample

A text or a paragraph is supposed to be coherent, which means that the sentences should be so closely connected to one another that they flow smoothly without gaps between them or jumps in logic. When the sentences of a paragraph flow smoothly, growing out of another, the paragraph is said to be coherent. Sentences in a paragraph are coherent when there is a clear transition from one to another. There are three main methods of transition to achieve coherence. They are transitional words and phrases, repetition of key words, and pronoun reference, which are identical respectively with the cohesion methods of conjunction, lexical cohesion, and reference (Xie, 2002) .

Lexical cohesion is concerned with repeated occurrences of the related item (synonym, reiteration). A connective is one familiar type of marked connective relationship in text. According to Halliday and Hasan (1976), there are four types of connective relations, but sometimes it is difficult to group them. Consequently, we analyzed the students' writings according Halliday's categories of the connectives: elaboration, extension, and enhancement.

Samples were randomly chosen for further survey based on the students' corpus. The following was the student's original article. A cohesion analysis by coding scheme was conducted to present the factors that caused the failure or success in coherence.

S1   *As the economy of our society develops faster and faster, cities are growing bigger and now cities are appearing.*

S2   *However, a serious problem came along with this phenomenon, that is the change in population distribution, which is to say there are more and more people crowding into big cities; and less people are willing to stay in countryside.*

S3   *Recently, I've heard a piece of news about the change of population in big cities.*

S4   *To be exact, it is not "change", it should be "proliferation".*

S5   *There was a big increasing of population in large cities during last year, which made the population line much steeper.*

S6   *Nevertheless the population of countryside is becoming smaller and smaller.*

S7   *What's the reason for this phenomenon?*

S8   *The answer is very simple, that is the development of economy.*

S9   *Although our nation has been stressing "the development of western area" recent years and has taken some measures, many people in big cities are reluctant to go to the west part of China, (in which there are mainly rural areas) while people in rural areas are growing into big cities in large numbers.*

S10   *The underlying reason is that, people think the big cities such as Beijing, Shanghai, are modernized, and they can easily make money there, but if they stay in rural areas, they will be a farmer and will not lead a rich life.*

S11   *However, the fact is that not many people who come from rural areas and go into urban areas can live a life as rich as the expected, with their migration, the population in countryside becomes smaller and smaller, which results in that there are not enough people in countryside who grow crops.*

S12   *To sum up, the process of urbanization results in the change of population distribution, which becomes a serious problem in our society.*

The following table uses the coding scheme in Halliday & Hasan (1976), where

R=Reference

S=Substitution

E=Ellipsis

C=Connective

L =Lexical Cohesion

S=Sentence Number

T=Number of Cohesion Ties

O=Immediate (presupposing an item in continuous sentence)

M=Mediated (having one or more intervening sentences that enter into a chain of presupposition)

N=Remote, Non-mediated (having one or more intervening sentences not involved in the presupposition)

S	T	Cohesive item	Type	Presupposed item	Coherent relatedness
1					0.83
2	4	However	C	S1	
		problem	L	city	
		this	R	growing and appearing	
		phenomenon	L	growing and appearing	
3	3	change	Ll	Change	0.89
		population	L	population distribution	
		cities	L	Cities	
4	3	to be exact	C	S3	0.79
		it	R	Change	
		change	L	Change	
5					0.21
6	1	Nevertheless	C	S5	0.09
7	2	this	R	Becoming	0.42
		phenomenon	L	Becoming	
8	2	answer	L	Reason	0.71
		that	R	Phenomenon	
9	1	Although	C	S8	0.33
10	4	people	L	People	0.79
		cities	L	Cities	

		Beijing Shanghai Guangzhou	L	Cities	
		they	R	People	
11	4	however	C	S10	0.03
		people	L	People	
		urban areas	L	Big cities	
		countryside	L	Rural areas	
12	1	To sum up	C	All the above sentences	0.16

Table 3 shows that each of the categories (i.e. reference, connective, substitution, and lexical cohesion) occurred, except for ellipsis.

