



Tan Arda Gedik

Friedrich-Alexander Universität Erlangen-Nürnberg, Germany

 <https://orcid.org/0000-0003-1429-9675>

Yağmur Su Kolsal

Middle East Technical University, Turkey

 <https://orcid.org/0000-0002-2659-4447>

A Corpus-based Analysis of High School English Textbooks and English University Entrance Exams in Turkey

Abstract

This study explores the disconnect between the English textbooks studied in high schools (9th–12th grades) and the English tested on Turkish university entrance exams (2010–2019). Using corpus linguistics tools such as AntWordProfiler, TAALED, and the L2 Syntactic Complexity Analyzer (L2SCA), this paper analyzes the lexical diversity and syntactic complexity indices in the sample material. A comparison of official textbooks and complementary materials obtained from the Ministry of National Education against the official university entrance exams demonstrates that: (i) differences in lexical sophistication level can be observed between the two corpora, the lexical sophistication level of the exam corpus was higher than that of the textbook corpus, (ii) there is a statistically significant difference between the two corpora in terms of lexical diversity, the exam corpus has a significantly higher level of lexical diversity than the textbook corpus, (iii) statistically significant differences also existed between the two corpora regarding the syntactic complexity indices. The syntactic complexity level of the exam corpus was higher than that of the textbook corpus. These findings suggest that Turkish high school student taught English with official textbooks have to tackle low-frequency and more sophisticated words at a higher level of syntactic complexity when they take the nationwide exam. This, in turn, creates a negative backwash effect, distorting their approach to L2, and raising other concerns about the misalignment between the official language education materials and nationwide exams.

Keywords: corpus linguistics, lexical diversity, syntactic complexity

Textbooks and Exams

English language teaching in Turkey has been a topic for long hours of debate in many layers of the society. With this in mind, the English curriculum in Turkey has witnessed many changes over the years (Hatipoğlu, 2016). The most drastic change in the recent years has been the lowering of the grade in which students learn English, the first foreign language to be taught at schools, from 4th to 2nd. In addition, the change in educational model which experienced a shift from a eight years of elementary school and four years of high school type of division of grades to four years of primary, four years of middle and four years of high school. This has required many to adopt a different approach to language learning. The national curriculum claims that the new model accommodates these changes, and the textbooks used in Turkish English as a Foreign Language (EFL) setting have also been tweaked and enhanced over the years. The national curriculum for English language for the term of 2018–2019 by Ministry of Education also states that the new curricular model puts emphasis on the use of authentic language in an authentic context, a consideration, the importance of which Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR) emphasizes.

The main goal of the new English curriculum for secondary schools is engaging learners of English in stimulating, motivating, and enjoyable learning environments to render them independent, fluent, and effective users of the language (Milli Eğitim Bakanlığı, 2018). Rather than adopting a singular teaching methodology, the curriculum sets recurring teaching and language principles which are based on the acknowledgment of the international status of English, the components of communicative competence and the integration of four main language skills. These claims of an enhanced educational model for the textbooks is very important in an EFL context, since textbooks are among the most widely used EFL teaching materials (Allen, 2008). The marked presence of textbooks in EFL classrooms signifies the need for analyzing the content and problems associated with the success of the EFL programs (Choi, 2008).

Textbooks can be considered a route map for any English language teaching (ELT) program: not only sources of information but also a factor influencing the program's structure and destination. A wrong selection can later be a source of regret. That holds true for government-imposed books, which give little opportunity for modification (Sheldon, 1998). In a wide variety of occasions in many countries, textbooks are designed with the aim of preparing the students for standardized tests, and while this widespread tendency in EFL can be a source of criticism, textbooks need to fulfill that aim. In Turkey, textbooks are mainly

used to prepare students who are to take high stakes exams. These exams are also referred to as nationwide university entrance exams. The textbooks are provided across Turkey at the beginning of each semester, free of charge, to establish equality (Gençoğlu, 2017). Some scholars have analyzed the discrepancies and a lack of correspondence between English textbooks and high stakes university entrance exams for English in various other contexts (Underwood, 2010; Tai & Chen, 2015; Nur & Islam, 2018). Although the English textbooks used at Turkish high schools are not directly aimed at addressing the English university entrance exam, the textbooks are handed out as an aid to improve students' overall proficiency. The exam, on the other hand, is a multiple-choice proficiency exam without subsections that test productive skills such as speaking and writing.

