BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

The Study on Sentence Difficulty of the Chinese Components

Jia-Fei Hong [1*], ***Jia-Ni Che*** [1], ***Yao-Ting Sung*** [2]

[1] *Department of Chinese as a Second Language, National Taiwan Normal University, Taipei 106, Taiwan*

[2] *Department of Educational Psychology and Counseling, National Taiwan Normal University, Taipei 106, Taiwan*

ABSTRACT

In recent years, reading comprehension has gradually become a proficiency indicator of interest in lexical and grammar. As sentences are the basic units of discourse structure, sentence difficulty is often applied to the study of text difficulty. Although there have been a number of studies on sentence difficulty, the lack of consistency in the indicators chosen or the discussion of specific grammatical issues have limited the research on sentence difficulty. Therefore, this study adopts a corpus-based approach, using a corpus as an objective and scientific data source. The study utilizes the Digital Platform for Chinese Grammar and the 8000 Chinese Words as important reference sources. Additionally, the CRIE 3.0 is employed to validate the texts and establish sentence difficulty indicators. However, due to the incomplete development of certain indicators in the "Chinese Grammar Digital Platform", the study refers to the Chinese Proficiency Grading Standards for International Chinese Language Education and Hanyu Shuiping Kaoshi to establish comprehensive sentence structure and sentence component difficulty indicators. Subsequently, the established difficulty indicators are validated by conducting comparative analyses using corpora as the basis. Native speaker corpora are used as benchmarks, while Mandarin learner corpora are used for comparison, and then validate objectively through the machine learning model. These validation aims to examine the validity and reliability of the selected indicators and establish a calculation method involving "level of grammar * point distribution ratio of grammar " to determine the difficulty indicators for Chinese sentences, Additionally, expert reliability is accessed to ensure the credibility of indicators.

*CORRESPONDING AUTHOR:
Jia-Fei Hong, Department of Chinese as a Second Language, National Taiwan Normal University, Taipei 106, Taiwan;
Email: jiafeihong@ntnu.edu.tw

ARTICLE INFO
Received: 11 June 2024 | Revised: 29 July 2024 | Accepted: 31 July 2024 | Published Online: 14 October 2024
DOI: https://doi.org/10.30564/fls.v6i4.6737

CITATION
Hong, J.-F., Che, J.-N., Sung, Y.-T., 2024. The Study on Sentence Difficulty of the Chinese Components. Forum for Linguistic Studies. 6(4): 434-448. DOI: https://doi.org/10.30564/fls.v6i4.6737

Keywords: Corpus-based; Sentence difficulty; Indicators; Lexical complexity; Syntactic complexity

1. Introduction

Reading comprehension is considered an important language proficiency indicator (Diana et al., 2011; Sung et al., 2013; Wu et al., 2020). Carrey & Connie (2015) found that there has a close relationship between syntactic proficiency and reading comprehension abilities. Not only learners' native language, proficiency in a second language also significantly predicts the correlation between comprehension of second language sentences and texts. When talking about reading texts, readability refers to the degree to which learners can understand a text. High readability articles are easier for readers to comprehend and typically feature lower difficulty vocabulary, simpler sentence structures, fewer compound words, and content that aligns with the reader's prior knowledge. However, previous studies on sentence difficulty have some limitations. Firstly, the selected indicators vary, leading to inconsistent research result. Some studies only consider vocabulary features as indicators, while others focus on specific grammar discussions (Ye & Qiu, 2008; Hong et al., 2016). Secondly, research has tended to be qualitative in the past, lacking consensus on how to quantify sentence difficulty and which features to select for quantification (Cheng, 2005; Wu et al., 2020). Thirdly, insufficient quantitative analysis has resulted in a lack of relevant applications and teaching research.

This study aims to build a corpus-based sentence difficulty indicators upon previous research by examining the complexity of vocabulary and grammar in Chinese texts. It seeks to confirm vocabulary and grammar indicators that affect sentence difficulty and establish a readability index for sentences. Native speaker corpus serve as standard, while corpus of Chinese as second language's learners are used to validate text readability. Machine learning model is employed to validate the effectiveness of the sentence difficulty indicators, and the reliability of the indicators is examined through Likert scales and tests. The ultimate goal is to assist teachers in conducting classroom teaching for learners of different proficiency levels and rewriting texts to meet the learning needs of Chinese as second language's learners at different levels.

2. Review of literature

2.1 Studies on sentence difficulty

Sentence difficulty is essentially the assessment of how well a sentence can be understood. Therefore, when discussing text readability, many studies also approach the topic from the perspective of sentence difficulty (Klare, 2000; Diana et al., 2011; Lu & Ai, 2015; Sun & Wan, 2016; Pang, 2016; Lu et al., 2019). Quantitative research on sentence difficulty has historically considered factors such as sentence length and the length of sentence components (Flesch, 1948; Gunning, 1952; Fry, 1963; McLaughlin, 1969; Brown, 1971; Klare, 2000; Lu, 2010; Wu et al., 2020). Additionally, many studies have taken into account the complexity of both vocabulary and sentence structure (Spache, 1953; Thoma & Gilbert, 1967; Kenneth & Susan, 1973; Laufer & Nation, 1995; Cheng, 2005; Ye & Qiu, 2008). However, there is still a lack of clear research on the impact of vocabulary and syntactic complexity on sentence difficulty. Even though it is acknowledged that both factors affect sentence difficulty, the absence of related quantitative studies prevents the effective confirmation of their respective impacts.

Next, the research will discuss the impact of lexical complexity and syntactic complexity, and then explore effective quantitative analysis methods on sentence difficulty.

Studies on lexical complexity

Among the various factors influencing sentence difficulty, vocabulary plays a significant role. Lexical complexity often involves the breadth and depth of lexical knowledge (Laufer & Nation, 1995). The main measurement indicators for lexical complexity include lexical sophistication, lexical diversity, lexical density, and lexical originality. And thus, vocabulary, being the smallest unit of meaning that learners encounter, fundamentally influences text readability. Diana et al. (2011) pointed out in their experiment with third-grade students reading science-themed texts that too many rare words in a text might prevent students from comprehending it.

In studies related to the complexity of Chinese vocabulary, Cheng (2005) identified three factors influencing Chinese sentence difficulty: sentence length, word frequency, and lexical semantics. It showed that the number of semantic categories of words affects sentence difficulty, and semantic categories refer to the classification of words based on their meanings, and words with multiple meanings can lead to semantic ambiguity in sentences. Additionally, the greater the number of semantic categories, the more significant the variation in word meanings. This indicates that lexical complexity significantly impacts sentence difficulty.

