International Journal of English Linguistics; Vol. 13, No. 5; 2023
ISSN 1923-869X E-ISSN 1923-8703
Published by Canadian Center of Science and Education

A Corpus-Based Approach to Investigate the Cohesive Features Across Different Levels of CEFR

Jiexin Chen[1]

[1] School of Foreign Languages, Neusoft Institute, Guangdong, Foshan, China

Correspondence: Jiexin Chen, School of Foreign Languages, Neusoft Institute, Guangdong, Foshan, China.

Received: July 25, 2023 Accepted: September 17, 2023 Online Published: September 27, 2023

doi:10.5539/ijel.v13n5p122 URL: https://doi.org/10.5539/ijel.v13n5p122

Abstract

Despite plenty of previous studies pointing out the importance of validating the CEFR scale, scant attention has been given to the validation of the CEFR cohesion scale based on learners' corpus. This study aims to examine the cohesive features of written texts at different levels of the CEFR using a corpus-based approach. Employing the TAACO and Coh-Metrix tools, this study identified seven categories of key cohesive features, namely connectives, lexical overlap (sentence), Type-token ratio (TTR) and Density, givenness, semantic overlap, hypernymy and deep cohesion of the CEFR. The results showed that hypernymy and deep cohesion were the strongest predictors to distinguish CEFR levels and these categories generally kept a nonlinear relationship with CEFR levels. This study provides empirical evidence to further validate and refine the CEFR cohesion scale and casts light on the development of cohesive competence across different levels of the CEFR from the perspective of second language acquisition. More importantly, this study can provide pedagogical implications for learning and assessing cohesive competence.

Keywords: cohesive features, CEFR, corpus-based approach, empirical evidence, developmental trajectory

1. Introduction

In the field of language learning and language teaching, the Common European Framework of Reference for Language, Learning, Teaching and Assessment (CEFR) has become a widely accepted standard for assessing language proficiency. As one of the best-known and the most commonly used instruments to present a set of reference levels, the CEFR categorizes language proficiency into six levels from A1 to C2. Each level of CEFR has its distinctive features and criteria that learners should achieve. Among the scales of different CEFR levels, cohesion, which refers to how different parts of a text are connected, is a fundamental aspect of communicative language ability. Cohesion plays a crucial role in shaping the coherence of a text and previous research has shown that cohesion is an important feature in determining language proficiency levels (e.g., Crossley, 2020; Crossley et al., 2016a; Halliday & Hasan, 1976). Additionally, the level of cohesion in the second language (L2) writing can be measured from the ability to use different cohesive devices, which can further reflect different degrees of proficiency.

However, looking back to the previous CEFR-related studies, there exist several gaps. Firstly, even though a considerable amount of research examined how CEFR can support language learning (Glover, 2011) and language teaching (Díez-Bedmar & Byram, 2019), as well as exploring the correlation between the vocabulary and fluency scale across CEFR levels (Tavakoli et al., 2020; Tavakoli & Uchihara, 2019; Wisniewski, 2017), till now there is a paucity of research on the cohesive features that differentiate different CEFR levels. Secondly, the extant CEFR scale was made with a top-bottom method, which might lead to a lack of empirical evidence to validate the CEFR scale (Davidson & Fulcher, 2007; Fulcher, 2004). Meanwhile, some criticism pointed out that there was a dearth of empirical foundation in the CEFR scale (Hulstijn, 2007) and the descriptors of CEFR levels were unconvincing (Alderson, 2007). Therefore, with the necessity to fill in the aforementioned gaps, it is essential to examine the validity of the CEFR cohesion scale.

In total, the purpose of this study is to provide empirical evidence to support the validity of the CEFR cohesion scale of different levels with learners' corpus. In addition, the significance of this study is to provide a better understanding of the cohesive features that differentiate different proficiency levels from the perspective of second language acquisition (SLA) and provide valuable insights for language teaching and assessment.

2. Literature Review

2.1 Conceptualization of Cohesive Competence

To begin with, cohesive competence belongs to an aspect of language competence. In the literature of applied linguistics, Bachman (1990) provided a hierarchical framework for understanding and defining language competence. In general, language competence refers to a set of specific knowledge components, which are particularly utilized in communicative tasks via language. Language competence encompasses organizational and pragmatic competence (Bachman, 1990). Moreover, organizational competence further includes grammatical competence and textual competence. According to Widdowson (1978), grammatical competence focuses on the competencies in language usage, such as the knowledge of vocabulary, morphology, syntax, phonology and so on, while textual competence includes the knowledge of the conventions for joining utterances to form a text, which particularly emphasizes the cohesion organization in a text. Therefore, cohesive competence is specifically a kind of textual competence, which is of particular importance in L2 writing because it can directly influence the general comprehension of a text.

To be further, as a crucial textual component, cohesion can occur at different levels (Crossley, 2020). Cohesion usually can be classified into local cohesion at the sentence level, global cohesion at the paragraph or chapter level and text cohesion within inter-document. In this study, we mainly focus on local cohesion and text cohesion. Additionally, previous studies have shown that cohesive competence can reveal the quality of text organization and influence the readability of the text, as well as comprehensibility to the reader (Ghasemi, 2013). Crossley and McNamara (2012) indicated that L2 writers with more textual knowledge usually had a better command of coherent manner and cohesive competence. The ability to employ these cohesive textual devices in language learning, language teaching, and testing has been of great interest in students' writing (Ghasemi, 2013; McCutchen & Perfetti, 1982; Witte, 1981).

2.2 Measuring Cohesive Competence

The most commonly used method to measure cohesive competence is through the analysis of cohesive devices, which are useful and effective in helping establish connections between different parts of a text. Over the past decades, a pool of devices has been put forward to measure cohesive competence. For instance, Halliday and Hasan (1976) pointed out that cohesion comprised ways of marking semantic relationships directly by using reference, sub-situation, ellipsis, conjunction, as well as lexical cohesion. On this basis, Canale and Swain (1980) and Bachman (1990) added that cohesion consisted of conventions governing the ordering of old or new information in discourse. In addition, cohesion can also be measured by detecting how propositions were linked structurally in a text and how the literal meaning of a text was interpreted (Widdowson, 1978). Moreover, Ehrlich (1988) argued that several conditions can be used to measure cohesion in L2 learners' discourse more accurately, such as referential linking and semantic connectors. These cohesive devices are often used to measure L2 learners' cohesive competence.

In addition, another approach to measuring cohesive competence is through the analysis of lexical chains. Lexical chains refer to a series of related words or phrases that are used consistently throughout a text to maintain cohesion (Paltridge, 1994). Examining the strength and continuity of lexical chains provides insights into a writer's ability to establish connections at the lexical level. Researchers have developed automated tools, such as the Tool for the Automatic Analysis of Cohesion (TAACO; Crossley et al., 2016b) and Coh-Metrix (McNamara et al., 2010) to measure the cohesive strength of lexical chains and assess a writer's cohesive competence. The common devices extracted via these tools encompass lexical overlap (sentence), Type-token ratio (TTR), connectives and so on. In this study, seven cohesive categories were also selected using TAACO and Coh-Metrix.

