Universe Scientific Publishing

DOI:10.18686/ahe.v7i23.9803

A Comparative Analysis of Syntactic Complexity of Middle School English Textbooks

Lezi Qiu

School of Foreign Studies, South China Normal University, Guangzhou, China 510000

Abstract: The syntactic complexity of textbooks plays a pivotal role in language learning, influencing not only students' comprehension but also their reading motivation and capability. While earlier textbook analyses often depended on subjective methods or relied on isolated measures, they seldom offered a holistic view of syntactic complexity. This study employs the L2 Syntactic Complexity Analyzer, Python 3.11, and SPSS 26.0 to assess and compare 14 syntactic complexity indicators across two prevalent high school English textbook series from the People's Education Press (PEP) and the Shanghai Foreign Language Education Press (SFLEP). Findings indicate that while PEP has a lower average syntactic complexity, it demonstrates a steadier increase compared to SFLEP. Moreover, PEP presents a more discernible hierarchical pattern in its syntactic complexity relative to SFLEP. This research furnishes significant insights into the quantitative evaluation of syntactic complexity in English textbooks, offering a foundational framework for textbook assessment and selection in China.

Keywords: Syntactic complexity; Middle school English textbooks; L2SCA

Introduction

The English Curriculum Standard of Compulsory Education (ECSCE) in China[1] mandates specific requirements for linguistic complexity in textbook compilation. Similarly, Western research emphasizes reading text complexity and its influence on comprehension. Studies aligned with the "ZPD"[2] have demonstrated the profound impact of teaching materials' complexity on language learners' development[3].

Syntactic complexity, a central determinant of text difficulty, affects students' reading capabilities[4-5]. Defined as the range and sophistication of syntactic structures in a text[6-7], it is an integral component in evaluating the readability of reading texts for both first and second language (L2) readers[8-9] and an index of language proficiency[10].

Considering its multi-dimensional nature[11], a plethora of measures has been proposed for syntactic complexity, and tools like the L2 Syntactic Complexity Analyzer (L2SCA) have been developed[12]. The L2SCA, applying 14 measures, is extensively used, particularly in analyzing written English compositions and teaching materials in China, particularly for junior high schools and colleges[13-14]. It covers five dimensions: Length of Production Unit, Amount of Subordination, Amount of Coordination, Degree of Phrasal Sophistication and Overall Sentence Complexity. The 14 syntactic complexity measures(SCMs) are summarized in Table 1.

Despite this, there appears to be a literature gap concerning the input provided to middle school learners via textbooks, especially when comparing PEP and SFLEP, with previous studies often relying on subjective methods[15-18].

To adress the gap, the study employs quantitative research using L2SCA to gauge syntactic complexity in two middle school English textbook series (PEP and SFLEP). It aims to identify syntactic complexity variations between the series and ascertain if the textbooks' complexity aligns with expected academic progression. As previous researches have not provided consistent results on which measures can be used to analyze the syntactic complexity of teaching materials, all 14 indexes of five dimensions will be included in this study.

Table 1 Syntactic Complexity Measures in L2 Syntactic Complexity Analyzer

Type	Code	Measure
Length of Production Unit	MLC	Mean length of clause
	MLS	Mean length of sentence
	MLT	Mean length of T-unit
Amount of Subordination	CT/T	Complex T-units per T-units
	DC/C	Number of dependent clauses per clause
	DC/T	Number of dependent clauses per T-unit
Amount of Coordination	CP/C	Number of coordinate phrases per clause
	CP/T	Number of coordinate phrases per T/unit
	T/S	Number of T-units per sentence
Degree of Phrasal Sophistication	CN/C	Number of complex nominals per clause
	CN/T	Number of complex nominals per T-unit
	VP/T	Number of verb phrases per T-unit
Overall Sentence Complexity	C/S	Number of clauses per sentence

1. Research Questions

This research aims to investigate the syntactic complexity of middle school English textbooks in two different series. The research questions are designed to explore the overall syntactic complexity of the two series, any differences in complexity between the five-volume textbook series, and how the complexity changes over time.

RQ 1: What is the syntactic complexity of the two middle school English textbook series? Does the syntactic complexity, as measured, align with anticipated growth in complexity as students progress to higher grades?

RQ 2: Are there any significant differences in syntactic complexity between the two compared five-volume middle school English textbook series? If so, which specific syntactic measures account for these differences?

2. Method

2.1 Data collection and description

The data for this study originates from ten textbooks spanning two distinct series: the People's Education Press (PEP) and the Shanghai Foreign Language Education Press (SFLEP) English editions. Each series contains five textbooks tailored for middle school students, described as follows:

Grade 7: Volume 1 (specified as "Book 1")

Grade 7: Volume 2 (specified as "Book 2")

Grade 8: Volume 1 (specified as "Book 3")

Grade 8: Volume 2 (specified as "Book 4")

Grade 9: Volumes 1 & 2 (specified as "Book 5")

The study's corpus for PEP comprises reading sections titled "Section A (3a)" and "Section B (2b)", as well as other continuous textual content such as letters. Any visual elements like drawings, tables, and supplementary notes have been excluded from the dataset.