On the other hand, Hong et al. (2020) conducted an automatic readability analysis of Chinese language textbooks, selecting 12 linguistic features from vocabulary, syntax, semantics, and discourse cohesion. The vocabulary aspect included the number of characters, the number od advanced words, and the number of two-character words. Advanced words were selected from the "8000 Chinese Words" . Content words, which have semantic meaning and can stand alone as sentences, were included in the semantic indicators.

From past research, it is evident that vocabulary significantly impacts sentence difficulty. Unfamiliar or highly complex words can affect learners' reading comprehension. Factors such as word frequency, the number of times a word appears in a single text, and learners' prior knowledge all contribute to this complexity. Besides, lexical semantics and phrase structure also influence comprehension. However, these studies lack consistency in indicator selection, rational scientific explanation for the chosen indicators, and thorough exploration of the influence of lexical complexity. These gaps highlight areas for further research.

Studies on syntactic complexity

In examining sentence difficulty, syntactic complexity holds a significant influence. Thoma & Gilbert (1967) explored the impact of sentence structure on reading comprehension, and Shen & Tao (2011) indicated that even if learners possess a large vocabulary, understanding sentence structures is crucial for text comprehension. This suggests that grammatical knowledge can more accurately predict reading ability. Carrey & Connie (2015) further demonstrated the importance of grammatical ability in developing reading skills for second language learners. They found that grammar and reading comprehension are not limited to the learner's native language; second language grammar competence significantly predicts the correlation between sentence and text comprehension, reaffirming the importance of grammar.

In the context od Chinese as a second language, analysis of syntactic complexity often focuses on individual grammars. For instance, Zhang (2005) suggested a teaching sequence for sentences using the character "bi (compare)" based on teaching grammar theory and a corpus of sentence structures with comparative functions. Her research found that introducing comparison sentences was the most commonly used expression. Hong et al. (2016) included the usage frequency of "bei (was/were done by)" sentences in their readability analysis indicators based on children's acquisition theory.

In summary, regarding the impact of syntactic complexity on sentence difficulty, studies by Thoma & Gilbert (1967) and Shen & Tao (2011) found that grammatical knowledge outperformed vocabulary knowledge in predicting text readability. Research on Chinese syntactic difficulty often focuses on individual grammar. This indicates that there is no consensus on evaluating syntactic complexity, and indicator selection is based on individual research experiences. Therefore. Establishing unified syntactic difficulty indicators is both important and urgent.

2.2 Corpus-based studies

In previous research, the majority of studies have focused on single linguistic features (Spache, 1953; Kenneth & Susan, 1973; Ye & Qiu, 2008). The selection of linguistic features often relies on qualitative research or personal experience, lacking objectivity. Therefore, this study aims to establish sentence difficulty while conducing corpus-based quantitative research to avoid subjective influences on objective development.

A corpus is a large collection of texts gathered according to specific principles and stored in computer files. It allows users to search for words, phrases, or sentences to observe particular linguistic phenomena (Chen, 2017). Chen & Bai (1998) proposed using a corpus-based learning approach to detect unknown words in Chinese and highlighted the advantages, including automatic inductive learning, automatic rule performance evaluation, and dynamic rule balancing of recall rates and precision rates.

Hong et al. (2012) investigated the event structures of transitive verbs such as "chi (eat)" , "wan (play)" , "huan (change)" and "shao (burn)" , under the MARVS theory (module-Attribute Representation of Verbal Semantics). They utilized the "Chinese Word Sketch" to observe specific nouns commonly used with these transitive verbs and analyzed their event structures.

Wu et al. (2019) argued that to standardize and systematize Chinese language teaching and establish clear grading standards and ability indicators, a corpus that provides real language contexts as an essential tool. The frequency of words in a corpus indicates their commonality. Furthermore, language learners can use the real contexts provided by the corpus to review their errors and gaps in language learning.

Last few years, to combine basic theories and practical applications in language teaching, more and more research would combine digital technology with the research methods of corpus linguistics. Hong (2021) built the Digital Platform for Chinese Grammar through the Chinese hierarchical grammar bank. Wang & Williams (2024) explore how artificial intelligence and language models can effectively assist in Chinese language teaching, then use Chat GPT as language model to design professional Chinese language courses, and to explore solutions to related challenges.

A corpus combines technology to analyze linguistic patterns and usage, making it a core element of language technology and computer-assisted language learning. Based on the aforementioned discussion, This research would compare different corpus, using learning model to validate the effectiveness of lexical and grammatical indicators.

2.3 Machine learning model

Lexical complexity and grammatical complexity have a significant impact on the difficulty of Chinese sentences, which in turn influences text readability. Research on text readability using machine learning models began as early as the 1950s, with continuous updates to the development and types of models. Since machine learning models can be used to predict text readability and the key factors influencing readability stem from understanding lexical and grammatical indicators of sentence difficulty, this project aims to validate the effectiveness of lexical and grammatical indicators on sentence difficulty using machine learning models.

The commonly used machine learning models for text readability include the generalized linear model (GLM), support vector machine (SVM), and models enhanced through representation learning, such as Bidirectional Encoder Representations from Transformers (BERT) and Hierarchical Attention Network (HAN) (Devlin et al., 2018; Yang et al., 2016).

Traditional readability models such as Flesch Reading Ease (Flesch, 1948), Dale-Chall readability formula (Dale & Chall, 1948), and SMOG (McLaughlin, 1969) focus on matching suitable reading materials to appropriate readers (Dubay, 2004). The most widely used readability formulas are Flesch's Reading Ease Formula and the Dale-Chall Formula, both of which are generalized linear models (GLM). These models use regression analysis with various linguistic indicators to estimate text difficulty or provide a formula to convert ability scores to estimate suitable reading levels (Fry, 2002; Klare, 1984).

In recent years, many researchers have adopted SVM as a machine learning model to address text readability issues (Feng et al., 2010; Sung et al., 2015). These studies demonstrate that using Structural Risk Minimization (SRM) (Vapnik & Chervonenkis, 1974) allows for the identification of small sample data (support vectors) that represent the entire training model. By finding the hyper-plane, data can be classified and effectively predicted. SVM is an artificial intelligence learning model that can correspond to the relationship between data features and defined text categories, train on relevant input indicators, and validate the difficulty levels of defined texts. SVM can map linearly nonseparable data into multi-dimensional space, providing better accuracy in classification.

In summary, applying machine learning models to text readability analysis is not only suitable but also a growing trend. Therefore, this study will use machine learning models to validate the effectiveness of sentence difficulty indicators. Specifically, this study intends to use SVM as the training algorithm for the prediction model. This choice is due to its successful application in text readability research, and because SVM, compared to deep learning algorithms, is a shallower machine learning algorithm (Kim et al., 2018). Thus, it requires a relatively smaller amount of training data, making it more suitable for this study.