2.3 Empirical Studies on Cohesive Competence

There have been a great many empirical studies on the application of cohesive competence and the development of cohesive competence. For instance, Liu and Braine (2005) investigated how Chinese undergraduate non-English majors were able to use a variety of cohesive devices in their writing. In addition, Yang and Sun (2012) explored the differences and similarities in the use of cohesive devices by Chinese EFL learners at different proficiency levels. Moreover, Abdi Tabari and Wind (2023) probed into the evolution of cohesive competence in L2 writing from the perspective of Complex Dynamic Systems Theory (CDST), pointing out that the development of certain local, global, and text cohesive features showed a non-linear manner. Meanwhile, the result highlighted the importance of examining L2 learners' cohesion development in the L2 writing context.

Apart from the aforementioned studies, the most important empirical studies lie in exploring the important

influence of cohesive competence on the overall quality of written production, which has become the subject of a considerable amount of empirical cohesion research (e.g., Chiang, 2003; Dastjerdi & Talebinezhad, 2006; Green et al., 2000; Johnson, 1992; Palmer, 1999). To be specific, some studies explored detailed cohesive devices to impact writing quality. For instance, the appropriate use of pronoun references, ellipses, and demonstratives positively correlated with essay quality (Cox et al., 1990), and the use of various connectives had a significant effect on young students' writing quality (Cameron et al., 1995). The presence of word overlap between sentences presented a positive correlation with writing quality (Jafarpur, 1991; Yang & Sun, 2012). Moreover, Crossley et al. (2016a) explored the development of local, global and text cohesion in writing and unpacked how to distinguish different L2 writing qualities by using cohesive devices. Scott and McNamara (2016) also examined the links between text quality, text elaboration and text cohesion. Besides, Kim and Crossley (2018) found that the cohesive features were reflected from both lexical and syntactic levels, which would influence writing quality.

However, even though adequate previous studies revealed that cohesion had a close relation with writing quality, the results might be inconsistent due to the different genres of the target writing tasks, such as the difference between argumentative and narrative genres (Abdi Tabari & Johnson, 2023). In addition, in previous CEFR studies, there was little research exploring to what extent these cohesive features would influence writing quality. Thus, the study will further delve into the relationship between cohesive competence and writing quality (namely the CEFR levels).

2.4 CEFR Cohesion Scale

With the powerful influence and great popularity of the CEFR, the CEFR cohesion scale has provided ample descriptors to measure cohesive competence. According to the descriptors by the Council of Europe (2018), coherence and cohesion refer to using linguistic devices to connect the separate elements of a text into an integrated text. These devices encompass referencing, substitution, ellipsis and other forms of textual cohesion, plus logical and temporal connectors and other forms of discourse markers. Furthermore, referring to these devices, the CEFR coherence and cohesion scale ranging from A1 to C2 has been constructed. The CEFR cohesion scale covers both linking words or elements and various types of cohesive devices. The specific cohesion scale is displayed in Appendix A.

Afterward, the importance of the cohesion scale stood out. On the one hand, the cohesion scale functions as an effective standard for raters to evaluate writing quality. For instance, Chiang (2003) found that cohesion was the best predictor to evaluate writing quality, and meanwhile, most raters preferred to adopt coherence and cohesion as rating indices, regardless of native or non-native speakers. On the other hand, the cohesion scale is a guide and reference for language educators to figure out the corresponding cohesive ability at different levels. Through the use of cohesive devices or cohesive items, the learners' writing at different proficiency levels can be identified (Yang & Sun, 2012).

However, some limitations exist in the CEFR cohesion scale. First, despite the popularity of the CEFR scale, there is an overwhelming lack of empirical evidence to support its validity (Wisniewski, 2017). This might be due to the scale calibration methodology conducted in the "Swiss Project" (North, 2007). In this project, teachers' decisions played a key role in the scaling process, which was inevitably subjective. Therefore, the validity of the CEFR scale, to some extent was questioned. In addition, the CEFR scale lacked evidence to distinguish L2 development and L2 proficiency (Hulstijn, 2011). It ignored the effect of CDST (de Bot & Larsen-Freeman, 2011; Freeman & Cameron, 2008), which indicates that the process of language ability development is complex, dynamic and nonlinear.

Moreover, one criticism of the CEFR is from the field of SLA. A paucity of connections existed between the SLA development stages (Pienemann et al., 1988) and the CEFR proficiency levels (Wisniewski, 2020). Meanwhile, individual differences and variability among non-native speakers were seldom considered during the CEFR scale construction (Hulstijn, 2007). To be specific, through the CEFR scale, the developmental trajectory of different levels was not clear. In other words, it was hard to know whether the developmental trajectory was linear or nonlinear and whether there were some differences among different levels.

Considering the aforementioned limitations in the CEFR scale, this study employs a corpus-based approach to validate and refine the CEFR scale, which is due to the effectiveness of previous studies. For instance, Boulton and Cobb (2017) examined the effectiveness of corpus use for second language learning or use. Language corpora can enrich reality (Gavioli & Aston, 2001). Meanwhile, researching learners' corpus enabled us to track and find out what learners achieved at different levels (McCarthy, 2016). A corpus-based approach can also provide insights into the sources of variation in the process of language use and dig out language learners'

proficiency (Gablasova et al., 2017). As a consequence, the corpora from learners' writing of different CEFR levels can present learners' real writing proficiency and the result can objectively reveal the developmental trajectory of learners' cohesive competence by extracting cohesive parameters.

Above all, this empirical study with a corpus-based method aims to lend support to the validity of the CEFR cohesion scale and unveil to what degree the most cohesive features differentiate CEFR levels. The specific questions of this study are as follows:

1) Which cohesive features correlate with the CEFR levels most?

2) To what extent do the most cohesive features differentiate different levels of CEFR?

3. Methodology

3.1 Corpus

The corpus of this study was obtained from the EF-Cambridge Open Language Database (EFCAMDAT; Geertzen et al., 2014), which is an open-access corpus available at http://corpus. mml.cam.ac.uk/efcamdat. In addition, the data of this corpus is very comprehensive and it continues to grow annually. Therefore, this corpus can provide SLA research with a huge treasure trove for CEFR-related empirical studies (Alexopoulou et al., 2017).

The data was extracted from CEFR writing essays from A1 to C1 levels because we considered that the language proficiency of the majority of ESL learners ranges from A1 to C1 levels. Furthermore, to obtain a well-balanced distribution of the final result, we extracted 200 essays at each level.