For comparative analysis via the "Independent-Sample T-Test" in SPSS, the extracted texts from each book are combined and segmented into sections of 1000 words each.

Each book in the SFLEP series includes 8 units. Each unit usually contains one text from the "Reading" section, one from the "More Practice" section, and other continuous discourses (excluding illustrations and other graphic information). Simultaneously, the corpus for each volume will be chunked for data analysis.

2.2 Data processing

The study employs NeoSCA, a refined version of Xiaofei Lu's L2 Syntactic Complexity Analyzer (L2SCA), to assess the syntactic complexity of Middle School English textbooks. NeoSCA, compatible with Windows and with an enhanced command-line interface, analyzes written English samples, computing 14 syntactic complexity indices. Using these analyses, distinctions in syntactic complexity between PEP and SFLEP are identified. Subsequently, syntactic complexity measures (SCMs) are processed in SPSS 26.0, calculating descriptive statistics before conducting t-tests to contrast PEP and SFLEP. Lastly, using Python 3.11, the study

chunks the corpus texts every 1000 words, runs one-way ANOVAs, and executes post-hoc tests, with Bonferroni's t-tests identifying developmental indices.

3. Results

3.1 Comparison of Overall Syntactic Complexity: PEP vs. SFLEP

3.1.1 Analysis of Syntactic Complexity Measures Between PEP and SFLEP

Upon evaluating 14 SCMs using NeoSCA, it is observed that, in comparison to SFLEP, the PEP series has lower mean values for 8 of the 14 measures, namely MLS, MLT, MLC, T/S, CP/T, CP/C, CN/T, and CP/C. However, PEP does not have lower mean values for the measures C/S, VP/T, C/T, DC/C, DC/T, and CT/T (See Figure 1).

Subsequent independent-sample t-tests(Table 4) reveal significant differences between the PEP and SFLEP series in two specific measures. Specifically, CP/T exhibits significant differences at a 95% confidence level and CP/C at a 99% confidence level. Both these measures fall under the category of "Amount of Coordination."

3.1.2 Developmental Patterns of SCMs in PEP and SFLEP Textbooks

To address RQ2, the progression of the 14 measures across the five volumes of each series is analyzed. While PEP textbooks largely demonstrate a pattern of increasing complexity from Book 1 to Book 5, there's a noticeable stagnation between Books 2 and 3in most SCMs. This suggests a potential plateau in challenge levels between Grades 7 and 8. On the other hand, SFLEP series manifests marginal variations and even presents an irregular rise between grades.

For example, in PEP, Figure 2 reveals an upward trend in Overall Sentence Complexity as grade levels ascend. To illustrate, Book 1 exhibits the most straightforward sentence structure with a C/S value of 0.9255, whereas Book 5 peaks with a value of 1.5119. This escalation is especially pronounced in the initial grades, evidenced by the marked increase from Book 1 to Book 2 and from Book 3 to Book 4.

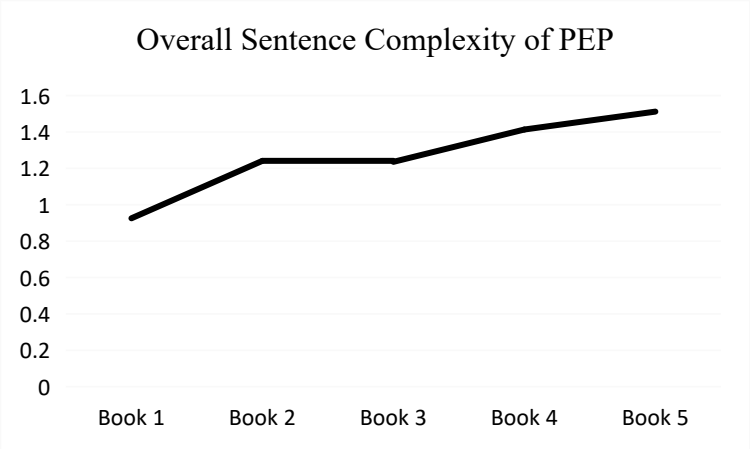

Figure 1 The Consistency of Overall Sentence Complexity of Five PEP Volumes

In comparison, there is a general trend of increasing Overall Sentence Complexity with higher grade levels in SFLEP(Figure 3), but an unexpected decrease from Book 3 to Book 2 stands out, which goes against the anticipated pattern of increasing complexity with higher grades.