3. Research methodology

3.1 Research design

"Digital Platform for Chinese Grammar" and "The 8000 Vocabulary List" are the foundation of this study. "Digital

Platform for Chinese Grammar" has a clear frame of the lexical complexity and syntactic complexity, and "The 8000 Vocabulary List" provides high-frequency words for CSL learning through comprehensive integration.

Since the difficulty of "The 8000 Vocabulary List" was graded from CEFR, to increase objectivity, the HSK and CPGS was used for comparison. Additionally, considering that the actual context involves expressions in Chinese sentences, the language learning model of CRIE 3.0 was employed to determine the difficulty of various grammatical elements in practical application. However, the standards for various difficulty indicators exhibit slight differences. SRM was used to validate the grammatical difficulty of all the aforementioned indicators, ensuring all the grammars exist in both teaching and life, and that their difficulty levels are appropriate. Based on the results above, this study establishes a method for calculating sentence difficulty. Considering that the ability to construct complex sentences generally requires a proficiency level of at least CEFR B1, difficulty levels "The Fourth-Grade Chinese Textbooks Used in Taiwan" and "The third and fourth volumes of A Course in Contemporary Chinese" were used as the basis for native speaker corpora and CSL learner corpora, respectively, for paraphrasing tests.

Subsequently, sentences with varying levels of lexical and grammatical complexity were randomly assigned for testing by CSL learners, and the Likert scale was used to provide feedback to instructors, comparing the difficulty of sentences before and after paraphrasing. Through the test and the Likert, this study established the validity and reliability of the difficulty calculation method.

3.2 Materials

Digital platform for Chinese grammar

In establishing sentence difficulty indicators, this study examines past literature and the semantic features of Chinese, integrating both lexical complexity and syntactic complexity to comprehensively measure sentence difficulty. The "Digital Platform for Chinese Grammar" aims to integrate "teaching and learning" by compiling and analyzing grammars. This platform has identified and categorized 339 grammars based on frequency and practicality into three main types: sentence types, sentence structures, and main sentence components. Each category is further subdivided into sublevels: grammars, subpoints, and detailed categories. Therefore, both lexical complexity and syntactic complexity can be referenced for their frequency of use by it and serve as the corpus source for this study.

The 8000 vocabulary list

"The 8000 Vocabulary List" is a core vocabulary selection referencing various international Chinese learning indicators, Chinese teaching materials, and corpora. These core words, frequently encountered by learners of Chinese as second language, are tagged for difficulty according to the Common European Framework of Reference for Languages (CEFR). In this study, besides consulting past literature and the "Digital Platform for Chinese Grammar," "The 8000 Vocabulary List" serves as an important reference for examining the difficulty standards of various vocabulary levels.

The study categorizes the "8000 Chinese Words" into seven levels, and these classifications are compared with the "Digital Platform for Chinese Grammar" and used as the basis for categorizing lexical complexity

Chinese readability index explorer 3.0 (CRIE 3.0)

Readability refers to the degree to which reading materials can be understood by readers. The readability of an article not only affects the reader's understanding of the content but also influences the learners' absorption of knowledge and grasp of information. The Chinese Readability Index Explorer 3.0 (CRIE 3.0) is an automated system for analyzing various features of texts, making it a practical tool for text analysis. Notably, the system can effectively identify the difficulty of Chinese vocabulary, which helps confirm the detailed difficulty levels of "sentence structures" and "sentence components" in the "Digital Platform for Chinese Grammar."

Chinese proficiency grading standards for international Chinese language education(CPGS)

The "Chinese Proficiency Grading Standards for International Chinese Language Education"(CPGS) serves as a reference standard for determining difficulty levels. This standard clearly defines specific quantitative indicators for syllables, Chinese characters, vocabulary, and grammar. The difficulty levels are categorized into three main categories: elementary, intermediate, and advanced, which are further subdivided into nine levels, forming a comprehensive three-category, nine-level Chinese proficiency grading system. This standard is used as the foundational reference for verifying the difficulty of grammars in this study.

Chinese proficiency test (Hanyu Shuiping Kaoshi, HSK)

The Hanyu Shuiping Kaoshi (HSK) is administered by the Chinese Proficiency Test Center of Beijing Language and Culture University based on the "Chinese Proficiency Grading Standards for International Chinese Language Education." The old version of the HSK divided the levels into six categories, while the new version is entirely based on the three-category, nine-level concept of the "Standards." This new version subdivides the levels from six to nine categories. Therefore, this study will use the HSK level configuration as the basis for determining the difficulty scores of various grammars within sentence structures and components.

Machine learning model

Among the various types of machine learning models, recent years have seen experts and scholars in the field of text readability employing the SVM (Support Vector Machine) model (Feng et al., 2010; Sung et al., 2015). The main advantage of this model lies in its ability to classify data and make effective predictions by identifying a hyperplane through training on small sample data. This characteristic makes SVM a highly accurate artificial intelligence learning model for predictive data analysis. Therefore, to verify whether the selected lexical and grammatical indicators are effective in influencing sentence difficulty, this study intends to use Smart Reading Model (SVM) as the machine learning model for validation. SRM include Chinese as second language (CSL) books for Chinese Learners, and it can provide suitable books recommendation according to learners' Chinese level. Thus, the study used SRM to validate the difficulty levels.

The fourth-grade Chinese textbooks used in Taiwan

The design of Taiwanese elementary school textbooks emphasizes diverse literary forms and cultural enrichment for the education of Chinese as a Native Language (CNL). The fourth-grade textbooks, in particular, introduce the concept of article passages, focusing on learners' abilities to handle passage structures and enhance reading comprehension skills. Since second language learners at the B1 to B2 level of the CEFR possess a substantial vocabulary and solid grammatical knowledge, their learning process shifts from short sentences to longer sentences and extended paragraphs, aligning with the learning trajectory of Taiwanese fourth-grade students. Additionally, their language proficiency levels match. Therefore, the texts from fourth-grade Mandarin textbooks are selected as the native language texts for this study.

The third and fourth volumes of a course in contemporary Chinese

"A Course in Contemporary Chinese" is currently one of the widely used textbooks for second language learners. Published by the Mandarin Training Center at National Taiwan Normal University, it is specifically designed for learners studying at overseas high schools or universities. The series consists of six volumes, with the levels aligned with the CEFR standards from A1 to C1, and the ACTFL standards from Novice to Superior. This study selects volumes three and four as the texts for second language learners. The content of these volumes primarily consists of extended dialogues and includes extensive training in written language and passages. Learners need to have reached at least a B1 level, possessing a substantial vocabulary and language knowledge, as well as the ability to handle longer sentences, to serve as indicators for assessing second language learners' language proficiency in this study.

