English Today

cambridge.org/eng

The syntactic complexity of the Inner Circle English varieties

A data-driven approach

Research Article

Cite this article: Long,Yijun and Haitao Liu. 2026."The syntactic complexity of the Inner Circle English varieties." *English Today* 1–13. https://doi.org/10.1017/S0266078426101333

Keywords:
English varieties; syntactic complexity; language history; population mobility; language contact

Corresponding author: Haitao Liu; email: lhtzju@yeah.net

Yijun Long[1] and Haitao Liu[2]

[1]School of International Studies, Zhejiang University, Hangzhou, China and [2]College of Foreign Languages and Literature, Fudan University, Shanghai, China

Abstract

Drawing data from *the International Corpus of English (ICE)* and employing 14 quantitative syntactic complexity metrics, this study investigates the relationships between British English (BrE), Canadian English (CanE), American English (AmE), Northern Ireland English (IreNE), Southern Ireland English (IreSE), Australian English (AusE) and New Zealand English (NZE). The study found that: 1) Compared with BrE, the other six varieties demonstrate a measurable trend towards structural simplification, primarily characterized by a lower proportion of subordinate clauses (DC/T, DC/C, CT/T, C/T), fewer verb and complex noun phrases (VP/T, CN/T), shorter mean length of sentence and T-unit (MLS, MLT), and simpler sentence structure (C/S). Meanwhile, they exhibit complexification features in coordination structures (CP/C, T/S) and clause length (MLC). 2) CanE, AmE, IreNE, IreSE, AusE and NZE show similar syntactic complexity patterns, which differ from BrE. Language history and language contact resulting from population mobility are important factors influencing CanE, AmE, IreNE, IreSE, AusE and NZE to present more similar syntactic complexity patterns. The insight provided by this study is that although these English varieties all belong to the Inner Circle in Kachru's Three Circles Model, they still show differences in syntactic complexity. This finding once again shows that language is a human-driven complex adaptive system.

1. Introduction

Nowadays, English is the world's lingua franca and the language of international communication, politics, commerce, travel, the media and so on (Schneider 2007, 1). However, English has diversified into new localized varieties that coexist with indigenous languages in some countries (Crystal 2012, 29–66). As a result, there are significant differences between such institutionalized varieties (Kachru 1985, 13; Kachru 2005, 211–220). Scholars have proposed various theoretical models to classify and explain the differences within English varieties (Kirkpatrick 2007, 27; Schneider 2003, 233–281). One of the more influential models is Kachru's 'Three Circles of English', which divides English varieties into the Inner, Outer and Expanding Circles based on their historical spread, acquisition patterns and functional domains (Kachru 1985, 366–367). The present study focuses on the Inner Circle English varieties (i.e., English as a native language (ENL) varieties). The Inner Circle of the model refers to the traditional cultural and linguistic bases of English, where English is the native language of residents in regions such as Britain, the United States, Canada, Australia and Ireland (Kachru 1985, 366).

The similarities and discrepancies between various English varieties have received increasing attention in recent decades. Extensive studies have compared vocabulary, syntactic and morphological complexity across English varieties to gain new insights into their relationships (Szmrecsanyi 2009; Xing et al. 2018). Studies have found that postcolonial Englishes exhibit simpler structures than existing ones (Mwangi 2004; Akinlotan and Housen 2017). Moreover, Szmrecsanyi and Kortmann (2009) employed quantitative analysis techniques to examine 76 morphosyntactic features across various English varieties. Their large-scale comparison revealed systematic differences in morphosyntactic complexity and analyticity across varieties. In addition, Kortmann and Szmrecsanyi (2009) analyzed the morphosyntactic features of 50 English varieties. They found that high-contact English varieties are characterized by structural simplification, while low-contact English varieties are characterized by a process of complexification.

© The Author(s), 2026. Published by Cambridge University Press.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333

As a result, they proposed that not all languages (and language varieties) are equally complex. Further studies have examined the similarities and differences of various English varieties across different language structures (Brunner 2014; Markus 2020). The above studies reveal that different English varieties have varying levels of complexity. However, while previous studies have predominantly examined relationships between various varieties, the relationships among ENL varieties remain underexplored. Szmrecsanyi (2012) further explored 30 ENL varieties based on the World Atlas of Varieties of English (WAVE), revealing subtle differences in their overall morphosyntactic similarities. Existing evidence suggests distinct complexity profiles across English varieties, highlighting the need for further systematic comparison of relationships in syntactic complexity within the ENL varieties.

The syntactic complexity of Inner Circle English varieties can be effectively analyzed through quantitative measurement tools. This study measures the syntactic complexity of English varieties using the L2 Syntactic Complexity Analyzer (L2SCA). L2SCA provides a suitable framework for comparing native English varieties due to its comprehensive metrics, which assess the length of production units, subordination, coordination, phrasal sophistication and overall sentence complexity (Lu and Ai 2015). The metrics in L2SCA combine multiple structural and syntactic features into a single quantitative standard. Although this approach may seem oversimplified, it demonstrates strong predictive power, making it valuable for classifying language output (Liu and Afzaal 2021). Although previous studies on English varieties have employed various metrics, the L2SCA offers a complementary perspective on syntactic complexity. Notably, English varieties exhibit distinct complexity regardless of the specific metrics employed, supporting the conclusion that syntactic complexity varies across different varieties (Kortmann and Szmrecsanyi 2009). Recently, some scholars have utilized L2SCA to compare syntactic complexity across various language varieties in detail (Liu et al. 2023; Chen et al. 2024), demonstrating its reliability in distinguishing between language production (Khushik and Huhta 2020). This study aims to explore and explain the relationships in syntactic complexity across several Inner Circle English varieties. Thus, using L2SCA is effective in capturing the essence of syntactic complexity. Therefore, this instrument's ability to quantify structural features makes it particularly valuable for examining the syntactic complexity of different English varieties.

Studies have utilized the L2SCA to investigate syntactic complexity across different language varieties. Liu et al. (2023) explored the syntactic complexity of interpreted speech, non-native English speech and native English speech, finding that interpreted speech and non-native English speech exhibited significantly lower complexity on most measures compared to native English speech. Furthermore, Chen et al. (2024) investigated the syntactic complexity of translated English and English as a foreign language (EFL varieties) in comparison with English as a native language (ENL varieties), demonstrating distinct complexity patterns across these varieties. The above studies have validated the feasibility and scientificity of L2SCA applied to world English studies.

Language is a human-driven complex adaptive system whose changes are influenced by multiple factors (Liu 2018). These factors may include the internal properties of language structure, population size and drift, the percentage of adult learners, the genealogical relationship and geographical distribution of languages, etc. (Bentz 2018). Although British English established the linguistic foundation for global English varieties, unique linguistic features inevitably emerge as English is nativized in different contexts (Schneider 2007). Studies have found the influence of Ulster Scots and Ulster English in Appalachia (Montgomery 2006), and Irish English may have contributed to the formation of English in the Caribbean (Rickford 1986; Harris 1990), as well as in Australia, New Zealand (Trudgill 1986, 2004) and Cameroon (Bobda 2006), etc. It can be seen that some later-formed ENL varieties are influenced to some extent by the earlier-formed English varieties from the British Isles. However, previous studies have rarely investigated relationships across ENL varieties in terms of syntactic complexity. To address this gap, this study employs 14 quantitative metrics of syntactic complexity to analyze and compare seven varieties: British English (BrE), Canadian English (CanE), American English (AmE), Northern Ireland English (IreNE), Southern Ireland English (IreSE), Australian English (AusE) and New Zealand English (NZE). Drawing data from the International Corpus of English (ICE), this study employs 14 syntactic complexity metrics and interprets the findings in the context of historical and cultural influences. Combining quantitative analysis with digital humanities approaches, this study aims to provide a comprehensive understanding of the relationships between Inner Circle English varieties. The study seeks to address the following two research questions:

1. What are the syntactic complexity features of BrE, CanE, AmE, IreNE, IreSE, AusE and NZE?
2. How do BrE, CanE, AmE, IreNE, IreSE, AusE and NZE relate to each other in terms of syntactic complexity patterns?