In the light of these, to achieve academic success in Turkey, students are obligated to succeed in the nationwide exam, but are the textbooks adequately preparing the students to cope with the exams? To the best of the researchers' knowledge, no study has scrutinized the lexical and syntactic complexity of high school English textbooks and the university entrance exams from a statistical standpoint so far within the Turkish context. Hence, this study aims to analyze English high school textbooks and the complementary materials that are currently in use throughout the country and English university entrance exams that were administered in the past ten years in terms of lexical sophistication, lexical diversity, and syntactic complexity using corpus linguistics analysis tools.

In sum, the current study aims to serve as (i) a non-biased source of findings while bridging the research gap, (ii) a gateway between the exam preparation committee and the textbook writers, (iii) the voice of students who struggle with vocabulary item and syntactic differences between the textbooks and exams.

Literature Review

English Language Teaching and Testing Situation in Turkey

The situation of EFL teaching in Turkey is a troubled area. Kırkgöz (2007) mentions that with Turkey's negotiations with the EU, English saw a rise of importance (e.g., to comply with the EU regulations like CEFR leveled textbooks). Attempts at accommodating for the rising importance of English competencies include international collaborations with schools in the EU in addition to modification of textbooks according to the new model. These factors have been the primary influences on the EFL teaching situ-

ation in Turkey. Kırkgöz (2007) also mentions two phases: 1863–1997, 1997 and onwards. 1863 marks the beginning of ELT in what was back then the Ottoman Empire. The year 1997, on the other hand, was of great importance as the compulsory grade in which English was taught was lowered from 6th to 4th grade. In other words, the content of many textbooks had to be re-evaluated, and this was another significant change to the EFL teaching situation in the near past. This could be associated with the never-ending change of ELT policies which attempt to make foreign language education better and increase the level of proficiency among school-age children, and as a result, the general demographic in Turkey.

As previously mentioned, Hatipoğlu (2016) mentions that Turkey has “one important high-stakes exam, which determines whether students gain entry to prestigious colleges or tertiary institutions” (p. 2). The study done by Hatipoğlu (2016) also reveals that a big number of pre-service teachers believe that high stakes exams play a dramatically life-changing role in one’s future. Furthermore, it is revealed that due to the detrimental consequences of the negative backwash effect of unplanned high stakes exams and changes to the curriculum, many students regard English as a sum of the parts they separately learn. Hatipoğlu (2016) claims the following for the EFL teaching situation in Turkey:

The short historical overview presented in the first part of the paper reveals an unsettled and frequently changing system where, in majority of the situations, changes were not based on empirical research, educational theories, or assessment models but rather on political and practical reasons. This reveals an inadequate understanding and skewed interpretation of testing and assessment. (p. 142)

Comparing English Language Testing and Teaching Materials in Other Contexts

English language testing is a topic that cannot be overlooked. Using multiple-choice based exams has been widely accepted as a way of testing many subjects, and English is not an exception. Many countries conduct various university entrance exams that utilize multiple-choice questions. Moreover, the lack of correspondence between textbooks and university entrance exams seems to be a recurring theme among other countries. In a study done by Nur and Islam (2015) in Bangladesh, the findings highlighted a clear disconnect between the intended English assessment policy directions and the practiced pattern. The analysis of data also indicated that a backwash of such “disconnect between policy and practice substantially intercedes the overall quality of secondary English education” (p. 100).