There were twenty-six cohesive items in the passage: Fifteen among them were lexical. They fell into two types: collocation and the same item.

Lexical cohesions were used most frequently, that is to say, they were high-frequency devices in this composition. The number of connectives was five, "however" occurred twice, "to be exact" once, "nevertheless" once, "although" once and "to sum up" once.

Most Chinese students usually familiarize themselves with English lexical items by referring to bilingual lists of words in the textbooks. As a result, they tend to transfer meanings of their native language to the English language. However, due to their limited vocabulary, there is a propensity to literally "translate" their Chinese characteristic "verb + object" structure into English, as seen in the following examples: "use their jobs", "receive a better education" and "enjoy the wonderful green".

Although many cohesive items were evident in this composition, the reader was left with an impression of discontinuity of coherence. At the beginning of the composition, the writer focused on the reason why people in rural areas moved into big cities, yet after that, the topic quickly changed. Whether a composition is well written or not does not totally depend on the number of the cohesive items, but the degree to which the sentences are closely connected. So an analysis of coherence is done based on the semantic relatedness but not the absolute literal cohesion.

The coherent relation between S1 and S2 was higher, because it was the beginning and also the first paragraph of an article, just indicating the existing problem but not any ideas or content involved in the deeper context. Although S2 and S3 did not belong to the same paragraph, they described the same content, so the coherent relatedness was also higher. In terms of S3 and S4, a connective was used as "to be exact" which is an extensive cohesion, but there also existed a confusion of pronominal reference. The semantic relatedness was 0.21, and from the table we can see the content indicated by S4 is transferred to another topic, which is S5. There was only a connective "nevertheless" between S5 and S6, resulting in the low coherent relatedness of only 0.09. To all appearances, the word "population" appeared both in both the sentences but it indicated different means of "population". Between S6 and S7, the coherence was comparatively high from literal comprehension, and the figure was not considerably higher, at 0.79, not over 0.5. The figure 0.71 expressed the semantic relatedness between S7 and S8, which was very reasonable, because the word "answer" was just the answer to S7 and the R "that" was equal to the question raised in S7. Though there is a connective "although" used as a connecting link between S8 and S9, the topic was also changed, and so in this sense, the semantic relatedness cannot be higher, standing at only 0.34.

5. Summary

All the analyses of the data calculated by means of LSA and the presentation of cohesive devices were more or less based on a subjective point of view. In other words, the data could not be an exact description of the coherence. Simply viewed, it can be regarded as an instrument or standard supporting the estimation of the students' composition. For instance, the semantic relatedness between S10 and S11 was only 0.03, but there were many cohesive devices (such as lexical cohesion) to be found. It cannot be deduced that there is no coherent relatedness between them because 0.03 is not a qualified description. It can, however, be deduced that it lacked semantic relatedness from a deeper structure of the whole article but not from the surface structure. Therefore, the cohesive devices represented a surface description for the

coherence of the article. It should be noted that this is not the pivotal standard to estimate the coherence of an article, because the deeper semantic relatedness through the context or the article is more important. The problem exists as to whether the mind follows a procedure of processing one topic and then between the preceding and the subsequent topics.

From the table, we can also ascertain the relatedness between S_{11} and S_{12} was 0.16. The reason why the indication was rather lower is because the connective "to sum up" was not used to express the coherence between S11 and S12, but between S12 and all the above sentences, in other words, the coherence between paragraph 3 and paragraph 1,2.

In brief, the quality of a composition does not totally depend on the quantity of cohesive devices used in it but the semantic relatedness between sentences based on the same topic or context. The data calculated by LSA can only provide an approximate expression or the approximate degree used for reference.

5.1 Data analysis and results of macro coherence by LSA

Besides sentences, English writing consists of paragraphs which are meaningfully connected. The paragraph is the basic constituent to create sections and chapters that are composed of by sentences. The sentences are used as semantically related expressions and explanations. The paragraph can be used as a syntactic signal which possesses thematic unity (Li, 2001).

