

# A Corpus-based Study of Syntactic Complexity in CATTI (China Accreditation Test for Translators and Interpreters) Prescribed Textbooks

Weiping Zhang<sup>1</sup>, Yan Wang<sup>2</sup>

<sup>1,2</sup>School of International Studies, Hunan Institute of Technology, Hengyang, Hunan, China

<sup>1</sup>155270597@qq.com, <sup>2</sup>1799464275@qq.com

**Abstract:** ***Objective:** Through comparing the differences in syntactic complexity between CATTI English interpretation practice (Level 3) and general English, this study enables textbook developers, publishers and researchers to judge the value of CATTI English interpretation practice (Level 3) more accurately, improve syntactic measures and provide inspiration for the evaluation of the syntactic complexity of texts in CATTI English interpretation practice (Level 3). **Methods:** The research method of corpus has provided a brand-new research perspective and means for translation studies, making the research more objective and rational. The L2 Syntactic Complexity Analyzer (L2SCA) allows researchers to analyze the syntactic complexity of English language texts. By using 14 different measurement methods, it helps to improve the quality of both quantitative and qualitative research. **Conclusions:** This paper has investigated the syntactic complexity of CATTI interpretation textbooks and found that there are significant differences in ten indicators between it and general English as well as academic English. The results of this study are instructive for textbook compilers: textbook compilers must use the corpus as a tool to conduct stricter quality control over the materials they compile. For textbook users, they should break away from their dependence on textbooks and supplement learning materials outside textbooks for language input so as to improve their English level. Colleges and universities can independently determine the starting stage and teaching objectives according to their actual needs.*

**Keywords:** Teaching materials; Syntactic complexity; CATTI English interpretation practice; Corpus.

## 1. INTRODUCTION

With the further development of economic globalization, English major graduates will have more opportunities to work in a multicultural environment. To adapt to the new trend of the development of higher education in China, improve teaching quality and meet the needs of the society and the country for high-quality talents in the new era [1]. The China Accreditation Test for Translators and Interpreters (CATTI), entrusted by the Ministry of Human Resources and Social Security of the People's Republic of China, is specifically implemented and managed by the China Foreign Languages Publishing Administration. It is a national professional qualification examination and has been included in the category of national vocational qualification certificates [2]. As a unified translation professional qualification (level) certification fully implemented in China, it evaluates and judges the bilingual translation ability and level of examinees in interpretation or translation. The goal of compiling the interpretation practice textbooks is to comprehensively cultivate students' interpretation abilities according to the curriculum requirements of CATTI. Therefore, in interpretation teaching, it is necessary to improve students' autonomous learning ability and comprehensive cultural awareness to meet the needs of China's social development and international exchanges [3]. However, in actual English learning, for various reasons, some students have not achieved positive and satisfactory learning results. Some students cannot get satisfactory scores in the CATTI English interpretation practice (Level 3) examination. Finding out the reasons for the poor performance of English major students in the CATTI interpretation ability test has become an urgent and arduous problem.

## 2. RESEARCH SIGNIFICANCE

For a long time, people's enthusiasm for foreign language teaching research has mainly been focused on the language system itself, psycholinguistics and teaching techniques. Students' learning achievements are mainly attributed to teachers' teaching methods. As the carrier of knowledge, teaching materials play a key role in students' English language learning process. Thus, the status of English teaching materials cannot be underestimated. Studies show that foreign language learning is affected by many factors such as gender, age, personality traits,

learning strategies, learning atmosphere, motivation, language talent, cognitive style, and learning attitude [4]. Teaching is a process of interaction between students and teachers. Teaching materials, as the transmitter of knowledge, are worthy of in-depth exploration by scholars.

Currently, the official textbook of CATTI English interpretation practice (Level 3) still plays an incomparably important role in the entire CATTI interpretation curriculum system. The quality of the official textbook of CATTI English interpretation practice (Level 3) is of great significance for improving the quality of foreign language teaching in colleges and universities and cultivating students' ability to use language reasonably. Scientific and reasonable teaching materials play a decisive role in promoting interpretation teaching [5]. Therefore, the textbook corpus, as the language input source for learners, has become a new research focus after the corpus of native English speakers as language output and the corpus of English learners. However, so far, no research has explored college English textbooks from the perspective of sentence structure complexity, and the level of syntactic complexity is an important indicator for measuring English proficiency. This paper uses the "syntactic complexity analyzer" to conduct an in-depth analysis of the official textbook of CATTI English interpretation practice (Level 3) and compares it with the general English corpus to provide references for the compilation and application of teaching materials.