Underwood (2010) conducted a similar study to the present one in Japan, comparing English textbooks and the Japanese university entrance exam for English. Underwood (2010) states that over the years, there has been a greater alignment between the textbooks and exams in terms of readability and lexical sophistication. Nevertheless, Underwood (2010) notes that there is still more improvement required in terms of lexical overlap between the analyzed materials.

Another different approach to the same topic was carried out by Tai and Chen (2015) in Taiwan. Their study compared English textbooks in high schools to the national university entrance exam, and the frequency of marked structures, namely relative, adverbial, and passive clauses, was attained by utilizing AntConc and Readability Test Tool. In other words, their study scrutinized the two corpora from a syntactic analysis point-of-view. They reported statistically significant results between the corpora. Although there have been many studies analyzing the relationship between syntactic complexity and L2 writing (Lu & Ai, 2015, Kyle, 2016; Kyle & Crossley, 2018), studies that scrutinize syntactic complexity levels to compare exams to textbooks have been very few (Mirshojaee & Sahragard, 2015). Nevertheless, these findings, where textbooks and exams are compared, demonstrate a lack of correlation between the above-mentioned corpora and affirm the fact that “skewed interpretation of testing and assessment” (Hatipoğlu, 2016, p. 142) is a recurring theme in other parts of the world.

Lexical Sophistication, Diversity

Read (2000) determines four different ways of identifying lexical richness: lexical density, lexical diversity, lexical sophistication, and proportion of errors. Lexical sophistication and lexical diversity are two essential terms out of those four for the present investigation as lexical density and proportion of errors are more often researched in corpora that are produced by learners. To measure lexical sophistication, researchers have calculated the total number of advanced or sophisticated words in a text (Laufer & Nation, 1995). Nevertheless, there has not been a consensus on what a sophisticated/advanced word is. Yet, overall, many seem to agree that the use of word frequency as a tool to identify whether a word is advanced or not has been the widely accepted way of approaching this issue (Bardel et al., 2012). Namely, low-frequency words and how many times those appear in a text appear to stand out as the most reliable way of approaching sophisticated words (Hyltenstam, 1988; Laufer & Nation, 1995; Read, 2000; Vermeer, 2004).

Bardel et al. (2012) approach lexical sophistication as the percentage of sophisticated or advanced words in a text, including the first one thousand

(K1), the first two thousand (K2), the first three thousand (K3), and academic word list (AWL) words in the corpora. The researchers argue that the lexical sophistication level(s) of non-native speakers (NNS) of a language can prove to be a source of knowledge when it comes to testing L2 knowledge. In other words, lexical sophistication can be employed as a way of determining whether a NNS has reached native-like proficiency in terms of vocabulary size. Their argument also extends to the vocabulary size of the teaching material employed to teach L2 since the more low-frequency words the learners are exposed to, the higher native-like proficiency they are likely to have. To measure the lexical sophistication level of a text or corpus, a procedure called lexical frequency profiling first carried out by Laufer and Nation (1995), corpus linguistics tools such as AntwordProfiler (Anthony, 2012) are utilized. AntwordProfiler enables finding the coverage of aforementioned word lists in a corpus. In recently conducted studies of Kwary et al. (2018), Du (2019), Beauchamp and Constantinou (2020), AntwordProfiler was used to analyze lexical frequency profiles.

Lexical diversity, on the other hand, refers to “the range of different words used in a text, with a greater range indicating a higher diversity” (McCarthy & Jarvis, 2010, p. 381). The researchers also argue that lexical diversity can be used to determine the “writing quality of a text, vocabulary knowledge, speaker competence, Alzheimer’s onset, hearing variation as well as socioeconomic status” (p. 381) of interlocutors in a conversation. Lexical diversity introduces two different sub-terms: type-token ratio (TTR; RootTTR and LogTTR), and the measure of textual lexical diversity (MTLD). While RootTTR and LogTTR are basically calculation of the TTR level of a text using a root and a log formula, in the case of MTLD, the text is divided into segments based on the TTR value of each segment. Each segment finishes when the TTR level reaches .72 (Toruella & Capsada, 2013) and the calculation of MTLD is done by dividing the length of the text in number of words by segments.