4. Results and discussion

This study uses the "Digital Platform for Chinese Grammar" and "The 8000 Vocabulary List" as important reference bases to derive indicators related to sentence difficulty. The difficulty indicators, particularly those within the "sentence structure" and "sentence components" categories of the "Digital Platform for Chinese Grammar", are ranked as shown in the diagram below.

In **Figures 1** and **2**, indicators highlighted in yellow represent items that have not yet been established in the "Digital Platform for Chinese Grammar." These were sorted and organized by cross-referencing "The 8000 Vocabulary List".

To objectively validate the difficulty ranking of various grammars, we consulted the grading configuration of grammars by CRIE 3.0, HSK and "Chinese Proficiency Grading

Standards for International Chinese Language Education" and compared the difficulty levels of grammars in the CEFR and HSK, both of which primarily consider the learning process of second language learners.

Consequently, there are several differences in the difficulty ranking after comparison, as shown in **Figures 3** and **4**. For example, within "sentence structure," the "purpose complex sentence" is not the first category acquired under "modification complex sentence" but the last. In the "sentence components" category, the "time noun" under "noun" was adjusted from the third level of difficulty to the most difficult. Moreover, based on the difficulty level at which grammars first appear, various categories can be divided into 1 to 4 difficulty levels.

In establishing the difficulty of sentences, this study, drawing from previous literature, determined that assessing the difficulty of a sentence requires measuring both lexical complexity and syntactic complexity, this also validate by **Figures 3** and **4**. To establish the difficulty of every single grammar in this project, CEFR was used as a criterion for categorizing grammars by difficulty, and SRM would be utilized for validation. If the difficulty of the grammar was divided appropriately, then the grammar would be included in the difficulty table. These categorizations were then compared with those of "Chinese Proficiency Grading Standards for International Chinese Language Education" and the difficulty levels in HSK. Taking part of CEFR level B2 grammars from difficulty table as a example. Grammars from serial number 1 to serial number 5 are presented as an example in **Table 1**.

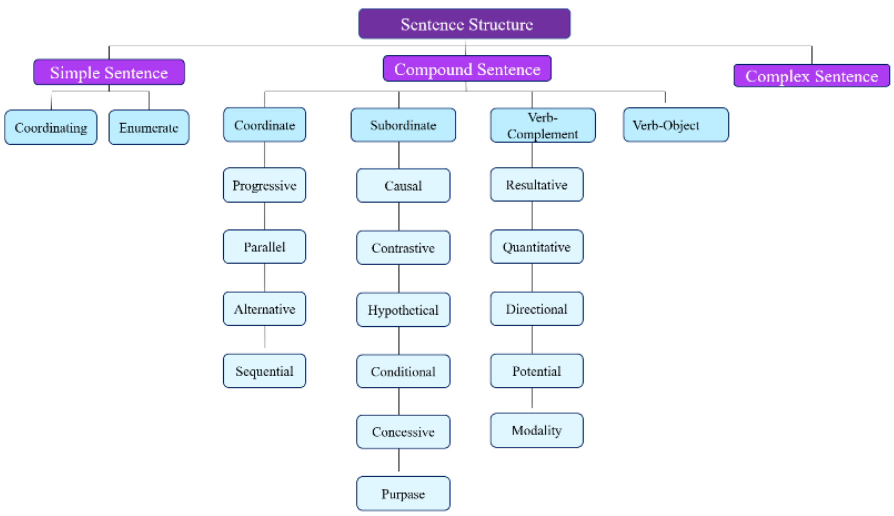

Figure 1. Difficulty Ranking of Sentence Structure.

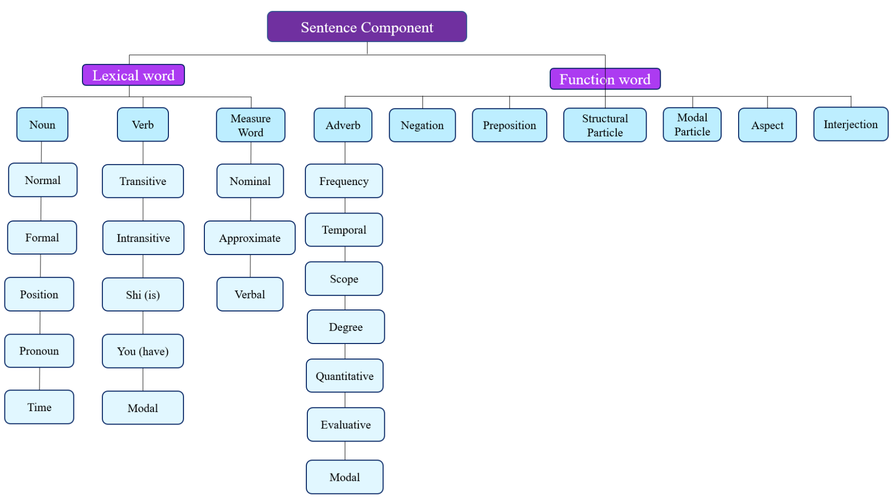

Figure 2. Difficulty Ranking of Sentence Component.

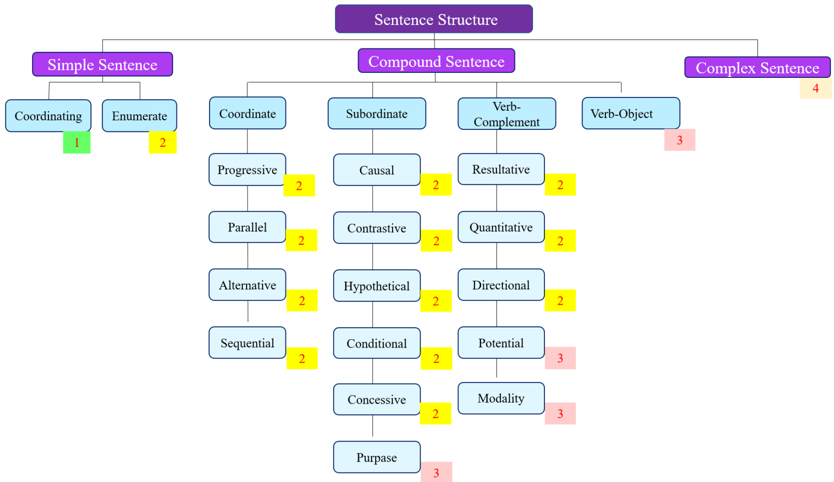

Figure 3. Adjustment of Difficulty Ranking of Sentence Structure.