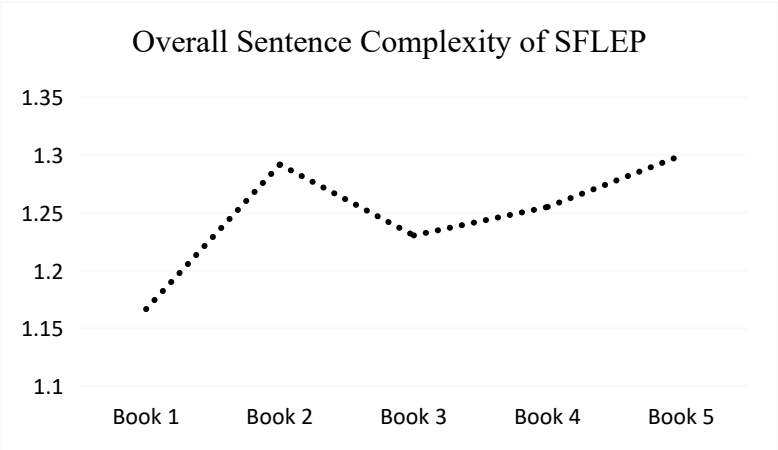

Figure 2 The Consistency of Overall Sentence Complexity of Five SFLEP Volumes

3.1.3 Differences in Syntactic Complexity of Five Volumes in Respective Series

In the second phase of the analysis, a one-way ANOVA is executed to discern potential discrepancies across the five volumes (from Book 1 to Book 5) within each series. This is followed by a post hoc Bonferroni test, aimed at identifying any notable distinctions between individual pairs among the five books.

The resuts of one-way ANOVA test for PEP(Table 2) implies that the textbooks present linguistic structures with diverse complexity levels. Remarkably, CP/T and CP/C emerge as consistent measures, with no discernible differences across the textbooks.

Post-hoc comparisons reveal discernible distinctions in measures such as MLS, MLT, and VP/T, even between adjacent textbooks, particularly between Books 1 and 2. For measures like MLS and MLT, Books 1 and 5 show pronounced differences, indicating a consistent progression of complexity across the series. Meanwhile, some non-adjacent books, like Books 1 and 4, are distinguished by the C/S, VP/T, and DC/C measures. In sum, while most measures can differentiate syntactic complexity both between adjacent and non-adjacent textbooks in the PEP series, CP/C and CP/C in Amount of Coordination remain still throughout.

Table 2 Summary of Significant ANOVA test of PEP

Measure	Levene Statistic (Homogeneity)	Sig. of Levene Test	ANOVA F-Value	Sig. of ANOVA	Post Hoc Differences Between Books（1−5）
MLS	1.683	0.187	9.427	0.000***	1-2, 1-4, 1-5, 3-5
MLT	0.587	0.675	10.361	0.000***	1-2, 1-3, 1-4, 1-5, 3-5
C/S	1.439	0.252	4.761	0.006***	1-2, 1-5
VP/T	2.305	0.088	7.682	0.000***	1-2, 1-3, 1-4, 1-5
C/T	1.900	0.143	4.258	0.010**	1-5
DC/C	0.706	0.596	6.969	0.001***	1-2,1-3, 1-4, 1-5
DC/T	0.855	0.505	4.257	0.010**	1-5
T/S	1.413	0.260	3.427	0.024**	1-2
CT/T	0.742	0.572	6.958	0.001***	1-3,1-4, 1-5
CN/T	4.704	0.006**	5.123	0.004***	1-5
CN/C	3.190	0.031**	3.644	0.019**	1-5

Note:***p<0.01; **p<0.05; *p<0.1

One-way ANOVA tests of SFLEP(Table 3) shows that the majority of measures do not significantly distinguish the textbooks at conventional levels. Notably, only the CN/T and CN/C measures of Degree of Phrasal Sophistication have p-values less than 0.05, indicating a significant difference. This significance is further localized between two nonadjacent volumes, Books 1 and 4, in post hoc multiple comparisons, with Book 1 having lower values than 4. In contrast, measures like MLS, MLT, and C/S (among others) do not distinctly differentiate any specific pairs of textbooks within the series. The consistent lack of significance across many of the indices might indicate a relative uniformity in terms of syntactic complexity for those specific metrics.

Table 3 Summary of Significant ANOVA test of SFLEP

Measure	Levene Statistic (Homogeneity)	Sig. of Levene Test	ANOVA F-Value	Sig. of ANOVA	Post Hoc Differences Between Books(1-5)
CN/T	2.400	0.088	4.302	0.013	1-4
CN/C	2.653	0.067	4.229	0.014	1-4

Note: **p<0.05

4. Discussion

Utilizing L2SCA, Python 3.11, and SPSS 26.0, this study scrutinizes the syntactic complexity of two middle school English textbook series. The intent is to discern patterns consistent with expected linguistic growth across grades and distinguish the intricacies of both series, laying a foundation for material selection and textbook comparison.