3.2 Tools for the Extraction of Cohesive Features

To analyze the cohesive features, TAACO and Coh-Metrix were employed. Based on the structure and content of the corpus, we finally extracted seven categories, namely connectives, lexical overlap (sentence), TTR and Density, givenness, semantic overlap, hypernymy and deep cohesion. These indices were chosen because a pool of studies has confirmed that they were productive predictors of L2 writing proficiency (Ferris, 1994; Grant & Ginther, 2000; Silva, 1993). A specific explanation of these categories is shown in Appendix B. Among these categories, connectives, lexical overlap (sentence), TTR and Density, givenness and semantic overlap were extracted via TAACO. To be further, Coh-Metrix, a comprehensive text analysis tool, was conducted to help analyze both surface-level and deep-level linguistic features of a text. Hypernymy and deep cohesion were extracted via Coh-Metrix.

3.3 Data Analysis

To begin with, all the selected data was transcribed and a detailed analysis of cohesive measures was conducted. To achieve an accurate measurement of cohesion, we selected 200 essays for each level of CEFR and therefore there were 1000 essays in total in our corpus. The chosen data was uploaded into the TAACO tool, and we extracted five cohesion-related parameters. Furthermore, Coh-Metrix was conducted in the meanwhile to extract hypernymy and deep cohesion. In the beginning, the descriptions of these indices should be covered. TAACO calculated 25 connectives indices and the top 20 indices with the closest relevancy were chosen in this study. The top 18 indices were chosen in lexical overlap (sentence), seven indices in TTR and Density, two in givenness and two in semantic overlap. Afterward, to answer the first research question, the Pearson correlation coefficient was calculated. Then, to answer the second research question, the statistical data was analyzed via analysis of variance (ANOVA) in Rstudio (version 4.2.1).

4. Results

In terms of the result of the first research question, cohesive features were reflected in the seven categories shown in Table 1. These categories were employed to unpack the cohesion of different CEFR levels. The degree of correlation was identified with the r value. According to Cohen (1988), the correlation values of $r < 0.1$ indicates a very weak effect, $0.1 <= r < 0.3$ weak, $0.3 <= r < 0.5$ moderate, and $r >= 0.5$ strong. Therefore, the correlation with each cohesive feature was described as follows.

Connectives encompass conjunctions, disjunctions, lexical subordinators and so on, which were necessary to link meaning groups and sentences. Using these link words correctly and flexibly was an important element in distinguishing different CEFR levels. As Table 1 shows, 10 indices were chosen from connectives, among which all_demonstratives, determiners, all_logical and unattended_demonstratives had strong correlations with CEFR levels. Besides that, the r values of the rest parameters were around 0.3 to 0.4, which indicated that these indices had a moderate correlation.

Lexical overlap (sentence), as a linkage between adjacent sentences of the text cohesion (Crossley et al., 2016b), includes adjacent sentence overlap, adjacent sentence overlap (sentence normed), binary adjacent sentence overlap, adjacent two-sentence overlap, adjacent two-sentence overlap (sentence normed) and binary two-sentence overlap. The most correlated features were adjacent_overlap_2_fw_sent_div_seg, adjacent_overlap_2_all_sent_div_seg and adjacent_overlap_fw_sent_div_seg. These indices highlighted the number of function lemma types that occurred at least once in the next sentence. The result showed they had a strong correlation with CEFR levels. Afterward, some features of binary adjacent sentence overlapping lemmas had a moderate correlation with CEFR levels, such as adjacent_overlap_binary_fw_sent, adjacent_overlap_binary_all_sent and adjacent_overlap_binary_2_all_sent.

TTR and Density focus on the lexical lemma, especially the number of lemmas (types) divided by the number of total running lemmas. To be specific, TTR and Density provide a measure of lexical diversity and lexical density. Lexical diversity conducted by TAACO reports various parts of speech categories, such as nouns, verbs and adjectives. In addition, lexical density indices indicate the proportion of the text that contains content words, like nouns, lexical verbs, adjectives, and adverbs derived from adjectives. The measure of lexical diversity can effectively discriminate essays of different CEFR levels (Treffers-Daller et al., 2016). The result showed that lexical_density_types and adv_ttr had a moderate correlation with CEFR levels. The result also found that function_ttr, lemma_ttr, function_mattr and argument_ttr had a strong negative correlation with CEFR, which meant high-level learners usually used fewer function words, argument_ttr and lemma_ttr.

Furthermore, the last categories include givenness, semantic overlap, hypernymy and deep cohesion. First, the concept of givenness reflects how much information can be retrieved from the preceding discourse, in particular from the usage of pronouns (McNamara et al., 2014). It is an index to count the number of third-person pronouns used in CEFR learners' writing, which showed a moderate correlation with CEFR levels. Then, semantic overlap is a factor in creating connections between sentences and conveying the intended meaning of the text. It functions to distinguish the similarity within a two-sentence span. From this level, the result showed that the r values of lsa_2_all_sent and word2vec_2_all_sent were about 0.3, which indicated a relatively moderate correlation. Next, hypernymy, as an index to count word specificity, is used to measure lexical sophistication and to predict writing quality. The result demonstrated that WRDHYPv had a strong effect on distinguishing the writing quality across different CEFR levels. At last, deep cohesion reflects the degree to which the text contains causal, intentional, and temporal connectives and conceptual links (Dowell et al., 2016). Deep cohesion whose r value reached .649 had a very strong correlation with CEFR levels.

Table 1. Correlation of selected cohesive indices

Category	Index	*r*
	all_demonstratives	.591**
	determiners	.568**
	all_logical	.509**
	unattended_demonstratives	.506**
Connectives	attended_demonstratives	.442**
	positive_causal	.408**
	all_positive	.408**
	lexical_subordinators	.386**
	positive_intentional	.365**
	all_connective	.321**
	adjacent_overlap_2_fw_sent_div_seg	.610**
	adjacent_overlap_2_all_sent_div_seg	.556**
	adjacent_overlap_fw_sent_div_seg	.517**
	adjacent_overlap_all_sent_div_seg	.465**
	adjacent_overlap_2_argument_sent_div_seg	.463**
	adjacent_overlap_binary_fw_sent	.444**
	adjacent_overlap_binary_all_sent	.441**
	adjacent_overlap_binary_2_fw_sent	.435**
Lexical Overlap (Sentence)	adjacent_overlap_2_pronoun_sent_div_seg	.425**
	adjacent_overlap_argument_sent_div_seg	.418**
	adjacent_overlap_binary_2_all_sent	.412**
	adjacent_overlap_binary_2_argument_sent	.395**
	adjacent_overlap_pronoun_sent_div_seg	.380**
	adjacent_overlap_binary_argument_sent	.377**
	adjacent_overlap_binary_2_cw_sent	.358**
	adjacent_overlap_2_fw_sent	.331**
	adjacent_overlap_binary_2_pronoun_sent	.311**
	adjacent_overlap_binary_pronoun_sent	.306**
	lexical_density_types	.485**
	adv_ttr	.475**
	prp_ttr	-.382**
TTR and Density	argument_ttr	-.524**
	function_mattr	-.547**
	lemma_ttr	-.693**
	function_ttr	-.722**
Givenness	pronoun_noun_ratio	.397**
	pronoun_density	.395**
Semantic Overlap	lsa_2_all_sent	.369**
	word2vec_2_all_sent	.311**
Hypernymy	WRDHYPv	.658**
Deep Cohesion	PCDCp	.649**

Note. ** stands for $p < 0.01$.

