World Journal of Education and Humanities
ISSN 2687-6760 (Print) ISSN 2687-6779 (Online)
Vol. 6 No. 4, 2024
www.scholink.org/ojs/index.php/wjeh

Original Paper

A Study on Text Cohesion in Senior High Students' Continuation Writing Based on Coh-Metrix

Ting Huang[1]

[1] School of Foreign Languages, Sichuan Normal University, Chengdu, China 610000

Received: June 17, 2024 Accepted: July 28, 2024 Online Published: August 09, 2024

doi:10.22158/wjeh.v6n4p85 URL: http://dx.doi.org/10.22158/wjeh.v6n3p85

Abstract

Continuation writing is a new type of writing task introduced in the Chinese college entrance examination reform. Text cohesion is essential for a well-written continuation. Cohesion in a text can be explicit, involving language-level textual cohesion, or implicit, involving semantic-level meaning continuity. This study uses the text analysis software Coh-Metrix to analyze the cohesion in high school students' English continuation writing. It explores the features of coherence in these writings from both the explicit language usage and the implicit semantic continuity, and the differences between the high-scoring group and the low-scoring group, aiming to provide effective teaching suggestions and references for English teachers.

Keywords

Text Cohesion, Continuation Writing, Coh-Metrix

1. Introduction

With the continuous advancement of the new English curriculum and the deepening of the college entrance examination reform, more and more provinces and cities have included Continuation Writing in the college entrance examination composition. Therefore, Continuation Writing has gradually become a hot topic of discussion among educators.

Continuation Writing is a comprehensive writing task that combines reading comprehension and writing (Wang, 2012). Specifically, students first read a text of about 350 words with the latter part removed, and then based on the given opening sentences of the two paragraphs of the latter part, they reasonably infer the development of the plot and write a continuation of about 150 words. The scoring criteria for Continuation Writing mainly evaluate the quality of the continuation from three aspects: content arrangement, logical structure, and language expression. In terms of the content arrangement, the continuation is required to create rich and reasonable content, effectively use cohesive devices

between sentences, have a clear overall structure and maintain coherence in meaning. Therefore, a good continuation must have appropriate cohesive devices, rich and reasonable plot development, and coherence in meaning.

At the same time, the *General High School English Curriculum Standards (2017 edition revised in 2020)* (hereinafter referred to as "*New Standards*") also made specific requirements for cohesion and coherence in the part of textual knowledge under language knowledge. In addition, the *New Standards* also propose that students should use cohesive devices to effectively improve the coherence of written texts in terms of language skills, learning strategies, and academic quality levels.

However, most students' Continuation Writing still has significant problems in overall textual cohesion. For example, the end of the first paragraph of the continuation cannot connect with the start of the second paragraph; the continuation cannot accurately echo the foreshadowing in the preceding text. The analysis of textual cohesion in high school students' Continuation Writing is helpful for teachers to understand students' weaknesses in discourse cohesion, so as to improve teaching plans and improve the quality of students' Continuation Writing.

With the development of computer technology, a new text analysis software called Coh-Metrix has started to be used for cohesion analysis (McNamara, Graesser, McCarthy, & Cai, 2014). Coh-Metrix is a text analysis tool developed by McNamara and others at the University of Memphis, which can automatically measure various surface and deep linguistic features, overcoming the past limitations of only being able to analyze surface or textual level cohesion (Du & Cai, 2013).

Graesser believes that five levels are related to text comprehension. These five levels include the surface, text, situational model, genre and rhetorical level, and the pragmatic level (Graesser, Millis, & Zwaan, 1997). The text and situational model levels are directly related to cohesion (Graesser & McNamara, 2011). The text level reflects the explicit information of the text, while the situational model level reflects the deep semantic meaning of the text (Van & Kintsch, 1983). In Coh-Metrix, the cohesion indices for the explicit language-level textual cohesion include connectives and referential cohesion, and for the implicit semantic-level textual cohesion, they include Latent Semantic Analysis (LSA) and situation model construction. The current study selected 22 indices directly related to textual cohesion from the four following categories: referential cohesion, connectives, latent semantic analysis (LSA), and situation model (see Table 1).