These two other terms are introduced because determining the lexical diversity level of a text has been problematic as lexical diversity indices may display sensitivity to the length of a text (McCarthy & Jarvis, 2010). Researchers like Biber (1989) have produced reliable analyses of corpora as they seem to have been aware of this sensitivity, however, researchers such as Ertmer et al. (2002) and Miller (1981) who have not demonstrated their awareness of this issue may have produced misleading analyses of corpora. McCarthy and Jarvis (2010), however, believe that MTLD, RootTTR, and LogTTR results are of a validating nature for analyzing a text and have corrective features and factors that help researchers yield a more reliable analysis. In this study, TAALED version 1.3.1. was used to this end. TAALED (The Tool for the Academic Analysis of Lexical Diversity) is used in calculating the lexical density of a corpus for types and tokens and eight indices of lexical diversity (Kyle, 2018). Studies of

Bulté and Roothoof (2020) and Skalicky et al. (2020) are recent examples of the use of TAALED for lexical diversity analysis.

With all of this mentioned, Crossley et al. (2011) draw on the importance of lexical proficiency explaining parts of lexical proficiency, as a cognitive construct, as exposure to lexically diverse corpora, lexical-semantic relations, and coherence of core lexical items. Thus, lexical proficiency is also a very salient indication of academic success in L2 (Daller, Van Hout, & Treffers-Daller, 2003) that is interconnected with the focus of this paper.

Given the context of the EFL teaching situation not only in Turkey but also in other countries, the following question arises: do English textbooks used in high schools and English university entrance exams correspond to each other in terms of lexical complexity? What is more important is that no matter what kind of approach the institutions follow, if the textbooks and exams do not match in terms of lexical richness (lexical sophistication and diversity in this paper's case), the students are left in a position of disadvantage where what they learn does not prepare them for the examinations. As mentioned, and demonstrated by many scholars (McCarthy & Jarvis, 2010; Crossley et al., 2011; Bardel et al., 2012), lexical richness goes hand in hand with the number of low-frequency words introduced in L2 textbooks and materials. It would be unimaginable to ignore this fact and create textbooks and exams disconnected from each other. This, in turn, would raise another important question in many readers' minds: do we test what we teach? When this is not the case, when what is not taught is being tested or vice versa, many students suffer from what is called a negative backwash effect. This, in turn, demotivates them and distorts their perception of and approach to L2, forcibly changing their notion of language from a tool of communication with which they can create and share to a distorted one on which they must (or are expected to) perform various assigned tasks to be considered proficient.

Syntactic Complexity

Syntactic complexity is one of the crucial elements in language testing and evaluation of L2 learners (Wang & Slater, 2016). To assess the syntactic complexity of a text, sentence level and word level measures have been proposed such as ratio of T-units to clauses and syntactic variety of tenses (Ellis & Yuan, 2004; Larsen-Freeman, 2006; Nelson & Van Meter, 2007; Norrby & Håkansson, 2007). This is because syntactic complexity seems to have become a vital indicator of a text's complexity and comprehensibility (Wang, 1970). Many scholars report that this complexity goes higher in more proficient L2 users (Lu, 2011; McNamara et al., 2010; Ortega, 2003). These L2 users, in correlation with their proficiency, produce syntactically lengthier pieces of texts

compared to less-proficient L2 users (Frase et al., 1999; Grant & Ginther, 2000; Ortega, 2003). A heightened use of subordination was also reported (Grant & Ginther, 2000). Therefore, it is fair to explain syntactic complexity in the lines of “measures such as length of production unit, amount of subordination or coordination, [and] range of syntactic structures” (Kim, 2014, p. 32). Park (2012) suggests that the mean length of clause and sentence as well as the number of complex nominals in clauses and T-units are of salient indicators for L2 proficiency. T-unit is one of the tiniest but most important indexes in evaluating syntactic complexity (Hunt, 1965). Wolfe-Quintero et al. (1998) in their study revealed that mean length of T-unit, dependent clauses, mean number of clauses per T-unit, and mean length of clause were the best indicators of syntactic complexity.