In terms of the result of the second research question, the descriptive statistics were demonstrated in Appendix C. ANOVA was adopted to understand how these cohesive features differentiated different CEFR levels. Post Hoc Tests were based on the Tukey HSD test, which can figure out more detailed information on the result of ANOVA. The results are summarized in Table 2.

For all_demonstratives, a significant difference was observed across different levels, $F(4, 995) = 164.8$, $p < .001$, $\eta^2 = .40$. For adjacent_overlap_2_all_sent_div_seg, $F(4, 995) = 137.7$, $p < .001$, $\eta^2 = 0.36$. For function_ttr, $F(4, 995) = 437.2$, $p < .001$, $\eta^2 = 0.64$. For pronoun_noun_ratio, $F(4, 995) = 145.0$, $p < .001$, $\eta^2 = 0.37$. As for the measures of WRDHYPv, $F(4, 995) = 341.2$, $p < .001$, $\eta^2 = 0.58$. In terms of PCDCp across different CEFR levels, a significant difference can be observed, $F(4, 995) = 219.6$, $p < .001$, $\eta^2 = 0.47$. Following the study of Cohen (1988), an effect size from .005 to .15 is considered small, .15 to .25 as medium and .25 and above as large. Therefore, the effect sizes of all of these six cohesive indices were relatively large, which indicated all of them

have significant distinctions with different CEFR levels.

Table 2. Summary of ANOVA results of all selected indices

Dimension	ANOVA	Eta²	95% CI
all_demonstratives	A1=A2<B1<B2=C1	0.40	[0.36, 1.00]
adjacent_overlap_2_all_sent_div_seg	A1<A2<B1=B2<C1	0.36	[0.32, 1.00]
function_ttr	A1>B1>A2>B2>C1	0.64	[0.61, 1.00]
pronoun_noun_ratio	A1<A2=B1<C1<B2	0.37	[0.33, 1.00]
WRDHYPv	A1<A2<B1=B2<C1	0.58	[0.55, 1.00]
PCDCp	A1<A2<B1<B2<C1	0.47	[0.43, 1.00]

Note. The *p*-value was set at 0.05. "= " signifies no significant differences; ">" stands for significantly larger; "<" indicates significantly less than.

Moreover, to depict the developmental trajectory of CEFR competence across different levels, the mean plots of these cohesive features were shown in Figure 1 to Figure 6, which can help us figure out the developmental trajectory of cohesive competence across different levels of CEFR (Tavakoli et al., 2020). It can be observed that most of these cohesive features showed a nonlinear relationship with the CEFR level. To be specific, as is shown in Figures 1, 5, and 6, the mean of all_demonstratives, WRDHYPv, and PCDCp generally presented an increasing tendency respectively. Figure 2 and Figure 4 showed that the mean of adjacent_overlap_2_all_sent_div_seg and pronoun_noun_ratio had some fluctuation, especially from B1 to C1 levels. Additionally, the mean of function_ttr shown in Figure 3 presented a negative correlation with CEFR levels. That is, with the language proficiency increasing, the mean of function_ttr decreased. In particular to A1 to A2 levels and B1 to C1 levels, the decreasing tendency of function_ttr was very obvious.

As a whole, the results indicated that the selected cohesive features had strong power to predict the different CEFR levels but the degree of each index may vary at different levels. Therefore, further discussion of the underlying reasons is needed.