2. Methods

This section describes the data sources and analytical procedures used in this study. It first introduces the corpora used in the analysis, followed by the syntactic complexity metrics calculated with the L2SCA tool.

2.1 Corpora

The accuracy of L2SCA is somewhat dependent on the quality of the text (Lu 2010). Because the inherent properties of spoken communication (e.g., lack of punctuation and frequent sentence fragments) may affect the reliability of syntactic complexity analysis, this study exclusively utilizes written texts from seven ICE sub-corpora: British (ICE-BrE), Canadian (ICE-CanE), American (ICE-AmE), Northern Ireland (ICE-IreNE), Southern Ireland (ICE-IreSE), Australian (ICE-AusE) and New Zealand (ICE-NZE) English corpora. The corpus text types include student writing, letters, academic writing, non-academic writing, reportage, instructional writing, persuasive writing and creative writing. All sub-corpora adhere to ICE's unified design framework (Greenbaum 1996), ensuring

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333

Table 1. Descriptive statistics of the seven corpora

Sub-corpora	Text count	Overall size	Mean size
ICE-BrE	199	412,726	2074
ICE-CanE	200	443,310	2217
ICE-AmE	200	444,254	2221
ICE-IreNE	101	222,207	2200
ICE-IreSE	99	211,826	2140
ICE-AusE	199	407,283	2047
ICE-NZE	200	584,813	2924

consistency for comparative studies between English varieties. The detailed information of the corpus is shown in Table 1.

2.2 Measures of syntactic complexity

Multiple metrics were developed to quantify syntactic complexity. As Altmann and Lehfeldt (1973) pointed out, the metric-based method provides a comprehensive view of the structure of a language. Sentence length is often used as a measure of syntactic complexity, as it is generally believed that changes in length reflect underlying structural changes (De Clercq and Housen 2017). In addition, complex sentences are typically composed of multiple simple sentences that employ subordination and coordination structures. The frequency and depth of embedding of these structures are also considered as a measure of syntactic complexity. Some language units, such as coordinate phrases and complex nominals, are considered more complicated than other language units (Lu 2010). Therefore, the high occurrence of these language structures may also be related to syntactic complexity. Based on this, we used L2SCA to calculate 14 metrics related to syntactic complexity and compared the relationships between different English varieties on these metrics.

The 14 syntactic metrics measured by the L2SCA tool can be categorized into five distinct dimensions of syntactic complexity, including length of production units (mean length of sentence (MLS), mean length of T-unit (MLT), mean length of clause (MLC)), sentence complexity (sentence complexity ratio (C/S)), subordination (T-unit complexity ration (C/T), complex T-unit ratio (CT/T), dependent clause ratio (DC/C), dependent clause per T-unit (DC/T)), coordination (coordinate phrases per clause (CP/C), coordinate phrases per T-unit (CP/T), sentence coordination ratio (T/S)), particular structure (complex nominals per clause (CN/C), complex nominals per T-unit (CN/T), verb phrases per T-unit (VP/T)). Table A1 in the Appendix is a detailed description of the 14 metrics in L2SCA. A higher score suggests greater syntactic complexity for all these measures. We use sentence (1) as an example to show how these metrics are calculated. The results are also shown in Table A1 in the Appendix.

(1) Although the statement may seem obvious, it does demonstrate the difficulties faced by any historian attempting to tackle the period. (W1A-001 from ICE-BrE)

An analysis has been conducted to answer the second research question. Cluster analysis is a multivariate method that aims to classify a sample of subjects (or objects) based on a set of measured variables into a number of different groups such that similar subjects are placed in the same group. Therefore, cluster analysis helps explore the relationships between the seven varieties in terms of syntactic complexity patterns.

3. Results

This section presents the results based on the 14 syntactic complexity metrics calculated from the seven ICE sub-corpora. First, these metrics are used to examine the syntactic complexity patterns of the seven English varieties. Statistical tests are then conducted to determine whether significant differences exist among these varieties. Subsequently, hierarchical cluster analysis is applied to further explore the relationships among the seven English varieties in terms of their syntactic complexity patterns.

3.1 Syntactic complexity across the seven English varieties

To address the first research question, this study explored the syntactic complexity of BrE, CanE, AmE, IreNE, IreSE, AusE and NZE. Table A2 in the Appendix provides the mean values and standard deviations for all 14 syntactic complexity metrics across the seven corpora. We conducted some tests to determine whether the mean complexity values differed significantly between the seven English varieties. Before performing statistical comparisons, we initially conducted the Shapiro-Wilk test for normality and Levene's test for equality of variance to assess data distribution. The results showed a lack of normal distribution, except for DC/C and CT/T metrics. Therefore, we conducted one-way ANOVA tests for normally distributed measures and Kruskal-Wallis H tests for the non-normally distributed measures. To visually represent the differences in the 14 metrics across BrE, CanE, AmE, IreNE, IreSE, AusE and NZE, a radar chart was created, as shown in Figure 1. This chart comprehensively compares the syntactic complexity patterns of English varieties by displaying all metrics simultaneously, with each color representing a specific English variety. Each axis of the radar chart corresponds to one of the 14 metrics. The shape of the seven English varieties in the radar chart visually reflects their overall distribution patterns across the 14 syntactic complexity metrics.

Metrics that showed statistically significant differences are marked with an asterisk (*) and a red dot. As shown in Figure 1, significant differences exist in 12 metrics between BrE, CanE, AmE, IreNE, IreSE, AusE and NZE (T/S, CP/C, DC/T, DC/C, CT/T, C/T, C/S, MLC, MLT, MLS, VP/T, CN/T). In comparison, no significance was found in 2 metrics (CN/C and CP/T), suggesting broadly similar rates of complex nominals per clause (CN/C) and coordinate phrases per T-unit (CP/T) across BrE, CanE, AmE, IreNE, IreSE, AusE and NZE.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333