Table 1. The Indices Selected from Coh-Metrix

Index	Sub-index	Description
Referential Cohesion	CRFNO1	Noun overlap, adjacent sentences, binary, mean
	CRFAO1	Argument overlap, adjacent sentences, binary, mean

	CRFSO1	Stem overlap, adjacent sentences, binary, mean
	CRFNOa	Noun overlap, all sentences, binary, mean
	CRFAOa	Argument overlap, all sentences, binary, mean
	CRFSOa	Stem overlap, all sentences, binary, mean
	CRFCWO1	Content word overlap, adjacent sentences, proportional, mean
	CRFCWOa	Content word overlap, all sentences, proportional, mean
Latent Semantic Analysis (LSA)	LSASS1	LSA overlap, adjacent sentences, mean
	LSASSp	LSA overlap, all sentences in paragraph, mean
	LSAPP1	LSA overlap, adjacent paragraphs, mean
	LSAGN	LSA given/new, sentences, mean
Connectives	CNCCaus	Causal connectives incidence
	CNCLogic	Logical connectives incidence
	CNCADC	Adversative and contrastive connectives incidence
	CNCTemp	Temporal connectives incidence
	CNCAdd	Additive connectives incidence
Situation Model	SMCAUSv	Causal verb incidence
	SMINTEp	Intentional verbs incidence
	SMCAUSr	Ratio of casual particles to causal verbs
	SMINTEr	Ratio of intentional particles to intentional verbs
	SMTEMP	Temporal cohesion, tense and aspect repetition, mean

2. Research Procedure

The present study primarily addresses the following two research questions:

1) What are the features of text cohesion in senior high students' continuation writing?

2) Are there any differences between high-scoring and low-scoring group continuation in text cohesion?

To address these two research questions, this study first employs a text analysis method by randomly selecting continuation writing pieces from 30 students at a certain high school. These works are converted into electronic documents and uploaded to the Coh-Metrix analysis software. Relevant indices from the analysis results are selected, and statistical data such as the mean and standard deviation of each indicator are calculated to analyze the cohesion and characteristics of the students' continuation writing. Subsequently, all samples are professionally scored by two experienced high school teachers and the score range is 0-25. All samples are divided into two groups based on their scores: the high-scoring group (20-25, referred to as HG) and the low-scoring group (0-19, referred to as HG). The cohesion data from the two groups were subjected to an independent samples t-test using SPSS software to ascertain whether the differences between the two sets of data were statistically significant. The cohesion differences of both the low- and high-scoring groups were then examined.

3. Results and Discussion

3.1 The Features of Textual Cohesion in Senior High Students' Continuation Writing

Table 2. The Results of Cohesion Indices

REFERENTIAL COHESION INDICES	MINIMUM VALUE	MAXIMUM VALUE	MEAN VALUE	STANDARD DEVIATION
CRFNO1	0.171	0.429	0.270	0.067
CRFAO1	0.455	0.727	0.597	0.070
CRFSO1	0.171	0.429	0.276	0.067
CRFNOa	0.123	0.290	0.186	0.042
CRFAOa	0.415	0.655	0.537	0.052
CRFSOa	0.178	0.338	0.242	0.020
CRFCWO1	0.100	0.180	0.125	0.018
CRFCWOa	0.086	0.134	0.106	0.010
LSA				
LSASS1	0.135	0.240	0.174	0.027
LSASSp	0.089	0.207	0.140	0.028
LSAPP1	0.259	0.345	0.299	0.020
LSAGN	0.312	0.397	0.341	0.017
CONNECTIVES				
CNCCaus	21.50	39.18	28.42	4.62
CNCLogic	33.33	53.54	40.00	4.27
CNCADC	11.86	23.81	16.30	3.40

CNCTemp	15.05	27.31	20.97	3.27
CNCAdd	29.54	49.60	40.55	5.64
SITUATIONAL MODEL				
SMCAUSv	23.26	48.52	33.73	5.18
SMINTEp	31.00	48.83	36.82	4.06
SMCAUSr	0.100	0.538	0.253	0.107
SMINTEr	0.524	0.900	0.708	0.115
SMTEMP	0.776	0.915	0.843	0.036

From the Table, we can observe that in referential cohesion, aside from the average value of 0.597 for argument overlap between adjacent sentences (CRFAO1) and 0.537 for argument overlap among all sentences (CRFAOa), all other referential cohesion indicators are below 0.5 (with the maximum value being 1), indicating a medium to low level. Argument overlap occurs when there is overlap between a noun in one sentence and the same noun (in singular or plural form) in another sentence; it also occurs when there are matching personal pronouns between two sentences (e.g., “he”/“he”) (McNamara, Graesser, McCarthy, & Cai, 2014). This suggests that students tend to use the same nouns or pronouns to refer to the same entities in their continuation writing. While this practice can enhance the readability of the text and strengthen semantic connections between sentences, it also indicates that students need to improve their ability to use a more diverse and flexible vocabulary.