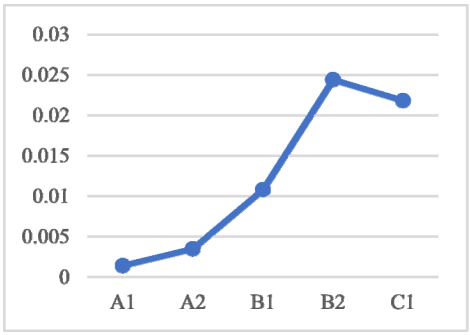

Figure 1. Mean of all_demonstratives

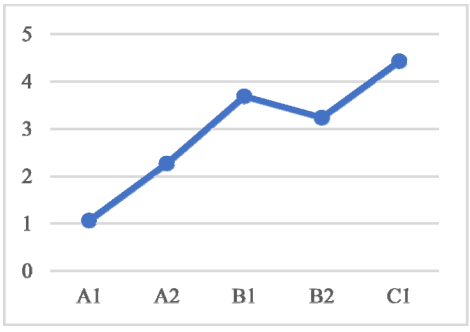

Figure 2. Mean of adjacent_overlap_2_all_sent_div_seg

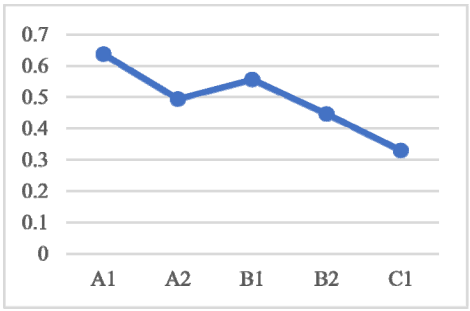

Figure 3. Mean of function_ttr

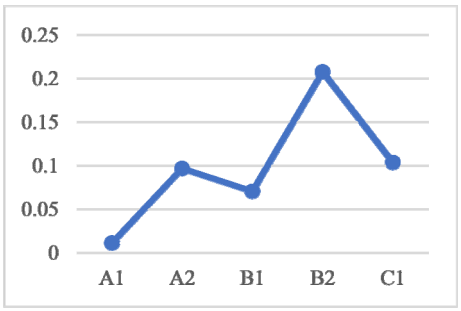

Figure 4. Mean of pronoun_noun_ratio

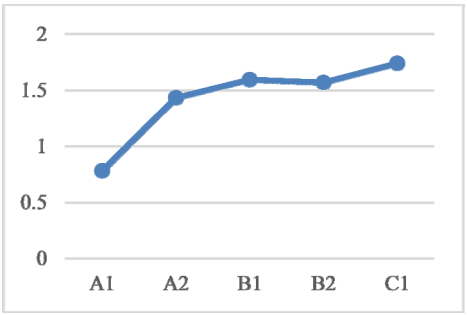

Figure 5. Mean of WRDHYPv

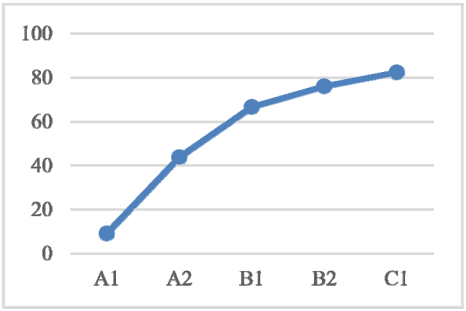

Figure 6. Mean of PCDCp

5. Discussion

5.1 Discussion on the Correlation Between Cohesive Features and CEFR Levels

The first research question examined which cohesive features correlated with the CEFR levels. In this study, seven categories of cohesive features were extracted. In line with previous studies (e.g., Englert & Hiebert, 1984; Struthers et al., 2013), the identified seven categories of this study had a close relationship with writing quality.

When it comes to the connectives and lexical overlap (sentence), the result of this study revealed that both of them positively correlated with CEFR levels. For connectives, it was important to predict the writing quality of L2 learners, especially the parameters of all_demonstratives, determiners and all_logical. The result aligns with the finding of Crewe (1990), which pointed out that the appropriate use of connectives in discourse was essential to improve cohesion. For lexical overlap (sentence), the finding showed that adjacent_overlap_2_fw_sent_div_seg was a very strong predictable parameter of the CEFR level, which indicated that L2 learners with higher language proficiency usually had a larger number of function lemma types, in that they were capable of shifting the usages of these lemmas, like article, pronoun, preposition, conjunction, modal verb, auxiliary verb and so on.

In terms of TTR and Density, the results aligned with previous findings that L2 learners with higher proficiency had more powerful lexical knowledge (Crossley et al., 2011). However, an unexpected finding was that function_ttr, lemma_ttr, function_mattr and argument_ttr had a very strong negative correlation with CEFR. This finding is consistent with the study of Crossley et al. (2016a) which pointed out that the repetition of function words (i.e., the TTR index) was a negative predictor of essay quality. One possible explanation might be that compared with the repetition of function words, more emphasis was put on lexical diversity to distinguish CEFR levels (Treffers-Daller et al., 2016).

Regarding givenness and semantic overlap, the study found that both of them positively correlated with different CEFR levels. The discovery is consistent with the study of Crossley et al. (2016a), which indicated givenness had a positive relation with text cohesion. In addition, McNamara et al. (2010) and McNamara et al. (2009) pointed out that semantic overlap can reveal learners' word knowledge and can be used as an index to judge writing quality.

Concerning hypernymy and deep cohesion, hypernymy had a strong correlation with writing proficiency, especially, noun hypernymy (Guo et al., 2013). One explanation might be that L2 learners with high proficiency tended to acquire hypernymic relations as their cognitive ability advanced (Anglin et al., 1993). For deep cohesion, it was also a strong cohesive predictor, in that as an intentional connective, deep cohesion can directly help readers to form a more coherent and deeper understanding of the text (Graesser et al., 2014).

The above aforementioned cohesive features were important indices to predict CEFR levels, but compared with the CEFR cohesion scale, there are some similarities and differences among the cohesive features. As mentioned in the literature review, the CEFR scale covers some key concepts, like referencing, substitution, ellipsis and logical connectors. These concepts are consistent with the connectives, lexical overlap (sentence) and semantic overlap of this study. However, some other different cohesive predictors, like hypernymy and deep cohesion, were extracted as cohesion predictors in this study. The result proved that hypernymy and deep cohesion can also function as strong predictors. It suggests that apart from the original indices from the CEFR cohesion scale, other specific and different parameters can also be used to complement the CEFR cohesion scale. In addition, considering the drawbacks of previous scale descriptors (Alderson, 2007; Hulstijn, 2007; Little, 2007), this study can fill in the gap by providing empirical evidence, which can enhance the validity of the CEFR cohesion scale.

5.2 Discussion on the Results of ANOVAs

In answer to the second research question pertaining to what extent the most cohesive features differentiated different levels of CEFR, there are some important findings. First, we found that the development of some cohesive features across different CEFR levels was generally nonlinear. This is in line with the finding of Tavakoli et al. (2020), which claimed that the progression of learners with higher proficiency levels might be nonlinear. This finding also caters to the previous research, which pointed out that cohesive features showed development over time, with most changes occurring in a non-linear manner (Abdi Tabari & Wind, 2023).

Second, some unexpected findings were observed. In terms of all_demonstratives, the ANOVA result indicated that A1 equaled A2 and there existed little difference between B2 and C1. One possible explanation might be that at the initial stage of language learning, there is no obvious difference between A1 and A2, while the lack of distinction between B2 and C1 might be due to a ceiling effect. In other words, the number of demonstratives used in the writing text varies from A2 to B2, but not any further. It is also possible that other parameters, like

determiners, all_logical may be able to distinguish these levels. The finding is in alignment with the study of Connor (1984), which lent support to the situation that there existed no significant differences in the number of cohesive devices used by high-proficiency and low-proficiency writers.

Afterward, the analysis of adjacent_overlap_2_all_sent_div_seg revealed that a very distinct increase appeared from A1 to B1, demonstrating a linear relationship between cohesion and proficiency level. This finding aligns with that of Crossley et al. (2016a), which proved adjacent overlap to be an essential cohesive feature. However, the result also showed there was no difference between B1 and B2. One reason for this might be that the learners at this level stay in a bottleneck period, but after that time, the learners can succeed in overcoming this obstacle with their proficiency level enhancing.

Regarding the analysis of pronoun_noun_ratio, its development in L2 writing was not consistent but presented a fluctuant pattern. It suggests that compared with A2, B1 and C1, the use of pronouns appears in the writing texts of the B2 level more frequently. However, there exists a distinct difference between B2 and C1. The C1 writer used fewer third-person pronouns later. One possible explanation might be that writers of high proficiency are more likely to find strategies to organize their texts instead of producing explicit cohesion cues (McCutchen, 1986; McCutchen & Perfetti, 1982). For instance, more proficient writers tend to use clause sentences to improve writing cohesion instead of using pronouns.

Theoretically speaking, the finding of this study caters to the core concept of CDST, which points out that language development is a nonlinear and emergent process (de Bot, 2008). Specifically, the development process would experience a stagnant period, but after that, the development can also speed up (Verspoor et al., 2011). Therefore, understanding the development stage can help us correctly perceive the alteration of different cohesion proficiency across different CEFR levels.