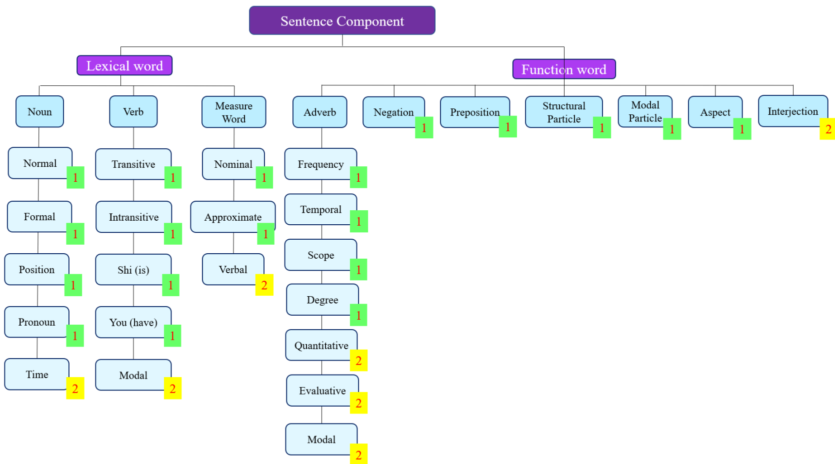

Figure 4. Adjustment of Difficulty Ranking of Sentence Component.

According to **Table 1**, CPGS and HSK are currently the most detailed reference standards for difficulty levels, which verify these CEFR level B2 grammars are effective in influencing sentence difficulty through the SRM modal. The difficulty score for each grammar is determined based on the difficulty levels of both standards. If there is a discrepancy between the scores, considering the learning situation of second language learners, the grammar is deemed to require a longer time to master and apply. Therefore, the higher difficulty score between the two standards is adopted as the difficulty score for that grammar.

To integrate the scoring indicators for various categories effectively and comprehensively reflect sentence difficulty, the calculation method of "Grammar Level * Distribution Proportion of Grammar " is designed for determining difficulty.

The distribution proportion of grammar is presented in **Table 2**, indicating that 291 grammars correspond to the difficulty levels of CEFR and HSK. The lower the level, the higher the number of grammars. For example, at level one, there are a total of 73 grammars, accounting for 25.09% of the overall proportion. Moreover, grammars at lower levels have higher frequencies of usage. Therefore, this project uses the distribution proportion of grammars as weights, where lower values indicate higher difficulty. An example of the calculation result is shown in **Table 3** below.

From the **Table 3**, the higher the score, the lower the ranking. Although the total grammatical difficulty of these five sentences added up to ten, they utilized different levels of grammars. Calculating these sentences with the calculation method of "Grammar Level * Distribution Proportion of Grammar", it was obvious that the scores were different.

To ensure the validity and reliability of the difficulty indicators, this study will rewrite texts from the fourth-grade

Table 1. CEFR B2 Grammars from Number 1 to Number 5.

Number	Degree	Grammar	SRM model	CPGS	HSK	Category	Difficulty level	Score
1	B2	wu lun (regardless)	V	4	4	Condition	2	4
2	B2	ji shi. . . sye. . . (even though)	V	5	5	Concession	2	5
3	B2	chu fei (unless)	V	5	5	Condition	2	5
4	B2	fo ze (otherwise)	V	4	4	Adversative	2	4
5	B2	ning ke. . . ye (bu) yao (rather than)	V	7	6	Alternative	2	7

Table 2. Distribution Ratio of Grammars' Level.

HSK / CEFR	1	2	3	4	5	6	789	Total
Pre A1	23	0	0	0	0	0	0	23
A1	32	14	1	0	1	0	0	48
A2	17	36	26	5	1	0	0	85
B1	1	7	21	20	18	5	3	75
B2	0	0	7	16	20	7	10	60
Total	73	57	55	41	40	12	13	291
Percentage	25.09%	19.59%	18.90%	14.09%	13.75%	4.12%	4.47%	100%

elementary school textbooks, as well as from the third and fourth volumes of the "A Course in Contemporary Chinese", corresponding to levels B1 and B2. The texts will be rewritten based on aspects such as lexical complexity, syntactic complexity, and changes in difficulty level. There will be eight types of revisions in total. Experienced Chinese language teachers with more than five years of teaching experience will be asked to distinguish between the original texts and the revised texts in terms of difficulty to assess the expert reliability.

The questionnaire design for expert reliability is presented in **Figure 5** below. The questionnaire comprises 200 sets of questions, each consisting of an original sentence and a revised sentence, and expert is a CFL teacher who has teaching over five years. Teachers will need to determine which sentence is more difficult and provide a rating of the difficulty level between the two sentences, along with relevant suggestions. **Table 4** is shown as the example question of Likert Scale of Expert Reliability for the formal questionnaire.

After the completion of the questionnaire by the Chinese language teachers, statistical analysis will be conducted on the responses of 10 teachers. Box plots will be used to illustrate the answering status of each type of question set. As an example, box plots will be presented for the question sets that involve revisions aimed at reducing the difficulty level in terms of lexical complexity compared to the original text. The box plot presentation is shown in **Figure 5** below.

According to the sentence difficulty index model, sentences with higher difficulty scores fall within the range of 3 to 5 points, while sentences with lower difficulty scores fall within the range of 1 to 3 points. The consistency rate for each question set reaches 96.70%. For question sets with significant score discrepancies, after revisions based on expert suggestions, expert evaluations were conducted again. The revised sentences aligned more closely with the difficulty scores indicated by the difficulty index and the discernment results of the teachers, as evidenced in **Table 5** below. The consistency rate among teachers' responses significantly improved.

Next, this study referenced the reading question types from the TOCFL (Test of Chinese as a Foreign Language). Base on the content of the 200 rewritten texts, corresponding questions and options were designed. These questions were evenly divided into four sets and distributed online as a questionnaire. Forty learners, each with a proficiency level of B1 or above, were asked to answer these questions randomly. The overall performance results were presented as **Table 6** below.

Table 3. Example of Sentence Difficulty Calculation.