Here, LSA was used to calculate the coherence of paragraphs in an article by the same method of calculation. The table below shows the results of the students' compositions (The samples here are also the same ones as the above based on the student composition corpus).

	Higher level	Middling level	Lower level
Rs or Rmicro	0.57	0.46	0.24
Rp or Rmacro	0.65	0.41	0.30
Actual score	18	13	6

Table 4 shows the comparison between coherence of paragraphs, sentences and actual score in three samples.

From the above table we can see the text with low local coherence 0.57or micro coherence but high macro coherence 0.65; while the high micro coherence 0.46 but low macro coherence 0.41or the micro coherence and macro coherence are in direct proportion. A detailed analysis follows.

5.1.1 The results of coherence techniques applied in English paragraph writing

Ni (2002) noted that the methods which can achieve coherence in paragraph writing are grammatical method, lexical cohesion, logical relationship and so on. There are two main parts: transition and sequence.

(1) Transition

Usually, one sentence, called the topic sentence, states the main idea of the paragraph. All the other sentences are related to this topic sentence. They further explain or support the main idea. Therefore, the transition of top sentences in each paragraph is very important because an overall meaning conveyed by the article is coming from the meanings of each paragraph while the topic sentences can present the general meaning of other sentences, in other words if the topic sentences are linked in series, the general idea of the writing can be gotten. In a word, meaning transition is laid in the most important position.

(2) Sequence

Sequence could be called logical sequence specifically. Generally speaking, the materials of an article should be arranged according to the logical sequence in order to achieve overall coherence. Such logical sequences are time and space, increase or decrease by degrees, deduction and induction.

After analyzing the students' compositions, they make better use of the sequence in their articles, usually the induction, which presents particular facts firstly and then come to a general conclusion keeping to the sequence that from specific to general.

	transition	Sequence
	12%	69%

statistics		Time and space	Increase or decrease by degree	deduction	induction
		1%			92%

Table 5: Per centum of coherence techniques applied in students' paragraph writing.

The blank indicates that the techniques applied by the students can't be identified clearly.

5.1.2 Cohesive techniques applied in English paragraph writing

Xu (2005), Zhang and Yuan (2003) believe that paragraph cohesion in an article can be strengthened in the following ways:

(1) The pronouns can be used to refer to someone, something or some idea mentioned in the last paragraph.

(2) The repetition of the key words or phrases occurs in the last paragraph.

(3) Making proper use of the transitional words throughout the article.

(4) Using a thesaurus to link with the last paragraph

(5) Using parallel structures to link with the last paragraph

(6) Using question sentences to link with the last paragraph

Among the above items, the third was widely used as the main means to create paragraph cohesion in the student compositions.

	Type1	Type2	Type3
Statistics	6%	30%	50% To Show Results: 94%

Table 6 Per centum of cohesive techniques
applied in students' paragraph writing.

The rest mean that the techniques applied by the students cannot be identified clearly.

6. Discussion

6.1 The analysis of statistical data in measuring textual coherence by using LSA

LSA is proved to be an objective and appropriate method in measuring both macro and micro textual coherence. A comparison between the degree of textual coherence calculated by LSA and the actual scores given by the experts is made to prove its validity; in addition, when a comparison between the degree of coherence calculated by LSA and cohesion analysis in a coding scheme is made, the results are identical in principle, except for some inevitable errors.

The data calculated by LSA can reflect the degree of the semantic relatedness between sentences or paragraphs and also indicates where the coherence is weak or strong.

Through the analysis, it is found that, on the one hand, the students under investigation had difficulty in forming clarity of thought, and that they were not good at arranging the meaning of the sentences logically at the macro-level; on the other hand, in terms of cohesive devices, good use was made of connectives, but in terms of lexical cohesion, their vocabulary size was relatively small.

As we know, a text is comprised of a structure that organizes sentences together into one meaningful context. Coherence and cohesion are two parts of the system of a language. One is expressed through semantic aspects while the other is expressed partly through grammar and partly through the vocabulary. Various devices are used to create the coherence in the text. In order to achieve coherent writing, neither of them can be ignored and in order to measure this aspect of the writing more thoroughly, the integration of LSA and a cohesive coding scheme might represent the best solution.