In closing, the above discussion compared the result of this study with the previous relevant studies and also demonstrated some explanations for the nonlinear developmental trajectory of cohesive features across different levels. The findings provide a key implication for SLA research that cohesion across different CEFR levels presents a nonlinear developmental trajectory. More importantly, even though prior studies usually constructed the scale employing a method of expert judgment (Galaczi et al., 2011) and noted a paucity of empirical evidence to support the validity of the CEFR scale by pointing out the underrepresentation of CEFR scale descriptors (McNamara et al., 2018), this study fills in the gap to fine-tune the scale descriptors and enhance the robustness of the cohesion scale, which provides more empirical evidence to further the validation of CEFR scale and casts light on the reformulation of more fine-grained CEFR cohesion scale.

6. Conclusion, Implications and Limitations

In the current study, we highlighted the importance of cohesive features in language proficiency and their relevance to the CEFR levels. We probed into seven categories of cohesive features presented in written texts at different levels of the CEFR using a corpus-based approach, which not only fills in the gap to provide empirical evidence to validate and refine the CEFR scale but also proves that the CEFR cohesion scale is effective to distinguish different levels. The study confirmed that the selected seven categories of cohesive features were associated with L2 learners' writing proficiency. More specifically, the finding demonstrates that hypernymy and deep cohesion are the strongest predictors of distinguishing different CEFR levels. The other five categories are also significant predictors of CEFR cohesion level but they generally keep a nonlinear relation with CEFR levels, which reveals that in terms of language proficiency levels, L2 learners' cohesion development at a certain stage could be either static or dynamic. The findings of the study have a significant contribution to the CEFR cohesion scale construction as this is the first study examining cohesion scale with a wide range of specific cohesive features with learners' corpus.

Most importantly, this study provides plenty of implications for future language learning, language teaching and language testing. Firstly, for language learning, this study can inform L2 learners that the development of cohesive competence is a nonlinear and dynamic process, therefore they need to follow the law of language development trajectory and master appropriate learning strategies. Secondly, from the perspective of language teaching, figuring out the important cohesive features of different CEFR levels can help teachers focus on teaching points more accurately in class. In addition, this study inspires teachers to know about the L2 learners' cohesion developmental trajectory, which can better inform teachers to provide effective instruction and intervention to L2 learners at a certain period. Specifically, this study can help teachers be aware of L2 learners' ceiling effect and take timely measures to enhance learners' writing proficiency. To be further, from the perspective of language testing, this study provides a further understanding of the cohesive features to distinguish different CEFR levels and helps with the development of the cohesion scale, which can cast light on

rater training programs and enhance rating efficiency and accuracy.

However, the study is not without limitations. In this study, we only extracted the index of lexical overlap at the sentence level rather than lexical overlap at the paragraph level due to the structure of the selected corpus. The reason lies in that each L2 writing text was combined into one paragraph and the index of lexical overlap at the paragraph level cannot be operated within one paragraph. Further studies can be conducted to overcome this limitation in the future.

References

Abdi Tabari, M., & Johnson, M. D. (2023). Exploring new insights into the role of cohesive devices in written academic genres. *Assessing Writing*, *57*. https://doi.org/10.1016/j.asw.2023.100749

Abdi Tabari, M., & Wind, A. M. (2023). Dynamic development of cohesive devices in English as a second language writing. *International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2022-0205

Alderson, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal*, *91*(4), 659–663. https://doi.org/10.1111/j.1540-4781.2007.00627_4.x

Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning*, *67*(S1), 180–208. https://doi.org/10.1111/lang.12232

Anglin, J. M., Miller, G. A., & Wakefield, P. C. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, *58*(10), 186. https://doi.org/10.2307/1166112

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, *67*(2), 348–393. https://doi.org/10.1111/lang.12224

Cameron, C. A., Lee, K., Webster, S., Munro, K., Hunt, A. K., & Linton, M. J. (1995). Text cohesion in children's narrative writing. *Applied Psycholinguistics*, *16*(3), 257–269. https://doi.org/10.1017/S0142716400007293

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*(1), 1–47. https://doi.org/10.1093/applin/I.1.1

Chiang, S. (2003). The importance of cohesive conditions to perceptions of writing quality at the early stages of foreign language learning. *System*, *31*(4), 471–484. https://doi.org/10.1016/j.system.2003.02.002

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.

Connor, U. (1984). A study of cohesion and coherence in English as a second language students' writing. *Paper in Linguistics*, *17*(3), 301–316. https://doi.org/10.1080/08351818409389208

Cox, B. E., Shanahan, T., & Sulzby, E. (1990). Good and poor elementary readers' use of cohesion in writing. *Reading Research Quarterly*, *25*(1), 47–65. https://doi.org/10.2307/747987

Crewe, W. J. (1990). The illogic of logical connectives. *ELT Journal*, *44*(4), 316–325. https://doi.org/10.1093/elt/44.4.316

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, *11*(3), 415–443. https://doi.org/10.17239/jowr-2020.11.03.01

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016a). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, *32*, 1–16. https://doi.org/10.1016/j.jslw.2016.01.003

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016b). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, *48*(4), 1227–1237. https://doi.org/10.3758/s13428-015-0651-7

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: the roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35*(2), 115–135. https://doi.org/10.1111/j.1467-9817.2010.01449.x

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2011). Predicting the proficiency level of language learners

using lexical indices. *Language Testing*, *29*(2), 243–263. https://doi.org/10.1177/0265532211419331

Dastjerdi, H. V., & Talebinezhad, M. R. (2006). Chain-preserving deletion procedure in cloze: A discoursal perspective. *Language Testing*, *23*(1), 58–72. https://doi.org/10.1191/0265532206lt318oa

Davidson, F., & Fulcher, G. (2007). The Common European Framework of Reference (CEFR) and the design of language tests: A matter of effect. *Language Teaching*, *40*(3), 231–241. https://doi.org/10.1017/S0261444807004351

de Bot, K. (2008). Introduction: Second language development as a dynamic process. *The Modern Language Journal*, *92*(2), 166–178. https://doi.org/10.1111/j.1540-4781.2008.00712.x

de Bot, K., & Larsen Freeman, D. (2011). Researching second language development from a dynamic systems theory perspective. In M. H. Verspoor, K. de. Bot & W. Lowie (Eds.), *A dynamic approach to second language development methods and techniques*. Amsterdam, Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/lllt.29.01deb

Díez-Bedmar, M. B., & Byram, M. (2019). The current influence of the CEFR in secondary education: teachers' perceptions. Language, *Culture and Curriculum*, *32*(1), 1–15. https://doi.org/10.1080/07908318.2018.1493492

Dowell, N. M., Graesser, A. C., & Cai, Z. (2016). Language and discourse analysis with Coh-Metrix: Applications from educational material to learning environments at scale. *Journal of Learning Analytics*, *3*(3), 72–95. https://doi.org/10.18608/jla.2016.33.5

Ehrlich, S. (1988). Cohesive devices and discourse competence. *World Englishes*, *7*(2), 111–118. https://doi.org/10.1111/j.1467-971X.1988.tb00224.x

Englert, C. S., & Hiebert, E. H. (1984). Children's developing awareness of text structures in expository materials. *Journal of Educational Psychology*, *76*, 65–74. https://doi.org/10.1037/0022-0663.76.1.65

Council of Europe. (2018). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment-Companion volume with new descriptors*. Cambridge, UK: Cambridge University Press.

Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly*, *28*(2), 414–420. https://doi.org/10.2307/3587446

Freeman, D. L., & Cameron, L. (2008). Research methodology on language development from a complex systems perspective. *The Modern Language Journal*, *92*(2), 200–213. https://doi.org/10.1111_j.1540-4781.2008.00714.x

Fulcher, G. (2004). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly*, *1*(4), 253–266. https://doi.org/10.1207/s15434311laq0104_4

Gablasova, D., Brezina, V., & McEnery, T. (2017). Exploring learner language through corpora: Comparing and interpreting corpus frequency information. *Language Learning*, *67*(S1), 130–154. https://doi.org/10.1111/lang.12226

Galaczi, E. D., Ffrench, A., Hubbard, C., & Green, A. (2011). Developing assessment scales for large-scale speaking tests: A multiple-method approach. *Assessment in Education: Principles, Policy & Practice*, *18*(3), 217–237. https://doi.org/10.1080/0969594X.2011.574605

Gavioli, L., & Aston, G. (2001). Enriching reality: language corpora in language pedagogy. *ELT Journal*, *55*(3), 238–246. https://doi.org/10.1093/elt/55.3.238

Geertzen, J., Alexopoulou, T., & Korhonen, A. (2014). *Automatic linguistic annotation of large scale L2 databases: The EF-Cambridge Open Language Database* (EFCAMDAT). Paper presented at the Selected Proceedings of the 2012 Second Language Research Forum: Building Bridges between Disciplines, Somerville, MA.

Ghasemi, M. (2013). An investigation into the use of cohesive devices in second language writings. *Theory and Practice in Language Studies*, *3*(9). https://doi.org/10.4304/tpls.3.9.1615-1623

Glover, P. (2011). Using CEFR level descriptors to raise university students' awareness of their speaking skills. *Language Awareness*, *20*(2), 121–133. https://doi.org/10.1080/09658416.2011.555556

Graesser, A. C., McNamara, D. S., & Cai, Z., Conley. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal*, *115*(2), 210–229. https://doi.org/10.1086/678293

Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe l2 writing differences. *Journal of Second Language Writing*, *9*(2), 123–145. https://doi.org/10.1016/S1060-3743(00)00019-9

Green, C. F., Christopher, E. R., & Mei, J. L. K. (2000). The incidence and effects on coherence of marked themes in interlanguage texts: a corpus-based enquiry. *English for Specific Purposes*, *19*(2), 99–113. https://doi.org/10.1016/S0889-4906(98)00014-3

Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, *18*(3), 218–238. https://doi.org/10.1016/j.asw.2013.05.002

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

Hulstijn, J. (2007). The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, *91*(4), 663–667. https://doi.org/10.1111/j.1540-4781.2007.00627_5.x

Hulstijn, J. (2011). Language proficiency in native and non-native speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, *8*(3), 229–249. https://doi.org/10.1080/15434303.2011.565844

Jafarpur, A. (1991). Cohesiveness as a basis for evaluating compositions. *System*, *19*(4), 459–465. https://doi.org/10.1016/0346-251X(91)90026-L

Johnson, P. (1992). Cohesion and coherence in compositions in Malay and English. *RELC Journal*, *23*(2), 1–17. https://doi.org/10.1177/003368829202300201

Kim, M., & Crossley, S. A. (2018). Modeling second language writing quality: A structural equation investigation of lexical, syntactic, and cohesive features in source-based and independent writing. *Assessing Writing*, *37*, 39–56. https://doi.org/10.1016/j.asw.2018.03.002

Little, D. (2007). The Common European Framework of Reference for languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal*, *91*(4), 645–655. https://doi.org/10.1111/j.1540-4781.2007.00627_2.x

Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System*, *33*(4), 623–636. https://doi.org/10.1016/j.system.2005.02.002

McCarthy, M. (2016). Putting the CEFR to good use: Designing grammars based on learner-corpus evidence. *Language Teaching*, *49*(1), 99–115. https://doi.org/10.1017/S0261444813000189

McCutchen, D. (1986). Domain knowledge and linguistic knowledge in the development of writing ability. *Journal of Memory and Language*, *25*, 431–444. https://doi.org/10.1016/0749-596X(86)90036-7

McCutchen, D., & Perfetti, A. C. (1982). Coherence and connectedness in the development of discourse production. *Text-Interdisciplinary Journal for the Study of Discourse*, *2*(1–3), 113–139. https://doi.org/10.1515/text.1.1982.2.1-3.113

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2009). Linguistic features of writing quality. *Written Communication*, *27*(1), 57–86. https://doi.org/10.1177/0741088309351547

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

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, *47*(4), 292–330. https://doi.org/10.1080/01638530902959943

McNamara, T., Morton, J., Storch, N., & Thompson, C. (2018). Students' accounts of their first-year undergraduate academic writing experience: Implications for the use of the CEFR. *Language Assessment Quarterly*, *15*(1), 16–28. https://doi.org/10.1080/15434303.2017.1405420

North, B. (2007). The CEFR illustrative descriptor scales. *The Modern Language Journal*, *91*(4), 656–659. https://doi.org/10.1111/j.1540-4781.2007.00627_3.x

Palmer, J. C. (1999). Coherence and cohesion in the English language classroom: The use of lexical reiteration and pronominalisation. *RELC Journal*, *30*(2), 61–85. https://doi.org/10.1177/003368829903000204

Paltridge, B. (1994). Genre analysis and the identification of textual boundaries. *Applied Linguistics*, *15*, 288–299. https://doi.org/10.1093/applin/15.3.288

Pienemann, M., Johnston, M., & Brindley, G. (1988). Constructing an acquisition-based procedure for second language assessment. *Studies in Second Language Acquisition*, *10*(2), 217–243. https://doi.org/10.1017/S0272263100007324

Scott, C. A., & McNamara, D. S. (2016). Say more and be more coherent: How text elaboration and cohesion can increase writing quality. *Journal of Writing Research*, *7*(3), 351–370. https://doi.org/10.17239/jowr-2016.07.03.02

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, *27*(4), 657–677. https://doi.org/10.2307/3587400