Sentence	Score	Ranking
Ta e **si(3) le, ba(4)** zhuo shang de shi wu **dou(1)** chi **wan(2) le**. (She was starving and ate all the food on the table.) (**3***18.90%+**4***14.09%+**1***25.09%+**2***19.59%)	1.77	5
Yin wei(2) ta e **le(2)** san tian, **suo yi ba(4)** zhuo shang de shi wu chi **wan(2) le**. (Because she hadn’t eaten for three days, she ate all the food on the table.)	1.74	4
Ta **tai(1)** e **le**, **ba(4)** zhuo shang de shi wu **dou(1)** chi le **hai shi(4)** e. (She was so hungry that she ate all the food on the table and was still hungry.)	1.63	2
Ta **ji shi(5)** chi **wan(2) le(1)** bing xiang li suo you de shi wu, **ye ke yi(2)** ji xu chi. (Even if she ate all the food in the refrigerator, she could continue to eat.)	1.72	3
Ta **jiu suan(6)** chi **wan(2) le(1)** bing xiang li suo you de shi wu, **ye bu(1)** jue de bao. (Even if she ate all the food in the refrigerator, she wouldn’ t feel full.)	1.14	1

Table 4. Likert Scale of Expert Reliability.

Example Question	Sentence	Shi qi sui de Xiao Ling shi jia li de du sheng nu, fang xue hou chang chang yi ge ren zai jia, dan ta bing bu ji mo, yin wei you xu duo wang you pei ban ta. (Xiao Ling, a seventeen-year-old girl, is the only child in her family. She is often home alone after school, but she doesn’t feel lonely because she has many online friends to keep her company.)	Jia li zhi you Xiao Ling yi ge xiao hai, xia ke hou chang chang yi ge ren zai jia, dan ta bu wu liao, yin wei you hen duo wang you pei ta. (Xiao Ling is the only child in her family. She is often home alone after school, but she doesn’t get bored because she has many online friends to keep her company.)
	Please identify the difficulty of the sentences	Easy ☐ Difficult ☐	Easy ☐ Difficult ☐
	Please rate the difficulty of the sentences	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐
	Suggestion		

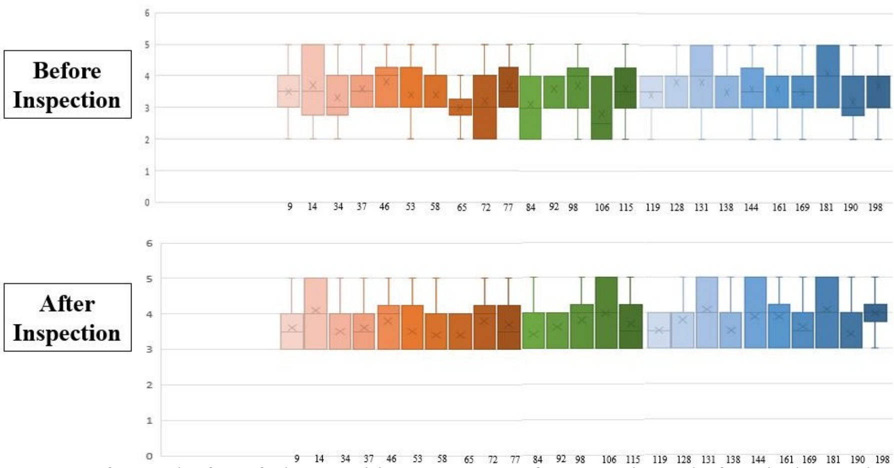

Figure 5. Before and After of The Rewriting CSL Texts of Increased Level of Lexical Complexity.

The purpose of this test was to understand the reading comprehension performance of Chinese as a second language learners when facing different levels of lexical and syntactic complexity. This would allow for further exploration or adjustment of questions with varying performance results. According to **Table 6**, the overall average accuracy rate of

Table 5. Comparison of Answer Agreement Rates Between Experts.

Teachers	Before Inspection	After Inspection
A	93.00%	100.00%
B	87.50%	99.50%
C	86.00%	97.50%
D	85.00%	99.50%
E	76.00%	97.50%
F	77.50%	95.50%
G	90.00%	98.50%
H	84.00%	98.00%
I	86.00%	96.00%
J	91.50%	97.00%

Table 6. Overall Accuracy.

		A	B	C	D	Average
Texts from CNL Textbook	Increased Difficulty	66%	73%	60%	62%	65%
	Decreased Difficulty	86%	85%	83%	77%	83%
Texts from CSL Textbook	Increased Difficulty	80%	89%	72%	72%	78%
	Decreased Difficulty	97%	90%	85%	92%	91%
Texts from CNL Textbook	Increased Difficulty	70%	82%	74%	75%	75%
	Decreased Difficulty	92%	77%	86%	90%	86%
Texts from CSL Textbook	Increased Difficulty	72%	67%	73%	63%	69%
	Decreased Difficulty	100%	83%	90%	86%	90%
Average		83%	81%	78%	77%	80%

the 40 learners was 80%, with the average accuracy rate for the four separate test sets ranging from 77% to 83%.

Based on the design of the questions, it was expected that when the complexity of vocabulary or syntax increased, the average accuracy rate for more difficult questions would be lower than for easier ones. However, upon closer examination, it was observed that in Test Set B, for native language texts focusing on syntactic complexity, the accuracy rates for both increased and decreased difficulty levels were high and very similar, at 89% and 90% respectively. Conversely, for second language texts, the accuracy rate for increased lexical complexity was higher than for decreased complexity.

To investigate these two phenomena, this study examined the response performance for each question in Test Set B. The results are presented in **Table 7**.

From the accuracy rates of each question in Test Set B, it can be seen that the overall performance accuracy is generally high. Therefore, it is necessary to examine each question individually to understand the reasons affecting the accuracy of responses. For example, regarding the aforementioned discrepancy in native language text with increased syntactic complexity, question number 40 had an unusual response pattern. The response details for this question are shown in **Table 8**.

Based on the investigation results, although the rewritten text of question number 40 contained some key grammatical elements at the B1 to B2 level, the overall difficulty was indeed significantly increased. Therefore, this was not the reason for the unusually high accuracy rate of this question. Upon closer examination of the question and its options, it was found that answer option C was more difficult than the other options. This allowed learners to find the answer through the process of elimination rather than comprehension. As a result, the difficulty of the options was adjusted

Table 7. Accuracy of Test B.

Texts from CNL Textbook (Lexical Complexity)	**Decreased Difficulty**	Number	6	12	22	29	43	48	
		Accuracy	100%	100%	10%	70%	80%	80%	
	Increased Difficulty	Number	3	11	15	32	38	41	
		Accuracy	50%	90%	90%	90%	100%	90%	
Texts from CNL Textbook (Syntactic Complexity)	**Decreased Difficulty**	Number	8	14	18	27	31	37	40
		Accuracy	90%	80%	90%	90%	90%	80%	100%
	Increased Difficulty	Number	5	9	13	23	34	39	47
		Accuracy	90%	90%	90%	90%	90%	90%	90%
Texts from CSL Textbook (Lexical Complexity)	**Decreased Difficulty**	Number	4	10	20	26	35	46	
		Accuracy	100%	80%	90%	80%	60%	80%	
	Increased Difficulty	Number	7	17	21	30	44	49	
		Accuracy	70%	50%	90%	70%	80%	100%	
Texts from CSL Textbook (Syntactic Complexity)	**Decreased Difficulty**	Number	2	16	24	33	45	50	
		Accuracy	70%	80%	50%	100%	50%	50%	
	Increased Difficulty	Number	1	19	25	28	36	42	
		Accuracy	60%	90%	90%	100%	90%	70%	

accordingly.