Mean length of clause (MLC) is the average number of words per clause. It can be referred to as a global measure of syntactic complexity. Many studies also point to a salient correspondence between MLC and proficiency levels (Cumming et al., 2005; Ortega, 2003; Wolfe-Quintero et al., 1998). In contrast to MLC, the mean length of T-unit (MLT) builds another layer of specific examination of the complexity. That is, dependent clauses might be indistinguishable in MLC, but MLT, due to its T-unit nature, specifies them. Ortega (2003) and Wolfe-Quintero et al. (1998) demonstrated that just like MLC, MLT also shows great correlation with high proficiency levels. T-units may not always be enough on their own, and another index may be required. A complex T-unit per T-unit (CT/T) is the proposed index by Casanave (1994) and Lu (2011). What makes this a complex T-unit is, this time the T-unit is expected to host an independent and a dependent clause at the same time. However, CT/T is not proven to be statistically significant in relation to language development; in other words, learners’ proficiency is not reflected through this index. Nevertheless, the studies (Casanave, 1994; Lu, 2011) done on CT/T only compared the production of L2 learners and thus their proficiency. CT/T has not been examined from the point of language testing and evaluation.

This study attempts to see whether there is a contrast between the two corpora. Complex nominals per T-unit (CN/T) is a syntactic construction that has nominal clauses, nouns with adjectives, possessives, prepositional phrases, and/or infinitives/gerunds. Despite studies not reporting a significant relationship between proficiency and CN/T numbers (Wolfe-Quintero et al., 1998; Lu, 2010), Dean (2017) demonstrates a significant connection between L2 proficiency and CN/T. Table 1 illustrates the definitions of the syntactic indices used in this study based on Lu’s (2010) article.

Table 1

Syntactic indices

	Explanation
MLC	Mean length of clauses
MLT	Mean length of T-units
CT/T	# of complex T-units per T-unit
CN/T	# of complex nominals per T-unit

Lu (2010) reported five categories of syntactic complexity measures. These were: length of production unit, amount of subordination, amount of coordination, level of phrasal complexity, and overall sentence complexity. The L2 Syntactic Complexity Analyzer (L2SCA) uses 14 indices based on Lu's (2010) categories. During this study, the following four indices were employed to examine the syntactic complexity levels: MLC and MLT identify the length of the production unit. CT/T identifies the amount of subordination and CN/T examines the degree of phrasal complexity. All these indices have been investigated to seek relations between proficiency and production. However, the current study assumes that textbooks should prepare students on all four indices and that exams should correspond to them. If the textbooks fall behind the exams in terms of syntactic complexity, this will ensure that proficiency levels of the students are not tested on the same level as the textbooks prepare them to be. Furthermore, the three categories addressed in the present study (lexical sophistication, lexical diversity, and syntactic complexity) would affect the comprehension of a text the most, especially in dealing with standardized tests. Quite clearly, comprehension and proficiency are cognitive heavy processes (Kalyuga, 2006). Thus, these indices, because they indicate complexity which affect comprehension and proficiency, may possibly indicate the relation between sentence complexity and syntactic processing of the sentences. Both corpora could be examined in relation to other ten indices as well, but to keep uniformity across the two corpora, the same set of indices were utilized, namely MLT, MLC, CT/T and CN/T.

Hence, the present study aims to examine the following research questions:

- (i) Are there statistically significant differences in terms of lexical sophistication and lexical diversity between the textbook and exam corpus?
- (ii) Are there statistically significant differences in terms of syntactic complexity between the textbook and exam corpus?