Latent Semantic Analysis (LSA) provides measures of semantic overlap between sentences or between paragraphs. Each measure varies from 0 (low cohesion) to 1 (high cohesion) (McNamara, Graesser, McCarthy, & Cai, 2014). In the LSA (Latent Semantic Analysis) section, we can observe that the semantic overlap between paragraphs in the continuation writing (LSAPP1) is stronger than the semantic overlap between sentences (LSASS1, LSASSp). This can be partly attributed to the fact that the continuation writing is based on two given topic sentences, which naturally increases the semantic overlap between paragraphs. Additionally, this indicates that students are beginning to consciously focus on cohesion and coherence between paragraphs, although their execution still needs improvement. Moreover, the LSA Given-New (LSAGN) metric shows the best performance within the LSA measures. Text constituents can be classified into three partitions: given, partially given (based on various types of inferential availability), or not given (i.e., new). This is a proxy for how much given versus new information exists in each sentence in a text, compared with the content of prior text information, for example, G/(N+G) ((McNamara, Graesser, McCarthy, & Cai, 2014)). In other words, a higher LSAGN value indicates that the continuation writing contains more information previously provided in the text. This suggests that students are consciously referencing the given information during the continuation writing process, but overall, their performance in this area still needs improvement.

In the connectives section, it is evident that the average values for logical connectives (CNCLogic) and additive connectives (CNCAdd) are the most prominent. In contrast, the lowest average value is for adversative connectives. Coh-Metrix analysis confirmed the higher incidence of causal and logical connectives in the high-cohesion text (McNamara, Graesser, McCarthy, & Cai, 2014). However, the average value for causal connectives in the continuation writing is quite low. This suggests that students are not using different types of connectives in a flexible or balanced manner.

The term "situation model" has been used by researchers in discourse processing and cognitive science to refer to the level of mental representation for a text that involves much more than explicit words [11]. In the situational model section, the average value for tense and aspect repetition (SMTEMP) is the highest. This measure tracks the consistency of tense and aspect throughout the text, indicating that students maintain tense consistency fairly well in their continuation writing. However, the ratio of causal particles to causal verbs (SMCAUSr) is relatively low (0.253), suggesting that students use causal particles and verbs less frequently. These ratios are calculated to indicate the importance of connectives in a text, with their necessity depending on the number of events described in the text. A text is considered more causally cohesive when it contains a higher proportion of connectives linking actions and events. If a text has many action, event, and intentional verbs but lacks causal connectives to guide the reader, the reader may have to make more inferences to understand the relationships between these actions and events in the sentences. The data reveals a lack of appropriate causal connectives in the continuation writing to link intentions and actions. This results in gaps in coherence and continuity between sentences, which corresponds with the relatively low average values for sentence-to-sentence semantic overlap (LSASS1, LSASSp) in the LSA section. With episodes in narrative text, the situation model would include the plot (McNamara, Graesser, McCarthy, & Cai, 2014). This also indicates that students' ability to construct a coherent storyline during the continuation writing process needs improvement.

Overall, the level of textual cohesion in students' continuation writing is relatively low. In terms of explicit language-level cohesion, students demonstrate basic narrative skills and are able to maintain tense consistency in their writing. However, their ability to use a diverse range of vocabulary and expressions flexibly, such as varying vocabulary and sentence structures, still needs improvement. In terms of implicit semantic-level cohesion, students make an effort to ensure cohesion between paragraphs and consciously refer to previously provided information, but their performance in these areas is still lacking. Additionally, their ability to construct a coherent storyline requires further development.

3.2 The Difference between high-scoring and low-scoring Group Continuation in Text Cohesion

Table 3. Independent Sample test of Cohesion Indices in HG and LG

REFERENTIAL COHESION INDICES	HG MEAN	LG MEAN	P VALUE
CRFNO1	0.324	0.250	0.550
CRFAO1	0.655	0.576	0.249
CRFSO1	0.330	0.257	0.434
CRFNOa	0.200	0.181	0.296
CRFAOa	0.562	0.528	0.910
CRFSOa	0.254	0.237	0.297
CRFCWO1	0.135	0.121	0.995
CRFCWOa	0.109	0.106	0.657
LSA			
LSASS1	0.190	0.169	0.953
LSASSp	0.145	0.138	0.440
LSAPP1	0.303	0.297	0.582
LSAGN	0.342	0.341	0.180
CONNECTIVES			
CNCCaus	29.46	28.05	0.324
CNCLogic	36.63	40.12	0.112
CNCADC	15.16	16.72	0.424
CNCTemp	20.97	20.96	0.041*
CNCAdd	39.93	40.78	0.968
SITUATIONAL MODEL			
SMCAUSv	31.36	34.59	0.601
SMINTEp	35.90	37.15	0.803
SMCAUSr	0.315	0.230	0.419
SMINTEr	0.724	0.700	0.456
SMTEMP	0.858	0.837	0.755

* $p < 0.05$.