This study was based on the Digital Platform for Chinese Grammar and The 8000 Vocabulary List, quantifying the usage frequency and error rate of various grammars. Considering the objectivity and comprehensiveness of the indicators, comparisons were made using CRIE, CPGS, and HSK. A machine learning model was then employed for validation to establish a comprehensive Chinese sentence difficulty index.

The Chinese sentence difficulty index was divided into four levels based on the difficulty levels of grammars when they first appear, their usage frequency, and error rates. Additionally, each grammar's difficulty influenced its classification. Through Likert scales filled out by experts and assessments conducted by learners, the reliability and validity of the Chinese sentence difficulty index were confirmed. Content with discrepancies was reviewed and revised accordingly.

As a result, the Chinese sentence difficulty index can serve as a reference for designing teaching materials and as an indicator for evaluating the learning progress of Chinese as a second language learners.

5. Conclusion

Sentences are the foundation of reading comprehension, and the key factors influencing the difficulty of Chinese sentences are lexical complexity and grammatical complexity. However, past research often focused on a single aspect or analyzed using a single type. In light of this, this study aimed to provide objective and level-appropriate grammatical indicators for learners by integrating all functionally grammatical words and grammars from the Digital Platform for Chinese Grammar and The 8000 Vocabulary List, based on corpus linguistics. Comparisons were made using CRIE, CPGS, and HSK to investigate the impact of lexical complexity and grammatical complexity on the difficulty of Chinese sentences, ultimately establishing a Chinese sentence difficulty index, which was then validated using a machine learning model.

Due to the different difficulty classifications and data sources of The 8000 Vocabulary List, CPGS, and HSK, discrepancies in difficulty levels across types were resolved by assigning the difficulty of a grammatical type based on the difficulty of the grammars when they first appear. For all functionally grammatical words and grammars, considering that learners cannot master them perfectly at the beginning, the higher difficulty level of the grammar was used as the score to establish a systematic and objective Chinese sentence difficulty index.

Furthermore, the reliability and validity of the index were verified by rewriting texts from the Fourth-Grade Chinese Textbooks used in Taiwan and the third and fourth vol-

Table 8. CNL Text Content of No. 40.

40	Texts from CNL Textbook (Increased Difficulty)	Before Inspection	Zhe ge qi gai zhi xiao lao tai tai mei yue shi wu yi ding yao dao miao li shao xiang, ran er ta zui pa tomg ban xiao de zhe shi, yin ci xiao xin yi yi de yin mi ci shi, ruo shao you shu hu, zhe ji mi kong pa jiu hui xie lou chu qu. (This beggar knew that the old lady would go to the temple to burn incense on the fifteenth of every month. However, he was afraid that his companions would find it out, so he carefully kept it a secret. If he was slightly careless, this secret might be leaked.)	Wei shen me zhe ge qi gai bu xiang rang bie ren zhi dao lao taitai qu miao li shao xiang de shi qing? (Why doesn't this beggar want others to know that the old lady went to the temple to burn incense?) (A) Ta bu xi huan lao tai tai. (He doesn't like the old lady.) (B) Ta xi wang neng bang zhu lao tai tai. (He wants to help the old lady.) **(C) Ta bu xiang ran zhe ge mi mi xie lou.** **(He doesn't want this secret to be leaked.)** (D) Ta xiang yao gen lao tai tai yi qi qu. (He wants to go to the temple with the old lady.)
		After Inspection	Zhe ge qi gai zhi xiao lao tai tai mei yue shi wu yi ding yao dao miao li shao xiang, ran er ta zui pa tomg ban xiao de zhe shi, yin ci xiao xin yi yi de yin mi ci shi, ruo shao you shu hu, zhe ji mi kong pa jiu hui xie lou chu qu. (This beggar knew that the old lady would go to the temple to burn incense on the fifteenth of every month. However, he was afraid that his companions would find it out, so he carefully kept it a secret. If he was slightly careless, this secret might be leaked.)	Wei shen me zhe ge qi gai bu xiang rang bie ren zhi dao lao taitai qu miao li shao xiang de shi qing? (Why doesn't this beggar want others to know that the old lady went to the temple to burn incense?) (A) Qi gai bu xi huan lao tai tai. (The beggar doesn't like the old lady.) (B) Tong ban da suan bang zhu lao tai tai. (The companion plans to help the old lady.) **(C) Qi gai xiang shou hao zhe ge mi mi.** **(The beggar wants to keep this secret.)** (D) Tong ban xiang gen zong lao tai tai. (The companion wants to follow the old lady.)

umes of A Course in Contemporary Chinese, understanding reactions and performance on rewritten content from different perspectives. Based on the results, the Chinese sentence difficulty index from this study can serve as a reference for instructional design. It can help adjust grammatical learning according to learners' proficiency levels or prepare more suitable learning content. Teachers can also use this index to track learners' progress and provide appropriate learning materials, thereby enhancing learners' sentence comprehension and reading abilities.

In the future, empirical teaching studies should be conducted to apply the Chinese sentence difficulty index in teaching Chinese sentences. Rewriting Chinese texts for empirical research in actual teaching settings will help examine the correlation between sentence difficulty indicators and text difficulty. Feedback from learners can be used to refine sentences and validate the effectiveness of text-based teaching, ultimately developing practical application for Chinese text teaching.

Author Contributions

Author contributions: Conceptualization, JF, JN and YT; methodology, JF and YT; software, JF and JN; validation, JF and JN; formal analysis, JF and JN; investigation, JN; resources, JF and YT; data curation, JF and JN; writing—original draft preparation, JN; writing—review and editing, JF and JN; visualization, JF and JN; supervision, JF; project administration, YT; funding acquisition, JF, JN and YT. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Data Availability Statement

Not applicable.

Funding

National Science and Technology Council.

Acknowledgement

This study is supported by the National Science and Technology Council, Taiwan, R.O.C., under Grant no. MOST 110-2511-H-003-034-MY3. It is also supported by National Taiwan Normal University's Chinese Language and Technology Center. The center is funded by Taiwan's

Ministry of Education (MOE), as part of the Featured Areas Research Center Program, under the Higher Education Sprout Project.