Struthers, L., Lapadat, J. C., & MacMillan, P. D. (2013). Assessing cohesion in children's writing: Development of a checklist. *Assessing Writing*, *18*(3), 187–201. https://doi.org/10.1016/j.asw.2013.05.001

Tavakoli, P., Nakatsuhara, F., & Hunter, A. M. (2020). Aspects of fluency across assessed levels of speaking proficiency. *The Modern Language Journal*, *104*(1), 169–191. https://doi.org/10.1111/modl.12620

Tavakoli, P., & Uchihara, T. (2019). To what extent are multiword sequences associated with oral fluency? *Language Learning*, *70*(2), 506–547. https://doi.org/10.1111/lang.12384

Treffers-Daller, J., Parslow, P., & Williams, S. (2016). Back to basics: How measures of lexical diversity can help discriminate between CEFR levels. *Applied Linguistics*, *39*(3), 302–327. https://doi.org/10.1093/applin/amw009

Verspoor, M., de Bot, K., & Lowie, L. (2011). *A dynamic approach to second language development: Methods and techniques*. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.29

Widdowson, H. G. (1978). *Teaching language as communication*. London: Oxford University Press.

Wisniewski, K. (2017). The empirical validity of the Common European Framework of Reference scales. An exemplary study for the vocabulary and fluency scales in a language testing context. *Applied Linguistics*, *39*(6), 933–959. https://doi.org/10.1093/applin/amw057

Wisniewski, K. (2020). SLA developmental stages in the CEFR-related learner corpus MERLIN. *International Journal of Learner Corpus Research*, *6*(1), 1–37. https://doi.org/10.1075/ijlcr.18008.wis

Witte, S. P., & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College Composition and Communication*, 189–204. https://doi.org/10.2307/356693

Yang, W., & Sun, Y. (2012). The use of cohesive devices in argumentative writing by Chinese EFL learners at different proficiency levels. *Linguistics and Education*, *23*(1), 31–48. https://doi.org/10.1016/j.linged.2011.09.004

Appendix A

CEFR Cohesion Scale Extracted from Council of Europe (2018)

Levels	Descriptors
C1	Can produce clear, smoothly flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices.
	Can produce well-organised, coherent text, using a variety of cohesive devices and organisational patterns.
B2	Can use a variety of linking words efficiently to mark clearly the relationships between ideas.
	Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse. Though there may be some 'jumpiness' in a long contribution.
	Can produce text that is generally well-organised and coherent, using a range of linking words and cohesive devices.
	Can structure longer texts in clear, logical paragraphs.
B1	Can introduce a counter-argument in a simple discursive text (e.g., with 'however').
	Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.
	Can form longer sentences and link them together using a limited number of cohesive devices, (e.g., in a story).
	Can make simple, logical paragraph breaks in a longer text.
A2	Can use the most frequently occurring connectors to link simple sentences in order to tell a story or describe something as a simple list of points.
	Can link groups of words with simple connectors like 'and, 'but' and 'because'.
A1	Can link words or groups of words with very basic linear connectors like 'and' or 'then'.

Appendix B

Additional Information on Selected Categories

Category	Index Name	In Text Name
Connectives	all_demonstratives	demonstratives
	determiners	determiners
	all_logical	all logical connectives
	unattended_demonstratives	unattended demonstratives
	attended_demonstratives	attended demonstratives
	positive_causal	positive causal connectives
	all_positive	all positive connectives
	lexical_subordinators	lexical subordinators
	positive_intentional	positive intentional connectives
	all_connective	all connectives
Lexical Overlap (Sentence)	adjacent_overlap_2_fw_sent_div_seg	adjacent sentence overlap function lemmas (sentence normed)
	adjacent_overlap_2_all_sent_div_seg	adjacent two-sentence overlap all lemmas (sentence normed)
	adjacent_overlap_fw_sent_div_seg	adjacent sentence overlap function lemmas (sentence normed)
	adjacent_overlap_all_sent_div_seg	adjacent sentence overlap all lemmas (sentence normed)
	adjacent_overlap_2_argument_sent_div_seg	adjacent two-sentence overlap noun and pronoun lemmas (sentence normed)
	adjacent_overlap_binary_fw_sent	binary adjacent sentence overlap function lemmas
	adjacent_overlap_binary_all_sent	binary adjacent sentence overlap all lemmas
	adjacent_overlap_binary_2_fw_sent	binary adjacent two-sentence overlap function lemmas
	adjacent_overlap_2_pronoun_sent_div_seg	adjacent two-sentence overlap pronoun lemmas (sentence normed)
	adjacent_overlap_argument_sent_div_seg	adjacent sentence overlap noun and pronoun lemmas (sentence normed)
	adjacent_overlap_binary_2_all_sent	binary adjacent two-sentence overlap all lemmas
	adjacent_overlap_binary_2_argument_sent	binary adjacent two-sentence overlap noun and pronoun lemmas
	adjacent_overlap_pronoun_sent_div_seg	adjacent sentence overlap pronoun lemmas (sentence normed)
	adjacent_overlap_binary_argument_sent	binary adjacent sentence overlap noun and pronoun lemmas
	adjacent_overlap_binary_2_cw_sent	binary adjacent two-sentence overlap content lemmas
	adjacent_overlap_2_fw_sent	adjacent two-sentence overlap function lemmas
	adjacent_overlap_binary_2_pronoun_sent	binary adjacent two-sentence overlap pronoun lemmas
TTR and Density	lexical_density_types	lexical density (types)
	adv_ttr	adverb lemma TTR
	prp_ttr	pronoun lemma TTR
	argument_ttr	argument lemma TTR
	function_mattr	function word MATTR
	lemma_ttr	lemma TTR
	function_ttr	function lemma TTR
Givenness	pronoun_noun_ratio	pronoun to noun ratio
	pronoun_density	pronoun density
Semantic Overlap	lsa_2_all_sent	lsa cosine similarity (two adjacent sentences)
	word2vec_2_all_sent	word2vec similarity (two adjacent sentences)

Appendix C

Descriptive Statistics Across Different CEFR Levels

	A1		A2		B1		B2		C1	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
all_demonstratives	0.001	0.006	0.003	0.009	0.011	0.010	0.024	0.017	0.022	0.013
adjacent_overlap_2_all_sent_div_seg	1.065	0.652	2.266	1.832	3.685	2.030	3.234	1.259	4.426	1.698
function_ttr	0.637	0.099	0.493	0.088	0.556	0.080	0.446	0.069	0.329	0.045
pronoun_noun_ratio	0.011	0.035	0.097	0.099	0.070	0.059	0.208	0.127	0.103	0.067
WRDHYPv	0.781	0.457	1.433	0.274	1.593	0.217	1.567	0.195	1.740	0.208
PCDCp	9.083	21.866	43.726	35.202	66.548	32.565	75.947	26.559	82.222	23.562