6.2 Disadvantages in LSA in measuring textual coherence

LSA is a statistical tool to obtain data which reflects the coherence between sentences or paragraphs which are already apparent in the paper. If the data is larger than 0.5, the data show a qualified coherent expression between sentences or paragraphs. In this way, the user can ascertain where the coherent relatedness is weak

or strong, and take effective measures to improve weak links. In this sense, the use of LSA can inform us about textual coherence, specifically where qualified coherence exists and where improvements are needed. It is especially useful for teachers to use it to measure the macro coherence and micro coherence of students' compositions, because they can note the weak links of all the students' compositions and then summarize generic errors; by analyzing these errors, teachers can see commonalities of error or place them into different categories, leading to more effective measures to be implemented to guide students. For example, if the students have difficulty in using cohesive devices properly, the teacher can give students the lesson about how to use cohesive devices appropriately in writing under different circumstances and contexts. From this point of view, we can put LSA forward as a possible means of error correction in English language teaching.

Fundamentally, LSA is a computational statistical method, so its scientific property and objective characteristics need to be considered. In terms of using LSA to measure textual coherence, however, some shortcomings should be noted. For example, the reason for the degree of coherence remains unclear, and also each value calculated by LSA cannot express any given meaning. The quantitative nature of the LSA method does not address these issues sufficiently, so further analysis is needed. Furthermore, LSA is not the only method that can be used to estimate textual coherence. Teachers cannot totally rely on the data from LSA analysis as it is not absolutely accurate. Under special circumstances, there may exist errors in judging the degree of textual coherence; for example, in the case where several cohesive devices can be used between two sentences, the ensuing calculated value may not be high. Accordingly, teachers should make a comprehensive analysis through the context. Despite these possible limitations, LSA can be nevertheless viewed as an objective statistical method.

6.3 Implications for second language teaching by using LSA to measure the textual coherence

The results of this research have some practical implications for the teaching of

English. LSA serves as a learning tool to conduct research in different fields, especially in psychology and education. For all educators, it is a credible and effective method for text-based research both in reading and writing, its use representing a means to provide insights into student compositions from the local structure (sentence coherence) to the global structure (paragraph coherence). It is helpful for teachers to identify problems in the writing process employed the students and take effective measures to address weaknesses. By measuring textual coherence, LSA can also help teachers to correct the students' examination papers (in writing) so that teacher work loads can be lightened.

Firstly, English learners should learn to employ clear ways of thinking and proper cohesive devices on which coherence is constructed. Clarity of thought can create a proper 'semantic space' to make the writing smooth and keep the relevance to the subject. At the same time, cohesive devices play an important role in English writing, thus suggesting it would be effective to introduce textual cohesion to students at an early stage in composition writing classes.

Secondly, the teachers should offer students model texts to analyze, paying attention to the writer's choice of words, such as connectives, personal pronouns, demonstrative pronouns and adjectives, etc. In this way, students can start to identify within a passage the various logical and linguistic links between sentences. At the same time, students need to examine the writer's organization of ideas, in other words, the writer's arrangements of text. Which sentence best expresses the main idea of the text? What supporting details are given to develop the main idea? How is the text organized?

Thirdly, after they have acquired an adequate vocabulary and fully comprehended the model texts, they are perhaps ready to embark upon their writing process with a heightened sense of awareness of cohesion and coherence. Once compositions have been completed, LSA can then be employed as a statistical method to estimate the writing coherence. Under guidance from the teacher, students can be shown where proper corrective measures are needed at the sentence or paragraph level. The more this process is practiced, the more students (and teachers) can familiarize

themselves with the techniques available to create coherence.