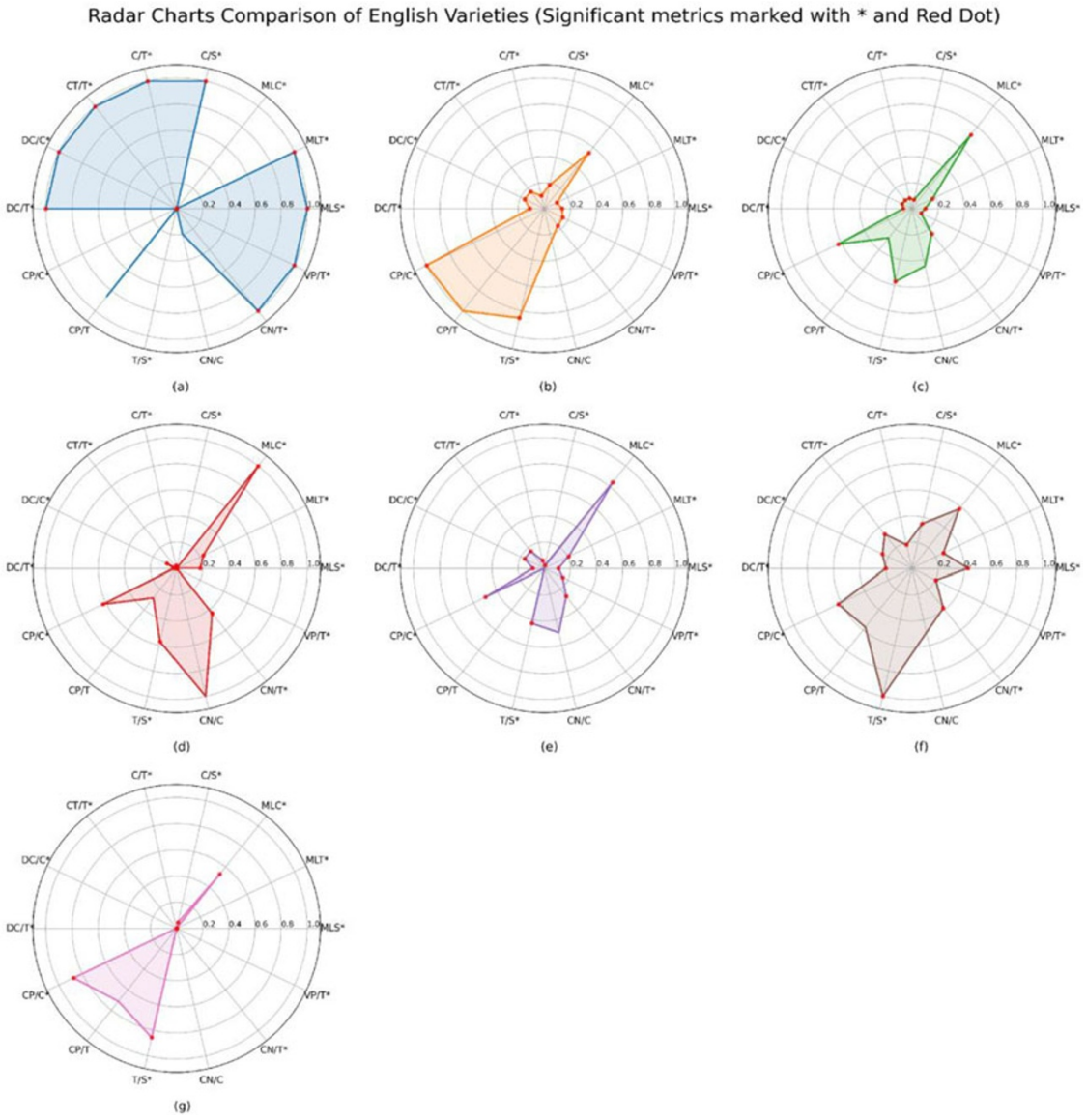


Figure 1. Normalized radar chart of 14 metrics for seven English varieties: (a) *BrE*, (b) *AmE*, (c) *CanE*, (d) *IreSE*, (e) *IreNE*, (f) *AusE*, (g) *NZE*.

Meanwhile, the distance from the center reflects the value of each variable. A closer distance to the edge indicates a higher value, while a closer distance to the center means a lower value. Among the 12 significant metrics, BrE exhibits the highest values in DC/T, DC/C, CT/T, C/T, C/S, MLT, MLS, VP/T and CN/T, indicating a tendency toward deep syntactic embeddings and longer sentence structures. High values for DC/T (dependent clauses per T-unit) and DC/C (dependent clause ratio) indicate that subordinate clauses within each unit frequently appear in BrE. Similarly, high values for CT/T (complex T-units ratio), C/T (T-unit complexity ratio) and C/S (sentence complexity ratio) reinforce this trend, reflecting that T-units or sentences contain multiple or complex clauses. High values for MLT (mean length of T-unit) and MLS (mean length of sentence) reveal that longer syntactic units are used in BrE more frequently. Moreover, the high scores in VP/T (verb phrases per T-unit) and CN/T (complex nominals per T-unit) indicate that BrE not only employs complex clausal embedding but also entails various complex phrases. Sentence (1) illustrates that BrE exhibits higher complexity in DC/T, DC/C, CT/T, C/T, C/S, MLT, MLS, VP/T and CN/T, and lower complexity in CP/C (coordinate phrases per clause), T/S (sentence complexity ratio) and MLC (mean length of clause) (see Appendix Table A1 for a detailed analysis). In sentence (1), the values of CP/C and T/S are both 0, indicating that coordinate phrases are rarely used in BrE. Compared with the sentence length (MLS=20) and T-unit length (MLT=20), the clause length of sentence (1) is much shorter (MLC=10). However, compared with BrE, CanE, AmE, NZE, IreNE and IreSE show higher complexity in CP/C, T/S and MLC, reflecting their preference for coordinate structures and longer clauses. Specifically, AmE and NZE have higher values in CP/C, AusE is more complex in T/S, and CanE, IreNE and IreSE are more complex in MLC. Three typical examples are used to illustrate the differences across these English varieties in CP/C, T/S and MLC.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333

(2) Quite plainly, however, Goethe bridged the gap between domestic worries and his international hierarchy of nations and peoples. (W1A-007 from ICE-AmE)
(3) Their findings have been summarized in Appendix one and are shown in Fig. 1C. (W1A-005 from ICE-AusE)
(4) By June of 1984 over 900 firms had been offered assistance in setting up an affirmative action program. (W1A-002 from ICE-CanE)

These examples demonstrate that while these English varieties use fewer subordinate structures than BrE, they achieve complexity through coordination or clausal length. For example, sentence (2) contains two coordinate phrases, resulting in a higher CP/C. Sentence (3) includes two coordinate sentences, hence a higher T/S. In contrast, sentence (4) is a longer clause, which contributes to a high MLC. These observations highlight distinct syntactic patterns across the English varieties. BrE exhibits a subordination-oriented character, characterized by deeply embedded syntactic structures. Its high values in DC/T, DC/C, CT/T, C/T, C/S, MLT, MLS, VP/T and CN/T collectively indicate a preference for embedded sentence structures. In contrast, AmE, NZE and AusE exhibit a coordination-oriented character, reflected in higher values on the CP/C and T/S compared with BrE. Furthermore, differences exist even within the coordination-oriented groups, such as CanE, IreNE and IreSE, which achieve high complexity through the use of longer clauses (high MLC). Like Kortmann and Szmrecsanyi (2009), the findings of this study also suggest that the syntax of all languages (and varieties of a language) is not equally complex.