From Table 3, it is evident that there is a significant difference in the use of temporal connectives (CNCTemp) between the high-scoring and low-scoring groups ($p < 0.05$). Connectives play an

important role in the creation of cohesive links between ideas and clauses and provide clues about text organization (Cain & Nash, 2011) This indicates that the high-scoring group places greater emphasis on using temporal connectives to enhance the cohesion between sentences in their continuation writing. Temporal connectives and temporal adverbials, are themselves cognitively motivated (Grisot, 2018). This suggests that students in the high-scoring group have a better understanding and more familiarity with using temporal connectives compared to those in the low-scoring group. Future research could explore this further from the perspectives of cognitive psychology and cognitive linguistics.

There are no significant differences in most of the cohesion indicators between the high-scoring and low-scoring groups ($p > 0.05$). This can be attributed to two factors. First, the samples for this text analysis were all taken from continuation writings produced by students in the same class. Second, it is likely that the English teacher in this class used effective teaching methods for continuation writing, which helped maintain a consistent level of performance among the students. The teacher frequently guided students to focus on the cohesion between paragraphs in their writing, analyzed important plot cues and foreshadowing from the original text, and encouraged extracurricular reading to build strong language skills. Although this teaching model has only been in place for one semester and students' cohesion abilities are still at a moderate level, it is evident that students are consciously improving their cohesion by using the given information effectively in their writing. Future empirical research could explore the effectiveness of this teaching model further.

4. Conclusion

4.1 Major Findings of this Study

The present study reveals several important findings. Firstly, the overall level of textual cohesion in students' continuation writing is fairly low. While students demonstrate basic narrative skills and can maintain tense consistency in their writing, their ability to effectively use a diverse range of vocabulary and expressions, such as varying vocabulary and sentence structures, needs improvement. In terms of implicit semantic-level cohesion, students make an effort to ensure coherence and continuity between paragraphs and consciously refer back to the previously given information, but they struggle to do so effectively. Moreover, their capacity to construct a coherent storyline requires further development.

Secondly, a significant difference was observed between the high-scoring and low-scoring groups in their use of temporal connectives (CNCTemp). This suggests that the high-scoring group places greater emphasis on using temporal connectives to improve cohesion between sentences during the writing process. These students are more adept at recognizing and using temporal connectives than their lower-scoring counterparts.

4.2 Implications

Based on the findings of this study, English teachers should focus on the following aspects in their continuation of writing instruction. First, concerning explicit language-level textual cohesion, teachers should guide students to pay attention to the flexible use of various connectives in their writing and

help improve their language habits. To enhance students' ability to use the target language flexibly, teachers should move away from simplistic and mechanical translation methods and monotonous practice routines. Instead, they should adopt an approach grounded in specific textual contexts, innovating teaching methods in line with new curriculum standards. This approach should aim to enrich students' learning experiences, focusing on deepening their understanding of the language and improving their precision in language use, thereby broadening their overall language proficiency. Additionally, teachers can encourage students to engage in extracurricular reading in English, which will enrich their exposure to the target language and enhance their sensitivity to it.

Additionally, regarding implicit semantic-level textual cohesion, English teachers should not only help students understand a broader range of cohesion and coherence skills but also emphasize both global and local cohesion within texts. Teachers should encourage students to use diverse cohesive language to strengthen the connections between paragraphs and sentences in their continuation writing. To improve students' ability to construct coherent plots, teachers can enhance students' inspiration through extracurricular reading. Furthermore, integrating continuation writing into daily lessons and employing various writing modes, such as paragraph expansion and group collaborative continuation writing, can provide students with more opportunities to practice and develop their plot-building skills.

In summary, teachers should focus not only on language-level cohesion but also on semantic-level cohesion in their instruction. It is essential to help students improve their ability to use the target language flexibly and guide them to pay attention to both global and local semantic coherence and plot development within their texts.

References

Wang, C. M. (2012). Reading Followed by Writing—An Effective Method to Improve the Efficiency of Foreign Language Learning. *Foreign Language World*, (05), 2-7.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

Du, H.Y., & Cai, J. T. (2013). A Study of Coh-Metrix-based Predictive Modeling of Chinese English Learners' Writing Quality in Argumentative Essays. *Modern Foreign Languages*, *3*, 293-300+331.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, *48*(1), 163-189. https://doi.org/10.1146/annurev.psych.48.1.163

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, *3*, 371-398. https://doi.org/10.1111/j.1756-8765.2010.01081.x

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.

Graesser, A. C., & McNamara, D. S. (2011). Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science*, *3*, 371-398. https://doi.org/10.1111/j.1756-8765.2010.01081.x

Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology*, *103*(2), 429-441. https://doi.org/10.1037/a0022824

Grisot, C. (2018). Cohesion, Coherence and Temporal Reference from an Experimental Corpus Pragmatics Perspective, *Springer Cham*, *XXIV*, 319. https://doi.org/10.1007/978-3-319-96752-3