Lastly, different analytical thinking modes between English and Chinese can also have a great effect on Chinese students' English writing. To address this, it is important to raise students' awareness of analytical thinking by focusing on cultural influences. For example, in traditional English teaching, teachers firstly provide detailed explanations about the teaching content, then assign writing tasks to the class. If such a teaching methodology is employed, students do not have an opportunity to participate in discussion of the topic as part of the pre-writing stage in the writing process, i.e. development of the students analytical thinking is lacking. To counter this, seminars, workshops and training sessions should be a regular component of a school's ongoing development program for fostering students' analytical thinking skills. Students often tend to prefer small seminars where they can talk with their peers and are allowed to form their own opinions. As a consequence, they are more likely to be better prepared for their writing assignment.

7. Conclusions

The cultivation of communicative skills among English learners is an important task in foreign language teaching. Writing, as a complex communicative skill, is receiving more and more attention in research. Results of such investigations show that LSA, as a computational statistical method, can actually measure the textual coherence objectively and appropriately. Additionally, through the analysis of student compositions, the use of appropriate cohesive devices is seen to contribute significantly to their overall writing coherence. As a conclusion, it is argued that analysis of textual cohesion should be employed more in actual teaching practice as an illustration of how important it is for students to be able to connect sentences smoothly and logically and, consequently, create better coherence. Only when students have the ability to understand the coherence in texts and learn to use proper cohesive devices can their own writing improve.

References

Adelstein, M. E. & Pival, J. G. (1980). *The writing commitment*. New York: Harcourt Brace.

Altenberg, B. (1987). Causal ordering strategies in English conversation. In: J. Monaghan (Eds.), *Grammar in the construction of texts* (pp. 50–64). London: Frances Pinter.

de Beaugrande, R., & Dressler, W. (1981). *Introduction to text linguistics*. London: Longman.

Foltz, P. W. (1996). Latent semantic analysis for text-based research*, behavior research methods. *Instruments and Computers.* 28 (2), 197-202.

Foltz, P. W., Kintsch, W. & Landauer, T. K.. (1998). The measurement of textual coherence with latent semantic analysis. *Discourse Processes.* 25, 285-308.

Guo, L. Q. & Wang, H. L. (2005). Analysis of error types in Chinese English learners' writing. *Sino-US English Teaching*. 1(5), 9-13.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English.* London: Longman.

Landauer, T. K., Foltz P. W. & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes.* 2*, 259-284.

Landauer, T. K., & Dumais, S..T. (1997). A solution to Plato's problem: the latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review.* 104*, 211-240.

Li, Ch. Zh. (2001). On cohesion and coherence in written texts. *Journal of Xuzhou Normal University. (Philosophy and Social Sciences Edition).* 27(4), 138-141.

Liang, H. & Jia, S. (2004). On the cultural difference which affects university students' English writing. *Journal of Dalian University.* 25(3)*, 55-57.

Mao, C. H.. (2002). *An analysis of the problems of coherence in Chinese students' English writings.* Unpublished Master thesis, Shanghai Normal University, Shanghai.

Ni, Z. M. (2002). Coherence techniques applied in English paragraph writing. *Journal of Luo yang University.*17(1)*, 99-100.

Seidlhofer, B., & Widdowson, H. (1999). Coherence in summary: The contexts of

appropriate discourse. In Bublitz, W., Lenk, U. & Ventola E. (Eds.), *Coherence in spoken and written discourse. How to create it and how to describe it. Selected papers from the International Workshop on Coherence, Augsburg,* 24-27 April 1997 (pp. 205-219). Philadelphia: John Benjamins.

Su, X. Y. (2000). Coherence in college students' English writing. *Journal of Zhengzhou University of Technology (Social Science).* 18 (2), 60-62.

Xie, P. (2002). Analysis of students' compositions in cohesion and coherence. *Journal of Hangzhou Educational Institute.* 13, 24-28.

Xu, X. H. (2005). On the relation between cohesion and coherence in discourse. *China Science and Technology Information.* 3, 166-167.

Zhang, M. P.& Yuan, L. P. (2003). On the subject of unity and coherence in college English writing. *Journal of Qizhou Normal College.* 19 (2), 77-79.