While all varieties share a common core grammar, they have developed distinct preferences for realizing complexity. Compared with BrE, CanE, AmE, IreNE, IreSE, AusE and NZE show simplification in sentence complexity (C/S), subordination (DC/T, DC/C, CT/T, C/T), sentence and T-unit length (MLS, MLT) and particular structures (VP/T and CN/T), but exhibit complexification in coordination (CP/C, T/S) and clause length (MLC). These differences may be attributed to communicative adaptation in multilingual settings, a phenomenon known as Foreigner Directed Speech (FDS). In Little (2011), two groups of participants had to learn an artificial language, and they reduced the morphosyntactic complexity of this language more when speaking to 'foreigners' in the experiment than when speaking to non-foreigners, demonstrating FDS under controlled experimental conditions. Likewise, native English speakers may employ strategies to simplify or complicate their language to better communicate with speakers from diverse cultures or backgrounds. These similar adaptations may occur in real-world multilingual interactions. In linguistically diverse environments, such as Canada, the United States, Ireland, Australia and New Zealand, native English speakers may unconsciously simplify or reconstruct specific sentence patterns to facilitate communication. Therefore, we believe that under the influence of the diverse language environment, Canadians, Americans, Irish, Australians and New Zealanders will simplify or complicate the English language structure to a certain extent to achieve the purpose of communication, thereby promoting communication between different language and cultural groups. Although we propose that adaptation may be one of the potential factors explaining the simplification or complexification of syntactic structures, it is important to note that this is only one among numerous possible factors. However, we recognize that the actual dynamics are far more intricate, and there could be multiple, interwoven influences contributing to the observed difference.

Furthermore, the shape of the polygon reveals the overall patterns of each English variety. The radar charts of the seven English varieties present different patterns of syntactic complexity, each variety exhibiting a unique shape that highlights its specific characteristics. BrE exhibits a unique fan-shaped pattern, clearly distinguishing it from the other six varieties. AmE, NZE and AusE occupy the outermost positions on the CP/C and T/S dimensions. In contrast, CanE, IreNE and IreSE show more consistent and aligned shapes, particularly in the MLC dimension. Consequently, it is possible to make a general assessment of the syntactic complexity patterns of the seven English varieties by observing the shapes of their radar charts. The closer the shapes of the charts, the more similar the syntactic complexity patterns appear to be across these varieties. In the next part, this study further conducts cluster analysis on seven English varieties, aiming to explore the relationships across these English varieties based on syntactic complexity.

3.2 The relationships between seven English varieties in syntactic complexity

Given the different syntactic complexity patterns of the seven English varieties shown in Figure 1, we performed hierarchical clustering on 12 significant metrics to systematically explore the relationships between them. These 12 metrics were specifically selected because they highlighted meaningful difference in syntactic complexity, and the results from the cluster analysis based on these metrics were consistent with the cluster analysis performed on all 14 metrics. This result suggests that the two non-significant metrics do not significantly impact the overall distinction between varieties. Ward D method and Euclidean distance are adopted for the clustering analysis. Hierarchical cluster analysis (HCA) is a type of unsupervised clustering method. Unlike k-means clustering, HCA does not require a prior specification to the number of clusters. Figure 2 presents the HCA results of English varieties based on the 12 syntactic complexity metrics jointly.

In a dendrogram (which looks like a family tree), the lower the point where two branches split apart, the more closely related those branches are. The clustering results reveal that BrE is clustered into one group, while the other six varieties are clustered into another group, indicating that CanE, AmE, IreNE, IreSE, AusE, and NZE exhibit similar syntactic complexity, which differs from that of BrE. This result is consistent with the findings of Szmrecsanyi and Kortmann (2009), whose research showed an apparent divergence between BrE and AmE. Moreover, the taxonomic relation between varieties found in this study is similar to the early Strevens' 'Model of English in the World' (Strevens, 1980), which shows the genealogical relationships between English varieties through a tree diagram overlaid on a world map. In this model, BrE and AmE formed the main branches, and different English varieties are grouped according to their geographical and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333

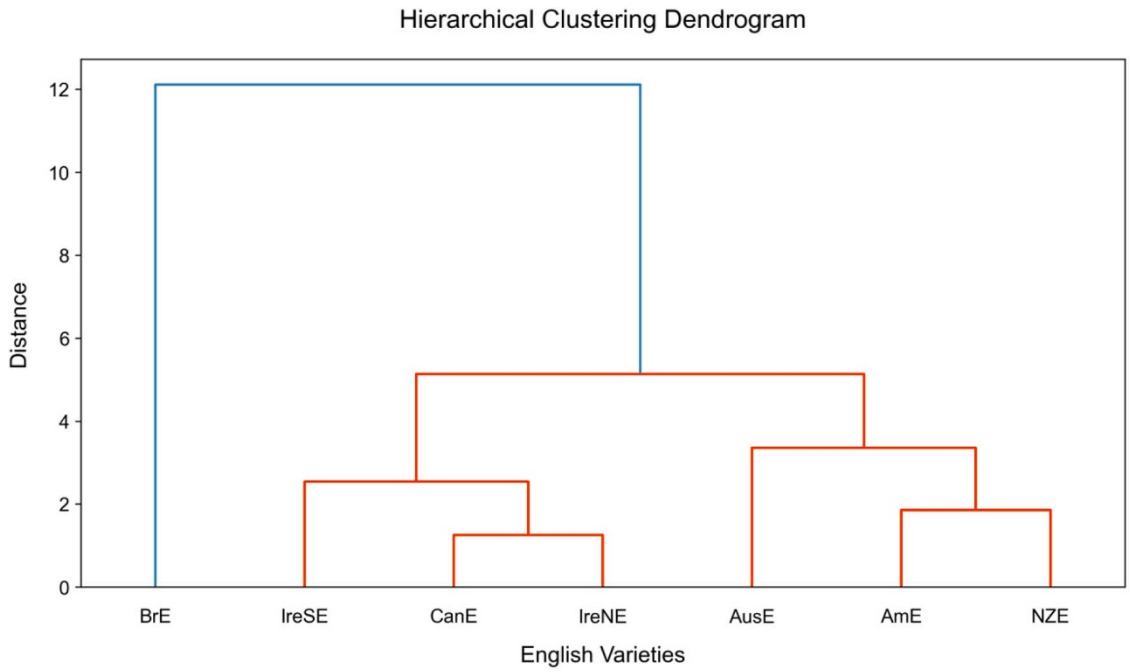


Figure 2. HCA based on 12 syntactic complexity metrics.

historical affinities. The clustering results in this study support this classification model, demonstrating the positive potential of quantitative indices in language classification tasks. Meanwhile, effect size analysis (η^2) revealed that all 12 syntactic complexity features contributed to distinguishing the three clusters, with 8 features exhibiting large effects ($\eta^2 \geq 0.93$). The most influential features were subordination metrics (C/T, DC/T, DC/C), coordination and sentence metrics (T/S, C/S) and unit length (MLT, MLC), collectively accounting for over 93% of the between-cluster variance. This result is consistent with our findings on the differences in various metrics across the seven varieties discussed in Section 3.1.

We attempt to explain why CanE, AmE, IreNE, IreSE, Aus and NZE exhibit similar syntactic complexity patterns by adopting a holistic perspective. Rather than focusing on which specific factors influence the performance of each variety on particular syntactic metrics, we aim to identify broader factors that connect these varieties and lead to similarities in their overall syntactic structure. Szmrecsanyi and Kortmann (2009) suggested that Irish English, Northern English and Newfoundland English cluster together, likely due to historical migration from Ireland and Northern England to Newfoundland. This finding provides a valuable perspective on the relationships among these English varieties, suggesting that historical population migrations and language contact may explain why different English varieties exhibit similar syntactic features. Based on this insight, we believe that CanE, AmE, IreNE, IreSE, AusE and NZE are also more similar in their syntactic complexity due to common historical factors, population mobility and language contact. Therefore, we attempt to explain why these English varieties cluster together, rather than focusing on individual factors that might influence the performance of each variety on specific syntactic metrics.