### **3. SYNTACTIC COMPLEXITY**

#### **3.1 Definition of Syntactic Complexity**

Syntactic complexity is an important concept in language acquisition and an important indicator of the level of foreign language development. It is an important indicator for measuring the richness of language expression and language ability. It mainly involves the complexity of sentence structure, including aspects such as sentence length, the number of clauses, the use of subordinate structures, and the proportion of specific phrase structures [6]. For example, in IELTS writing, syntactic complexity can be presented by 14 indicators in the L2SCA analyzer to show the syntactic complexity characteristics of students at different levels. In the study of the English reading comprehension texts of the National English Proficiency Test Level 2 (PETS-2), with the help of corpus methods and specific text analysis tools, 5 different indicators were selected to analyze the syntactic complexity of the reading comprehension texts.

#### **3.2 Measurement Methods of Syntactic Complexity**

##### **3.2.1 Corpus-based Measurement**

**Output Length:** It includes indicators such as the mean length of a single sentence (MLS) and the mean length of a T-unit (MLT). Usually, the values of intermediate-level students in terms of output length are higher than those of beginner-level students to varying degrees, which confirms that output length is an intuitive indicator for distinguishing the writing syntactic complexity of students at different language levels.

**Sentence Complexity:** Many similarities are reflected in the dimension of "sentence complexity", which is related to the ratios of T-units, clauses, complex nominal phrases, etc., and can be used as an important indicator for distinguishing the IELTS writing language ability of beginner and intermediate-level students.

**Subordinate Complexity:** For example, the ratios such as CT/T and DC/C in "subordinate complexity", if less than 1, indicate that the use of such is relatively rare.

**Coordinate Complexity:** The two indicators related to coordinate phrases (CP/C, CP/T) in "coordinate complexity" show different results in different writing tasks and among students at different levels, indicating that the requirements and design elements of writing tasks will lead to different performances of students in coordinate complexity.

**Specific Phrase Structure:** It is related to the ratio of subordinate clauses and is also a sign for distinguishing students at different levels.

### **4. RESEARCH SUBJECTS, RESEARCH METHODS AND RESEARCH DESIGN**

#### **4.1 Research Subjects**

This study creates an English text corpus of CATTI English interpretation practice (Level 3), analyzes the characteristics of the syntactic complexity of the English texts. Meanwhile, it uses the Crown general English corpus as a reference corpus. By comparing the differences in syntactic complexity between CATTI English interpretation practice (Level 3) and general English, it enables textbook developers, publishers and researchers to judge the value of CATTI English interpretation practice (Level 3) more accurately, improve syntactic measures and provide inspiration for the evaluation of the syntactic complexity of texts in CATTI English interpretation practice (Level 3). If the syntactic complexity indicators of the CATTI English interpretation practice (Level 3) textbooks are closer to those of the general English corpus, it indicates that the textbooks prescribed by CATTI are more in line with the general English standards and the texts selected for the CATTI-prescribed textbooks are more successful.

## **4.2 Research Methods**

### 4.2.1 Literature and data method

The author Used China network, CNKI, wan fang database and searched platform or collected books in library.

### 4.2.2 Specific Analysis Tools

The L2SCA analyzer for second language syntactic complexity: In the research, this analyzer is used to analyze CATTI texts, and 14 different indicators are selected for analysis to systematically present the syntactic complexity characteristics of CATTI.

### 4.2.3Corpus-based method

The corpus research approach has presented a completely novel research perspective and methodology for translation studies, rendering the research more objective and logical. The L2 Syntactic Complexity Analyzer (L2SCA) empowers researchers to dissect the syntactic complexity of English texts. Through the utilization of 14 distinct measurement techniques, it contributes to enhancing the caliber of both quantitative and qualitative research.

## **4.3 Research Design**

### 4.3.1 Research Questions

The basic assumption of this paper is that the official textbook of CATTI English interpretation practice (Level 3), as an important language input channel and learning example for foreign language learners, should have a syntactic complexity close to that of actual language use. The research question is: Is there a significant difference in the syntactic complexity between the official textbook of CATTI English interpretation practice (Level 3) and general English?