References

Arya, D.J., Hiebert, E.H., Pearson, P.D., 2011. The effects of syntactic and lexical complexity on the comprehension of elementary science texts. International Electronic Journal of Elementary Education. 4(1), 107–125.

Brown, H.D., 1971. Children's Comprehension of Relativized English Sentences. Child Development. 42(6), 1923–1936.

Cheng, C.C., 2005. Cihui yuyi yu juzi yuedu nanyidu jiliang [〈詞彙語義與句子閱讀難易度計量〉Semantics and sentence reading difficulty measurement.]. In Proceedings of the Sixth Chinese Lexical Semantics Workshop. Xiamen, China, April 21.

Chen, K.J., Bai, M.H., 1998. Unknown Word Detection for Chinese by a Corpus-based Learning Method. Computational Linguistics and Chinese Language Processing. 3(1), 27–44.

Dale, E., Chall, J.S., 1948. A formula for predicting readability: Instructions. Educational research bulletin. 37–54.

Devlin, J., Chang, M.W., Lee, K., et al., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Dubay, W.H., 2004. The principles of readability. Costa Mesa, CA: Impact Information.

Feng, L., Jansche, M., Huenerfauth, M., et al., 2010. A comparison of features for automatic readability assessment. In Proceedings of the 23rd International Conference on Computational Linguistics, Beijing, China. pp. 276–284.

Flesch, R., 1948. A New Readability Yardstick. Journal of Applied Psychology. 32(3), 221–233. [Fry2002] Fry, E.B., 2002. Readability versus leveling. Reading Teacher. 56(3), 286–292.

Gordon, P.C., Hendrick, R., Johnson, M., 2004. Effects of noun phrase type on sentence complexity. Journal of Memory and Language. 51, 97–114.

Gunning, R., 1952. The technique of clear writing. New York: McGraw-Hill.

Hong, J.F., 2021. A Corpus-based Study of Development of the Digital Platform of Chinese Grammar Bank and its Assistant in CSL. Journal of Chinese Language Teaching. 18(1), 59–87.

Hong, J.F., Ahrens, K., Huang, C.R., 2012. Event structure of transitive verb: a MARVS perspective. International Journal of Computer Processing of Languages. 24(01), 37–50.

Hong, J.F., Sung, Y.T., Tseng, H.C., et al., 2016. A multilevel analysis of the linguistic features affecting Chinese text readability. Taiwan Journal of Chinese as a Second Language. 13, 95–126.

Hong, J., Peng, C., Tseng, H., et al., 2020. Linguistic Feature Analysis of CEFR Labeling Reliability and Validity in Language Textbooks. Journal of Technology and Chinese Language Teaching. 11(1), 57–83.

Kenneth, I.F., Susan, M.C., 1973. Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior. 12, 627–635.

Kim, J., Zhou Y., Schiavon S., et al., 2018. Personal comfort models: predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning. Building and Environment. 129, 96–106.

Klare, G.R., 1984. Readability. Handbook of reading research. London: Routledge. Volume 1, pp. 681–744.

Klare, G.R., 2000. The Measurement of Readability: Useful Information for Communicators. Journal of Computer Documentation. 24(3), 11–25.

Laufer, B., Nation, P., 1995. Vocabulary Size and Use: Lexical Richness in L2 Written Production. Applied Linguistics. 16, 307–322.

Lu, X.F., Ai, H.Y., 2015. Syntactic complexity in college-level English writing Differences among writers with diverse L1 backgrounds. Journal of Second Language Writing. 29, 16–27.

Lu, X.F., 2010. Automatic analysis of syntactic complexity in second language writing. International Journal of Corpus Linguistics. 15(4), 474–496.

Pang, C., 2016. Hanyu juzi nanyidu yingxiang yinsu fengxi [〈漢語句子難易度影響因素分析〉Factors of the difficulty of chinese sentence]. Journal of Language and Literature Studies. 1, 18–19.

Shen, Y.L., Tao, W., 2011. The Relative Significance of Vocabulary Breadth and Syntactic Knowledge in the Prediction of Reading Comprehension Test Performance. Chinese Journal of Applied Linguistics. 34(3), 113–126.

Siu, C.T.-S., Ho, C.S.-H., 2015. Cross-Language Transfer of Syntactic Skills and Reading Comprehension Among Young Cantonese – English Bilingual Students. Reading Research Quarterly. 50(3), 313–336.

Spache, G., 1953. A new readability formula for primary-grade reading materials. Elementary school journal. 53, 410–413.

Sung, Y.T., Chen, J.L., Cha, J.H., et al., 2015. Constructing and validating readability models: the method of integrating multilevel linguistic features with machine learning. Behavior research methods, 47(2), 340–354.

Sung, Y.T., Chen, J.L., Lee, Y.S., et al., 2013. Investigating Chinese Text Readability: Linguistic Features, Modeling, and Validation. Chinese Journal of Psycology. 55(1), 75–106.

Sun, S.Y., Wan, Y., 2016. Yuedu ceshi nandu de yingxiang yinsu yanjiu — — cong wenben juzi tuijinfangshi rushou [〈閱讀測試難度的影響因素研究──從文本句子推進方式入手〉A research on factors of reading text test difficulty.]. Language Planning. 2, 55–58.

Thoma, A.G., Gilbert, K.K., 1967. Influence of Syntactic Errors on Sentence Recognition. Journal of Verbal Learning and Verbal Behavior. 6, 692–698.

Vapnik, V.N., Chervonenkis, A., 1974. Teoriya Raspoznavaniya Obrazov: Statisticheskie Problemy Obucheniya [Theory of pattern recognition: Statistical problems of learning]. Moscow: Nauka.

Wang, C.C., Williams, J., 2024. Utilizing Chat GPT in the Development of CSP Textbooks: A Case Study of Teacher-student Collaboration in a One-on-one Classroom. 4(1), 53–96.

Wu, S.Y., Yu, D., Jiang, X., 2020. Development of Linguistic Features System for Chinese Text Readability Assessment and Its Validity Verification. Chinese Teaching in The World. 1, 81–97.

Yang, Z., Yang, D., Dyer, C., et al., 2016. Hierarchical attention networks for document classification. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, San Diego, CA, USA. pp. 1480–1489.

Ye, W.X., Qiu, L.K., 2008. Hanyu fuheci lijie nanyidu de jisuan [〈漢語複合詞理解難易度的計算〉Computing the degree of readability in understanding Chinese compounds.]. Language and Linguistics. 9(2), 435–447.