4. Discussion

This section interprets the key findings presented in the Results section and explores potential factors that may contribute to the patterns of syntactic complexity observed in the seven English varieties. It discusses how factors such as historical development, population mobility and language contact may account for the syntactic similarities among CanE, AmE, IreNE, IreSE, AusE, and NZE, as well as their differences from BrE.

4.1 Language history

The basic language framework of CanE, AmE, IreNE, IreSE, AusE and NZE originated from BrE. Therefore, the fact that these six English varieties exhibit similar syntactic complexity is closely related to the historical background in which BrE has deeply influenced them. During the 17th and 18th centuries, dialects from different regions of England and the British Isles existed in every American colony, resulting in an extensive dialect mixture and levelling process that made the English varieties across the colonies more homogeneous compared to those in Britain (Kretzchmar 2004). Moreover, England had a significant influence on language change in Ireland (Kallen 1994). Ireland and England have always had close relations due to trade in the late Middle Ages and England's invasion of Ireland in 1169 (Hickey 2007, 30). Meanwhile, English was introduced into Australia on British settlement in the 18th century and in the following decades, gradually overtook Indigenous languages to become the majority language of Australia (Leitner 2004). At the same time, English arrived in New Zealand in the late 18th century. After the Treaty of Waitangi was signed, the Māori chiefs yielded sovereignty to the British Crown (Hundt 2012). There

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333

is no doubt that BrE has had an important influence on the formation and development of NZE.

Our findings indicate that CanE, AmE, IreNE, IreSE, AusE and NZE share greater syntactic similarity than with BrE, which may reflect their inheritance of a common normative foundation from the early stages of BrE. However, this shared lineage does not preclude considerable internal diversity. Our results also reveal the differences in syntactic complexity between them, for instance, higher CP/C in AmE and NZE, higher T/S in AusE and higher MLC in CanE, IreNE and IreSE. Therefore, while their shared historical origins may explain their clustering in terms of syntactic complexity, their respective developmental trajectories further shape their unique characteristics of complexity.

4.2 Population mobility

As language is a human-driven complex adaptive system, changes in people and populations are potential motivations for language change (Liu 2018). At the same time, from the perspective of language as a complex adaptive system, language contact is a significant factor in language change, and factors influencing language contact include the population size of language users and cultural preferences (Bentz 2018). With these factors, language change is intertwined with population contact, migration and diffusion. The increased movement of people worldwide 'has produced the continued diversity of the language itself, as it adapts to the emerging cultures of different communities' (Seargeant 2012, 184). As a result, English has become multiple, meaning it does not have only one centre (the United Kingdom) that influences its form and use, but rather various centres in numerous communities (Seargeant 2012, 186). Therefore, this study argues that population mobility has caused CanE, IreNE, IreSE, AmE, AusE and NZE to influence each other.

As shown in Figure 1, CanE, AmE, IreNE, IreSE, AusE and NZE exhibit more similar syntactic complexity than BrE. Specifically, the syntactic complexity of CanE appears to be closer to IreNE and IreSE, while AmE seems to be closer to AusE and NZE, which is related to the immigration between them. Firstly, the close relationship between CanE, IreNE and IreSE was influenced in part by the mass emigration caused by the Irish Potato Famine. The Irish Potato Famine, a period of starvation and disease in Ireland from 1845 to 1852, triggered a historic social crisis and had a profound impact on Irish society and history (Kinealy 1994, 7). In 1847, over a hundred thousand emigrants left, mainly for British North America (Woodham–Smith 1992, 238). The 1851 census reported that more than half of Toronto's residents were Irish, and in 1847 alone, 38,000 Irish people migrated to this city with a population of less than 20,000 (Winder 2000). Clearly, the first major period of British immigration to Canada was predominantly Irish people. Elliott (2004, 62) cites early immigration data for the decade after 1815, showing that 58% of the 77,000 immigrants to Quebec between 1816 and 1824 were Irish. While IreNE, IreSE and CanE cluster together in our analysis, which may reflect the historical role of Irish immigration, this does not imply a more substantial overall Irish influence on CanE than on other varieties, such as AusE. The development of each variety involved multiple factors, and the clustering observed in our study highlights one of several possible historical connections.

Second, population movement between the United States, Australia and New Zealand has provided great opportunities for mutual linguistic influence among AmE, AusE and NZE. These movements, especially during critical historical periods, laid the foundation for long-term linguistic, cultural and socio-political connections. A pivotal moment in history occurred during the gold rush era of the 1850s, which can be seen as the first real contact between Australia and the United States (Korhonen 2017, 38). During this period, both Australia and the United States experienced gold discoveries that triggered global waves of migration, resulting in two-way migration between the two countries (Korhonen 2017, 38). These American migrants often attained influential roles in Australia due to their specialized skills and prior experience (Mosler and Catley 1998, 13; Fritz 2007, 44). Their prominence in the mining communities gave them the ability to exert not only economic and technological influence but also linguistic influence on the local population. At the same time, miners returning to Australia from the United States brought with them various lexical items and expressions they had encountered in the United States (Mosler and Catley 1998, 12). This bidirectional migration facilitated a flow of linguistic features, particularly vocabulary, from AmE into AusE. The study also provided support for the influence of American vocabulary on AusE (Butler 2001). Beyond the gold rush era, the increasing American presence in Australian political thought, commerce and communication technologies contributed to further linguistic and cultural exchange (Korhonen 2017, 38). Although this type of migration was only temporary in nature, the social, cultural and linguistic effects were notable (Korhonen 2017). Additionally, the widespread popularity of American television shows introduced Australians to new idioms, accents, slang and cultural references, many of which were absorbed into everyday speech, particularly among younger generations (Korhonen 2017). This influence has become more pronounced in the era of globalization, as digital media further facilitates exposure to AmE in Australia.

Similarly, the discovery of gold also played a central role in driving immigration to New Zealand during the mid-19th century. As noted by Gordon et al. (2004, 42), the New Zealand gold rush prompted an influx of migrants not only from the United Kingdom and Ireland but also from Australia and the United States. These diverse migration patterns led to increased language contact, especially between speakers of AusE and AmE. In this context, the migration of Australians to New Zealand, and vice versa, was particularly impactful. Since the early days of European colonization, Australia and New Zealand have maintained strong socio-political and economic ties, reinforced by geographic proximity and shared colonial histories. Hundt (2012) explains that one explanation why AusE and NZE developed similarities is that the migration from Australia to New Zealand and vice versa contributed to the maintenance of close socio-political ties. The history of New Zealand, and consequently of NZE, is closely intertwined with that of Australia and AusE, since many influential linguists have interpreted the historical

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333

precedence of the colonisation of Australia as having linguistic consequences on the emergence and evolution of NZE, assuming an AusE precedence over the formation of NZE (Przewozny and Viollain 2016). According to this perspective, the settlement and linguistic developments in Australia predated those in New Zealand and provided a foundational influence for the latter. As a result, many early phonological and lexical features of NZE bear strong similarities to AusE.