Research Question 1

The data reveals that on 8 out of 14 measures, PEP has lower mean values than SFLEP. Notably, only two measures related to Coordinated Phrases (CP/T and CP/C) set the series apart, with PEP trailing SFLEP. This signifies PEP's minimal use of Coordinated Phrases. The PEP series generally escalates in complexity from Book 1 to Book 5. However, a stagnation from Book 2 to Book 3 implies a possible plateau in linguistic challenges. In contrast, SFLEP fluctuates with grade levels, showing a marked increase in Overall Sentence Complexity (C/S) and Degree of Phrasal Sophistication (CN/T, CN/C) from Book 2 to Book 3. This divergence underscores the need for meticulous material selection for different learner stages.

Overall, the significant difference between PEP and SFLEP confirm the role syntactic complexity plays in the linguistic complexity of ESL textbooks. The different patterns of increase and stabilization among the syntactic complexity measures show that students should confront with texts with increasing unlinearly SCMs; rather, it is necessary to pay attention to different dimensions of syntactic complexity at different (clusters of) grade levels.

Research Question 2

The investigation's aim is to verify if both series reflected an anticipated rise in complexity with advanced grades. In PEP, despite a general upward trend from Book 1 to Book 5, the plateau between Book 2 and Book 3 stands out. This deviates from Liao's findings[18], likely due to Liao's reliance on "average sentence length" and a lesser emphasis on clause and phrase density.

SFLEP, on the other hand, offers less predictability across grades, resonating with Cheng's[15] observation of repetitive grammar points. The present study reveals that the spike in complexity measures from Book 2 to Book 3 suggests a shift to more intricate nominal phrases, possibly necessitating fewer compound/complex sentences to balance reading difficulty.

The stark difference in the hierarchical structuring of syntactic complexity between the series becomes evident. While PEP showcases clear distinctions based on complexity, SFLEP lacks this grading pattern, suggesting possible challenges for lower-grade students. Teachers must exercise caution in material selection, and SFLEP compilers should aim for clearer gradations.

In sum, this analysis sheds light on the intricacies of syntactic complexity in English textbooks. The notable differences between PEP and SFLEP, and their varied progression patterns, can guide educators and curriculum developers in textbook selection for middle schoolers. Expanding this study by encompassing a broader textbook range and diverse syntactic complexity tools would be a valuable next step.

Implications

(1)Progression PatterThis analysis of the syntactic complexity in two middle school English textbook series underscores their varied linguistic progression. Key implications include:

ns: While PEP series typically advances in syntactic complexity, it stalls between Books 2 and 3, highlighting the need for educator vigilance. SFLEP's inconsistent pattern necessitates careful grade-level appropriateness checks.

(2)Complexity Metrics: A comprehensive set of measures is crucial for complete syntactic complexity evaluation, as seen when juxtaposed with Liao's findings.

(3)Teaching Approaches: Identified complexity variations suggest periods where reinforcing concepts might be prioritized over increasing challenges, indicating a need for adaptable teaching strategies.

(4)Textbook Selection: Differences between PEP and SFLEP, particularly in Coordinated Phrases, emphasize syntactic complexity's role in textbook selection.

Overall, a nuanced approach to English textbook evaluation is vital, with the findings guiding curriculum planning, textbook design, and pedagogy.

5. Limitation and future work

This study, analyzing the syntactic complexity of middle school English textbooks in China, has some boundaries. It focuses on two major textbook series, potentially limiting its broader applicability. Future research should examine a wider range of textbooks for more inclusive results. While this study emphasizes syntactic complexity using the L2SCA tool, a broader spectrum of textual complexity, including vocabulary and cultural nuances, should be explored. Incorporating qualitative methods like discourse analysis alongside quantitative data can offer a deeper understanding of student-textbook interactions. Additionally, the creation of advanced natural language processing tools suited for various learner levels can aid educators globally in textbook selection, optimizing student learning outcomes.

References:

[1]Ministry of Education of the People's Republic of China, 2022, (Compulsory Education) English Curriculum Standards [S]. Beijing: People's Education Press.

[2]Vygotsky, L. S. 1978. Mind in Society: the Development of Higher Psychological Processes[C]. Cambridge, MA: Harvard University Press.

[3]Ai, Haiyang & Xiaofei Lu, 2013. A Corpus-based Comparison of Syntactic Complexity in NNS and NS University Students' Writing [J]. Automatic Treatment and Analysis of Learner Corpus Data

[4]Lei Lei, Yaqian Shi, 2020. Syntactic complexity in adapted extracurricular reading materials[J]. The Modern Language Journal. Volume104, Issue1.