### 4.3.3 Corpora Used

The official textbook of CATTI English interpretation practice (Level 3) used in this paper is published by the Foreign Languages Press. This textbook is compiled according to the "Syllabus for the English Level 3 Translation (Interpretation and Translation) Examination (Revised Edition)". The content of the textbook is rich and diverse, involving interpretation exercises on different themes and scenes, such as ceremonial speeches, international exchanges, tourism, cultural education, sports, health care, exhibition introductions, performing arts, business talks, etc. Learning these contents can help candidates become familiar with various common interpretation topics, enhance interpretation skills and the ability to cope with different scenes.

The general English corpus used as a reference is the Crown corpus developed by Beijing Foreign Studies University. The development of the Crown corpus has gone through a certain period of time, and the final version was released in November 2022. It contains 700 texts. This corpus is a balanced American English corpus with a scale of about 1 million words. It collects contemporary American English texts. As an updated language resource of contemporary American English, it provides new materials and bases for language research and provides important references for contrastive studies involving diachronic changes (compared with Brown, Frown and other corpora), regional differences (compared with LOB, FLOB, CLOB and other corpora) and cross-language

comparisons (compared with LCMC, Torch series corpora, Globe series corpora and other corpora).

## 5. RESULTS AND ANALYSIS

In the realm of language learning and assessment, understanding the syntactic complexity of educational materials is of paramount importance. This is particularly true when it comes to textbooks like the CATTI (China Accreditation Test for Translators and Interpreters) English textbook (Level 3), which plays a significant role in the language education of English major students. To delve deep into its syntactic characteristics, a series of meticulous procedures were carried out.

The journey of this research began with the careful input of the texts from the CATTI textbook into the computer. Each article within the textbook was treated as an independent entity, a self-contained unit for processing. This approach was adopted to ensure that the analysis could capture the unique syntactic features of each individual piece of writing. By isolating each article, we could obtain a more accurate and detailed understanding of the syntactic complexity within the textbook as a whole.

Once the texts were ready for analysis, the "syntactic complexity analyzer" was employed. This sophisticated tool is designed to dissect language structures and extract a wealth of information about syntactic complexity. It was able to generate 9 basic indicators and 14 derived syntactic complexity indicators for each file. These indicators are not just abstract numbers; they represent different aspects of syntactic structure, such as the length of clauses, the distribution of different types of phrases, and the relationships between various syntactic elements.

After obtaining all these indicators for each file, the next step was to perform a series of statistical operations. The average value and standard deviation of these indicators were calculated. The average value provides a central tendency, giving us an idea of the typical syntactic complexity level within the textbook. The standard deviation, on the other hand, reveals the degree of variability among the texts. Additionally, a variance test was conducted. This test helps to determine whether the differences in the syntactic complexity indicators are statistically significant or just random fluctuations.

With a comprehensive understanding of the CATTI textbook's syntactic complexity, the research then moved on to a broader comparison. The entire set of textbooks was considered as one entity, and this was then compared horizontally with two well-established corpora: the general English corpus and the general academic English corpus. The general English corpus represents the everyday language used by native speakers in various contexts, while the general academic English corpus reflects the language used in academic settings.

**Table 1**

	CATTI(n=15)		Brown(n=15)		ACDEMIC(n=15)		ANOVA	
	Mean	SD	Mean	SD	Mean	SD	F	Sig.
MLC	23.40	7.96	27.74	3.88	21.56	2.58	7.47	0.00
MLS	21.64	7.07	25.42	3.17	19.25	2.84	8.05	0.00
MLT	16.15	7.13	12.22	1.79	11.52	1.87	5.42	0.01
C/S	1.51	0.27	2.31	0.39	1.90	0.25	0.42	0.95
C/T	1.96	0.35	3.03	0.53	2.28	0.28	0.63	0.81
CT/T	1.39	0.22	2.12	0.38	1.70	0.26	5.46	0.00
DC/C	0.30	0.11	0.44	0.06	0.51	0.52	1.94	0.16
DC/T	0.44	0.21	0.95	0.26	3.17	3.24	9.09	0.00
CP/C	1.08	0.10	1.10	0.12	1.13	0.06	0.64	0.53
CP/T	0.36	0.14	0.59	0.11	0.49	0.14	12.39	0.00
T/S	0.62	0.29	0.57	0.25	0.48	0.19	1.00	0.37
CN/C	0.46	0.24	0.27	0.11	0.29	0.12	5.21	0.01
CN/T	2.61	0.72	3.06	0.46	2.53	0.64	6.83	0.00
VP/T	1.89	0.71	1.46	0.30	1.50	0.34	4.21	0.00

The results presented in Table 1 are quite revealing. They show that there are significant differences in most of the syntactic complexity indicators among the CATTI textbook, general English, and academic English. In terms of MLC (Mean Length of Clause), for example, the CATTI textbook might have a distinct average clause length compared to the other two corpora. This could imply that the sentences in the CATTI textbook are structured in a

way that is different from the norm in native - speaker English. Similarly, differences in MLS (Mean Length of Sentence) and MLT (Mean Length of T - unit) further highlight disparities in sentence and text - unit lengths.