However, the influence of AmE on NZE should not be underestimated. As New Zealand increasingly participated in global cultural and economic systems, AmE began to exert its own influence on NZE. Studies conducted by Bayard (1989) provided some evidence of the growing influence of American English on New Zealand vocabulary. In summary, the close and historically rich economic, political, and cultural relationships among the United States, Australia, and New Zealand have facilitated sustained population movement and interaction. These movements, whether motivated by economic opportunities, such as the gold rushes, or geopolitical alliances, like those formed during World War II, have had profound effects on the English varieties spoken in each country. Migration flows have acted as channels for the exchange of vocabulary, pronunciation, idiomatic expressions, and even discourse conventions. Over time, these exchanges have contributed to the formation of distinct yet interconnected English varieties.

Through the above data, we can see that there have been varying degrees of migration between Ireland, Canada, the United States, Australia and New Zealand. The English spoken by people in these different countries has continuously interacted and integrated with the local English. This process inevitably influences these English varieties, shaping their own unique usage characteristics. Therefore, we believe that the reason why CanE, AmE, IreNE, IreSE, AusE and NZE exhibit more similar usage characteristics in syntactic complexity is the result of the mutual influence between these English varieties.

4.3 *Language contact*

Most languages in the world are the result of language contact and dialect contact. English is no exception. Kachru (1992, 6) pointed out that the international diversification of English is mainly due to its contact with other languages. Language contact primarily involves interactions between people, and the speaker subject is the primary driver of language change caused by language contact (Lim and Umberto 2016). Therefore, the language contact brought about by population movement is an important factor in the mutual influence and integration between CanE, AmE, IreNE, IreSE, AusE and NZE.

As Thomason (2001, 8) said: 'Language contact is ubiquitous, and there is no evidence that any language has developed in complete isolation from other languages.' This statement is undoubtedly true for English, which has developed from the Old English period to the present. Language contact between Germanic, Celtic and Romance dialects led to the evolution of varieties in the British Isles, making today's standardized and non-standardized English varieties contact languages (McWhorter 2008; Venneman 2011). Numerous studies have revealed the role of language contact or population contact in the mutual influence between different languages or different language varieties (Zhou and Liu 2023). Seven articles on language contact and world Englishes in the journal *World Englishes* in 2016 illustrated the connection between language contact and world Englishes from different levels, highlighting the importance of a deeper understanding of world Englishes from the perspective of language contact. In addition, Schreier and Hundt (2013) emphasized the importance of language contact in the history of English diffusion in a series of case studies on different English varieties. The study suggests that contact between Irish English and other English varieties, particularly with English dialect varieties in the British Isles, has led to its morphology and syntax primarily following the patterns found in other English varieties in the British Isles (Filppula 2008). It is reasonable to believe that population movement between Ireland, Canada, the United States, Australia and New Zealand provided opportunities for English language contact, leading to mutual influence between CanE, AmE, IreNE, IreSE, AusE and NZE, which may be an important factor in causing CanE, AmE, IreNE, IreSE, AusE and NZE to have similar patterns in syntactic complexity.

5. Conclusion

Using data from the ICE, this study analyzes and compares syntactic complexity across seven Inner Circle English varieties. Through the application of 14 quantitative syntactic complexity measures, this study seeks to identify both shared patterns and distinctive features across these Inner Circle varieties. The results indicate that, compared to BrE, CanE, AmE, IreNE, IreSE, AusE and NZE exhibit structural simplification in areas such as subordination, sentence and T-unit length, sentence complexity, verb phrases and complex nominals, while also showing increased complexity in coordination and clause length. These differences may reflect the communication strategies employed by native speakers of these varieties.

Language functions as a human-driven complex adaptive system, where change emerges from interactions among speakers shaped by social, cultural, and historical factors. In this study, we believe that the observed similarities in syntactic complexity pattern among CanE, AmE, IreNE, IreSE, AusE and NZE can be attributed to their shared linguistic heritage with BrE and extensive language contact resulting from large-scale immigration movements. The interpretation of these findings draws on digital humanities approaches, integrating historical, demographic and cultural dimensions, highlighting the value of combining quantitative corpus-based methods with contextual humanistic interpretation.

However, some limitations should be considered. First, this study only examined the syntactic complexity of several Inner Circle English varieties. Future research could expand the number and scope of English varieties to gain a

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333

deeper understanding of the syntactic complexity of different English varieties worldwide. In addition, the corpus used in this study was not subdivided into dialect varieties, and only the impact of English varieties in one country on those in other countries was considered. Future research can examine which dialect varieties in one country influenced English varieties in other countries. Meanwhile, future studies will consider incorporating more up-to-date and large-scale corpora, such as GloWbE, to enhance the breadth and relevance of the data.

Acknowledgements

This paper would not have been possible without the compilers of the respective components of the International Corpus of English who made the data used in this study available to the public through academic licenses. Also, this study is part of the work conducted by the Digital Humanities and Glottometrics Lab at Fudan University.

References

Akinlotan, Mayowa, and Alex Housen. 2017. "Noun Phrase Complexity in Nigerian English: Syntactic Function and Length Outweigh Genre in Predicting Noun Phrase Complexity." *English Today* 33 (3): 31–38. https://doi.org/10.1017/S0266078416000626.

Altmann, Gabriel, and Werner Lehfeldt. 1973. *Allgemeine Sprachtypologie: Prinzipien und Messverfahren*. München: Wilhelm Fink.

Bayard, Donn. 1989. "Me Say That? No Way!: The Social Correlates of American Lexical Diffusion in New Zealand English." *Te Reo: Journal of the Linguistic Society of New Zealand* 32 (32): 17–60.

Bentz, Christian. 2018. *Adaptive Languages: An Information-Theoretic Account of Linguistic Diversity*. Berlin: De Gruyter Mouton.

Bobda, Augustin Simo. 2006. "Irish Presence in Colonial Cameroon and its Linguistic Legacy." In *The Celtic Englishes IV: The Interface Between English and the Celtic Languages*, edited by Hildegard C. L. Tristram, 217–231. Potsdam: Potsdam University Press.

Brunner, Thomas. 2014. "Structural Nativization, Typology and Complexity: Noun Phrase Structures in British, Kenyan and Singaporean English." *English Language and Linguistics* 18 (1): 23–48. https://doi.org/10.1017/S1360674313000269.

Butler, Susan. 2001. "Australian English – An Identity Crisis." In *English in Australia*, edited by David Blair and Peter Collins, 151–162. Amsterdam: John Benjamins.

Chen, Jiaxin, Dechao Li, and Kanglong Liu. 2024. "Unraveling Cognitive Constraints in Constrained Languages: A Comparative Study of Syntactic Complexity in Translated, EFL, and Native Varieties." *Language Sciences* 102: 101612. https://doi.org/10.1016/j.langsci.2024.101612.

Crystal, David. 2012. *English as a Global Language*. Cambridge: Cambridge University Press.

De Clercq, Bastien, and Alex Housen. 2017. "A Cross-Linguistic Perspective on Syntactic Complexity in L2 Development: Syntactic Elaboration and Diversity." *The Modern Language Journal* 101 (2): 315–334. https://doi.org/10.1111/modl.12396.

Elliott, Bruce S. 2004. "Regional Patterns of English Immigration and Settlement in Upper Canada". In *Canadian Migration Patterns From Britain and North America*, edited by Barbara J. Messamore, 51–90. Ottawa: University of Ottawa Press.

Filppula, Markku. 2008. "Irish English: Morphology and Syntax." In *Varieties of English: 1 The British Isles*, edited by Bernd Kortmann, Clive Upton and Edgar W. Schneider, 328–359. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110208399.2.328.

Fritz, Clemens W. A. 2007. *From English in Australia to Australian English 1788–1900*. Frankfurt am Main: Peter Lang.

Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret Maclagan, Andrea Sudbury, and Peter Trudgill. 2004. *New Zealand English: Its Origins and Evolution*. Cambridge: Cambridge University Press.

Greenbaum, Sidney. 1996. *Comparing English Worldwide: The International Corpus of English*. Oxford: Clarendon Press.

Harris, John. 1990. "More on Brogues and Creoles: What's Been Happening to English Short U?" *Irish University Review* 20 (1): 73–90.

Hickey, Raymond. 2007. *Irish English: History and Present Day Forms*. Cambridge: Cambridge University Press.

Hundt, Marianne. 2012. "Varieties of English: Australian/New Zealand English." In *Historical Linguistics of English: An International Handbook*, edited by Alexander Bergs and Laurel Brinton, 1995–2012. Berlin: De Gruyter. https://doi.org/10.1515/9783110251609.1995.

Kachru, Braj B. 1985. "Standards, Codification and Sociolinguistic Realism: The English Language in the Outer Circle." In *English in the World: Teaching and Learning the Language and Literatures*, edited by Randolph Quirk and Henry Widdowson, 11–30. Cambridge: Cambridge University Press.

Kachru, Braj B. 1992. "Introduction: The Other Side of English and the 1990s." In *The Other Tongue: English across Cultures*, edited by Braj B. Kachru, 1–15. Champaign, IL: University of Illinois Press.

Kachru, Braj B. 2005. *Asian Englishes: Beyond the Canon*. Hong Kong: Hong Kong University Press.

Kallen, Jeffrey L. 1994. "English in Ireland." In *English in Britain and Overseas: Origins and Development: The Cambridge History of the English Language, Volume V*, edited by Robert Burchfield, 148–196. Cambridge: Cambridge University Press.

Khushik, Ghulam Abbas, and Ari Huhta. 2020. "Investigating Syntactic Complexity in EFL Learners' Writing across Common European Framework of Reference Levels A1, A2, and B1." *Applied Linguistics* 41 (4): 506–532. https://doi.org/10.1093/applin/amy064.

Kinealy, Christine. 1994. *This Great Calamity: The Irish Famine, 1845–52*. Dublin: Gill and Macmillan.

Kirkpatrick, Andy. 2007. *World Englishes Paperback with Audio CD: Implications for International Communication and English Language Teaching*. Cambridge: Cambridge University Press.

Korhonen, Minna. 2017. "Perspectives on the Americanisation of Australian English: A Sociolinguistic Study of Variation." PhD diss., University of Helsinki.

Kortmann, Bernd, and Benedikt Szmrecsanyi. 2009. "World Englishes between Simplification and Complexification." In *World Englishes: Problems, Properties and Prospects: Selected Papers from the 13th IAWE Conference*, edited by Lucia Siebers and Thomas Hoffmann, 265–285. Amsterdam: John Benjamins.

Kretzchmar, William A. 2004. "Standard American English Pronunciation." In *A Handbook of Varieties of English*, edited by Bernd Kortmann and Edgar W. Schneider, 257–269. Berlin: Mouton de Gruyter.

Leitner, Gerhard. 2004. *Australia's Many Voices: Australian English – The National Language*. Berlin: Mouton de Gruyter.

Lim, Lisa, and Umberto Ansaldo. 2016. *Languages in Contact*. Cambridge: Cambridge University Press.

Little, Hannah. 2011. "The Role of Foreigner Directed Speech in the Cultural Transmission of Language and the Resulting Effects on Language Typology." PhD diss., University of Edinburgh.

Liu, Haitao. 2018. "Language as a Human-Driven Complex Adaptive System." *Physics of Life Reviews* 26–27: 149–151. https://doi.org/10.1016/j.plrev.2018.06.006.

Liu, Kanglong, and Muhammad Afzaal. 2021. "Syntactic Complexity in Translated and Non-Translated Texts: A Corpus-Based Study of Simplification." *PLoS One* 16 (6): e0253454. https://doi.org/10.1371/journal.pone.0253454.

Liu, Yi, Andrew K. F. Cheung, and Kanglong Liu. 2023. "Syntactic Complexity of Interpreted, L2 and L1 Speech: A Constrained Language Perspective." *Lingua* 286: 103509. https://doi.org/10.1016/j.lingua.2023.103509.

Lu, Xiaofei. 2010. "Automatic Analysis of Syntactic Complexity in Second Language Writing." *International Journal of Corpus Linguistics* 15 (4): 474–496. https://doi.org/10.1075/ijcl.15.4.02lu.

Lu, Xiaofei, and Haiyang Ai. 2015. "Syntactic Complexity in College-Level English Writing: Differences among Writers with Diverse L1

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333

Backgrounds." *Journal of Second Language Writing* 29: 16–27. https://doi.org/10.1016/j.jslw.2015.06.003.
Markus, Manfred. 2020. "What Did Joseph Wright Mean by Meaning." *International Journal of English Studies* 20 (1): 1–25. https://doi.org/10.6018/ijes.403481.
McWhorter, John H. 2008. *Our Magnificent Bastard Tongue: The Untold History of English*. New York: Gotham Books.
Montgomery, Michael. 2006. *From Ulster to America: The Scotch-Irish Heritage of American English*. Belfast: Ulster Historical Foundation.
Mosler, David, and Robert Catley. 1998. *America and Americans in Australia*. Westport, CT: Praeger.
Mwangi, Serah. 2004. "Prepositions Vanishing in Kenya." *English Today* 20 (1): 27–32. https://doi.org/10.1017/S0266078404001051.
Przewozny, Anne, and Cécile Viollain. 2016. "On the Representation and Evolution of Australian English and New Zealand English." *Anglophonia: French Journal of English Studies* 21 (21). https://doi.org/10.4000/anglophonia.727.
Rickford, John. 1986. "Social Contact and Linguistic Diffusion: Hiberno-English and New World Black English." *Language* 62 (2): 245–289.
Schneider, Edgar W. 2003. "The Dynamics of New Englishes: From Identity Construction to Dialect Rebirth." *Language* 79 (2): 233–281. https://dx.doi.org/10.1353/lan.2003.0136.
Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.
Schreier, Daniel, and Marianne Hundt. 2013. *English as a Contact Language*. Cambridge: Cambridge University Press.
Seargeant, Philip. 2012. "English and Linguistics Globalisation." In *English in the World: History, Diversity and Change*, edited by Philip Seargeant and Joan Swann, 178–187. Abingdon: Routledge.
Strevens, Peter. 1980. *Teaching English as an International Language: From Practice to Principle*. Oxford: Pergamon Press.
Szmrecsanyi, Benedikt. 2009. "Typological Parameters of Intralingual Variability: Grammatical Analyticity versus Syntheticity in Varieties of English." *Language Variation and Change* 21 (3): 319–353. https://doi.org/10.1017/S0954394509990123.
Szmrecsanyi, Benedikt. 2012. "Typological Profile: L1 Varieties." In *The Mouton World Atlas of Variation in English*, edited by Bernd Kortmann and Kerstin Lunkenheimer, 826–843. Berlin: De Gruyter.
Szmrecsanyi, Benedikt, and Bernd Kortmann. 2009. "The Morphosyntax of Varieties of English Worldwide: A Quantitative Perspective." *Lingua* 119 (11): 1643–1663. https://doi.org/10.1016/j.lingua.2007.09.016.
Thomason, Sarah. 2001. *Language Contact: An Introduction*. Edinburgh: Edinburgh University Press.
Trudgill, Peter. 1986. "The Role of Irish English in the Formation of Colonial Englishes." In *Perspectives on the English Language in Ireland*, edited by John Harris, D. G. Little and David Michael Singleton, 3–7. Dublin: Centre for Language and Communication Studies, Trinity College.
Trudgill, Peter. 2004. *New-Dialect Formation: The Inevitability of Colonial Englishes*. Edinburgh: Edinburgh University Press.
Vennemann, Theo. 2011. "English as a Contact Language: Typology and Comparison." *Anglia* 129 (3–4): 217–257. https://doi.org/10.1515/angl.2011.084.
Winder, Gordon M. 2000. "Trouble in the North End: The Geography of Social Violence in Saint John, 1840–1860." *Acadiensis* 29 (2): 27–57. https://journals.lib.unb.ca/index.php/Acadiensis/article/view/10782.
Woodham–Smith, Cecil. 1992. *The Great Hunger: Ireland 1845–1849*. London: Penguin.
Xing, Frank Z., Danyuan Ho, Diyana Hamzah, and Erik Cambria. 2018. "Classifying World Englishes from a Lexical Perspective: A Corpus-Based Approach." In *Computational Linguistics and Intelligent Text Processing*, edited by Alexander Gelbukh, 564–575. Lecture Notes in Computer Science. Cham: Springer.
Zhou, Chenliang, and Haitao Liu. 2023. "Lexical Diversity as a Lens into the Classification of Slavic Languages: A Quantitative Typology Perspective." *Digital Scholarship in the Humanities* 38 (3): 1359–1371. https://doi.org/10.1093/llc/fqad042.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333