The differences in ratios such as CT/T (Complex Transitive Clauses per T - unit), CP/T (Clausal Phrases per T - unit), CN/C (Clausal Nodes per Clause), CN/T (Clausal Nodes per T - unit), and VP/T (Verb Phrases per T - unit) also paint a clear picture of the divergence in syntactic complexity. These ratios reflect the internal organization and complexity of sentences, and the significant differences suggest that the CATTI textbook has a unique syntactic profile.

Lu's research on the syntactic complexity level of Chinese students provides an interesting backdrop to our findings. Although Lu focused on the syntactic development of English major students rather than direct comparison with native - speaker language, his results are highly relevant. He found that the differences in syntactic complexity between high - level and low - level students were mainly concentrated in 9 indicators: MLC, MLS, MLT, CT/T, DC/T, CP/T, CN/C, CN/T and VP/T.

Given that the CATTI textbook is the main source of language input for English major students, the significant differences in syntactic complexity between the textbook and native - speaker English raise concerns. If students are exposed primarily to a form of English with a syntactic structure that deviates from the norm, it is likely to have a negative impact on their language acquisition. For example, students may struggle to understand native - speaker conversations or write in a more natural and fluent manner. They may also face difficulties in adapting to real - world language use, whether in academic or professional settings.

These findings emphasize the need for a more in - depth examination of the CATTI textbook's content. It is essential to consider how to bridge the syntactic gap between the textbook and native - speaker English to enhance students' language learning outcomes. This could involve revising the textbook content, supplementing it with additional materials that reflect native - speaker syntactic patterns, or providing targeted instruction on syntactic complexity. Future research could also explore how to incorporate these findings into teaching methods to better support students in developing a more native - like syntactic proficiency.

## 6. CONCLUSIONS

This paper has investigated the syntactic complexity of the CATTI interpretation textbook and found that there are significant differences in ten indicators from general English and academic English. The results of this study have implications for textbook compilers: textbook compilers must use the corpus as a tool to conduct stricter quality control on the materials they compile. For textbook users, they should break away from their dependence on textbooks and supplement learning materials outside textbooks for language input to improve their English level. Colleges and universities can independently determine the starting stage and teaching objectives according to their actual needs.

## ACKNOWLEDGEMENT

This paper is financially supported by College Students' Innovation and Entrepreneurship Projects in Hunan Province in 2023, with project name A Corpus-based Study of Syntactic Complexity in CATTI Prescribed Textbooks (S202311528138).

## REFERENCES

- [1] Chang, Jinhui. *The Representation and Development of Second Language Syntax of Chinese English Learners* [D]. Lanzhou University, 2020.
- [2] Jiang, Lin. Semantic Priming in the Priming of Double - Object Construction and Prepositional Dative Construction [J]. *Modern Foreign Languages*, 2009(1): 59 - 67.
- [3] Lu, X. Automated measurement of syntactic complexity in corpus - based L2 writing research and implications for writing assessment [J]. *Language Testing*, 2017, 34(4): 493 - 511.
- [4] Lu, X., & H. Ai. Syntactic complexity in college - level English writing: Differences among writers with diverse L1 backgrounds [J]. *Journal of Second Language Writing*, 2015, 29: 16 - 27.
- [5] Qin, Xiaoqing & Wen, Qiufang. *Research on the Development Laws and Characteristics of Chinese College Students' English Writing Ability* [M]. Beijing Social Sciences Press, 2007: 6.

- [6] Wu, Yufang. The Influence of Syntactic Complexity on the Intra - Linguistic Syntactic Priming Effect in Second Language [D]. Fujian Normal University, 2019.

### **Author Profile**

**Weiping Zhang** lecturer, School of International Studies, Hunan Institute of Technology, Hengyang, Hunan, China Research direction: translation practice.

**Yan Wang** undergraduate student of School of International Studies, Hunan Institute of Technology. Research direction: translation.