Appendix

Table A1. Syntactic complexity metrics based on Lu (2010, 479)

Measure	Code	Definitions	Analysis	Value
Type 1: Length of production unit				
Mean length of sentence	MLS	the average number of words per sentence	The sentence has 20 words.	20
Mean length of T-unit	MLT	the average number of words per T-unit	T-unit consists of one main clause and any subordinate clauses or non-clausal structures attached to it. The sentence consists of one main clause and one subordinate clause, which forms 1 T-unit.	20
Mean length of clause	MLC	the average number of words per clause	The sentence has 2 clauses: 1. "Although the statement may seem obvious" (8 words). 2. "It does demonstrate the difficulties faced by any historian attempting to tackle the period" (12 words). MLC = (8 + 12)/2 = 10.	10
Type 2: Sentence complexity				
Sentence complexity ratio	C/S	the average number of clauses per sentence	The sentence has 2 clauses.	2
Type 3: Subordination				
T-unit complexity ratio	C/T	the average number of clauses per T-unit	The sentence has 1 T-unit with 2 clauses.	2
Complex T-unit ratio	CT/T	The proportion of T-units that are complex (contain at least one dependent clause).	The sentence has 1 T-unit, which is complex (contains a dependent clause).	1
Dependent clause ratio	DC/C	the ratio of dependent clauses to total clauses	The sentence has 2 clauses: 1. "Although the statement may seem obvious" (dependent clause). 2. "It does demonstrate the difficulties faced by any historian attempting to tackle the period" (independent clause). DC/C = 1/2 = 0.5.	0.5
Dependent clause per T-unit	DC/T	the average number of dependent clauses per T-unit	The sentence has 1 T-unit with 1 dependent clause.	1
Type 4: Coordination				
Coordinate phrases per clause	CP/C	the average number of coordinate phrases per clause	The sentence has 0 coordinate phrases.	0
Coordinate phrases per T-unit	CP/T	the average number of coordinate phrases per T-unit	The sentence has 0 coordinate phrases.	0
Sentence coordination ratio	T/S	the proportion of coordinate sentences (sentences joined by coordinating conjunctions like "and," "but," or "or") relative to the total number of sentences.	The sentence has 0 coordinate sentences.	0
Type 5: Particular structure				
Complex nominals per clause	CN/C	the average number of complex nominals (noun phrases with modifiers or clauses) per clause	The sentence has 2 clauses with 2 complex nominals: 1. "the difficulties faced by any historian"; 2. "any historian attempting to tackle the period". CN/C = 2/2 = 1.	1
Complex nominals per T-unit	CN/T	the average number of complex nominals per T-unit	The sentence has 1 T-unit with 2 complex nominals. CN/T = 2/1 = 2.	2
Verb phrases per T-unit	VP/T	the average number of verb phrases per T-unit	The sentence has 1 T-unit with 2 verb phrases: "may seem" and "does demonstrate". VP/T = 2/1 = 2.	2

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333

Table A2. Mean values of 14 syntactic complexity metrics

Code	BrE		CanE		AmE		IreNE		IreSE		AusE		NZE	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
MLS	25.3	9.8	22.2	5.3	22.3	5.9	22.2	5.2	22.4	5.7	23.3	5.8	21.8	5.5
MLT	24.3	8.1	20.6	4.9	20.2	5.1	20.7	4.7	20.8	5.3	21.0	5.1	19.8	4.8
MLC	11.5	2.2	12.3	2.8	12.1	3.0	12.5	2.8	12.7	3.1	12.2	2.6	12.1	2.6
C/S	2.2	0.8	1.8	0.4	1.9	0.4	1.8	0.4	1.8	0.3	1.9	0.4	1.8	0.3
C/T	2.1	0.6	1.7	0.3	1.7	0.3	1.7	0.3	1.7	0.2	1.7	0.3	1.6	0.3
CT/T	0.6	0.1	0.4	0.1	0.5	0.1	0.5	0.1	0.4	0.1	0.5	0.1	0.4	0.1
DC/C	0.5	0.1	0.4	0.1	0.4	0.1	0.4	0.1	0.4	0.1	0.4	0.1	0.4	0.1
DC/T	1.1	0.5	0.6	0.2	0.7	0.3	0.7	0.2	0.6	0.2	0.7	0.2	0.6	0.2
CP/C	0.3	0.1	0.3	0.2	0.4	0.2	0.3	0.2	0.3	0.2	0.3	0.2	0.4	0.2
CP/T	0.6	0.3	0.5	0.2	0.6	0.3	0.5	0.2	0.5	0.3	0.6	0.3	0.6	0.3
T/S	1.0	0.1	1.1	0.1	1.1	0.1	1.1	0.1	1.1	0.1	1.1	0.1	1.1	0.1
CN/C	1.5	0.5	1.6	0.5	1.5	0.6	1.6	0.5	1.7	0.6	1.6	0.5	1.5	0.5
CN/T	3.2	1.3	2.6	0.9	2.5	0.9	2.6	0.9	2.7	1.0	2.7	1.0	2.4	0.8
VP/T	2.8	0.8	2.3	0.4	2.3	0.5	2.3	0.4	2.2	0.4	2.3	0.4	2.2	0.4

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333



YIJUN LONG is a PhD student in the School of International Studies, Zhejiang University, Hangzhou, China. Her research interests focus on the digital humanities and quantitative linguistics. Email: lyjlyj2021@126.com



HAITAO LIU, PhD is a University Professor in the Humanities at Fudan University, Shanghai, China. His research interests include the history of language and linguistics, language planning, data-driven linguistics. Email: htliu@163.com

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 19 Nov 2026 at 14:57:09, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0266078426101333