Methodology

To answer the questions above, all data were gathered online either from eba.gov.tr (for English textbooks) or from ösym.gov.tr (for English university entrance exams), ÖSYM being the Measurement, Selection and Placement Center, the sole body responsible for preparing and administering the nationwide entrance exams and the placement of students, while EBA is the online platform where students and teachers alike can access educational content, among which are textbooks. English textbooks and other complementary materials (i.e., corresponding workbooks and listening transcripts) that are currently in use from 9th through 12th grade were identified and downloaded in .pdf format. Meanwhile, English university entrance exams between the years 2010–2019 were identified and downloaded in .pdf format. In total, there were eight textbooks and ten exams. The textbooks covered each grade in high schools (9th–12th grade) and were published by the following publishing houses; (MEB) *Relearn*, *Teenwise*, *Progress* for 9th; *Count Me In*, *Gizem* for 10th; *Sunshine*, *Silverlining* for 11th; and *Count Me In* for 12th grades with their accompanying workbooks. Regardless of the publishing house of the books, the respective CEFR level for grades were as follows: A1–A2 for 9th grade, A2+–B1 for 10th grade, B1+–B2 for 11th grade and B2+ for 12th grade. The total number of tokens in the textbook corpus was 301.255. The ten exams were all prepared and released by ÖSYM between the years of 2010–2019 with a total token number of 66.913. While these books are produced by different publishing houses, they all have to follow the same regulations put forward by MEB, and their products (textbooks) have to go through a series of assessments and evaluation by a committee allocated by MEB itself.

Once the data collection was over, the followings were executed in a progressive order: (a) convert all the .pdf files into .docx files using an online document converter; (b) clean both corpora of any mistakes, typos and unnecessary signs or images which may have been caused by the conversion and may interfere with the results; (c) convert the clean .docx files into compatible .txt files for the analysis tools; (d) run both AntwordProfiler, TAALED and the L2 Syntactic Complexity Analyzer (L2SCA) on all the documents and save the results in .csv files; (e) run the .csv files' output through SPSS for statistical analysis, including descriptive analysis and a series of independent samples t-tests; (f) interpret the results.

While for lexical sophistication, AntWordProfiler (Anthony, 2012) was used to examine both corpora, for lexical diversity, Kristopher Kyle's TAALED version 1.3.1. was employed. TTR, RootTTR, LogTTR, and MTLT were selected as the indices to conduct the comparison between the two corpora. As mentioned in the literature review, because these indices have corrective

features that are required when working with longer texts, they were chosen reliable indices. As for syntactic complexity, the L2SCA (Lu, 2010) was employed to analyze MLT, MLC, CT/T and CN/T because of the following two reasons: (i) the researchers specifically wanted to focus on whether sentence and clause lengths were statistically different across corpora even though the token numbers are vastly different (thus MLT and MLC were selected), (ii) the amount of subordination, as mentioned in the literature review, would affect one's comprehension (hence, CT/T and CN/T were selected). Finding out the differences between the two would then show the researchers whether students are trained well enough for a timed examination regarding decoding syntactically heavily subordinated clauses. Another reason is that the scope of this study would need to be broader to examine all the syntactic indices at once.

Results

Lexical Sophistication and Lexical Diversity

The mean difference between the two corpora regarding the percentage of K1, K2, and AWL words were conducted with the SPSS software. For the following results, assumptions of equal variance and normality were met. Although the descriptive means results of K1 and K2 between the two corpora demonstrated means resembling each other, the means for AWL displayed a mismatch. As illustrated in Figure 1, the textbook corpus scored a higher mean in its use of K1 and K2 words (*MK1: 79.96%*, *SDK1: 1.93501*; *MK2: 6.64%*, *SDK2: .76213*) than the exam corpus (*MK1: 79.52%*, *SDK1: 1.65094*; *MK2: 6.15%*, *SDK2: .46871*). On the other hand, the exam corpus had a significantly higher coverage of academic words (*MAWL: 5.65%*, *SDAWL: 1.16101*) than the textbook corpus (*MAWL: 2.71%*, *SDAWL: 1.12163*). This finding was further proven with the following results. Independent t-tests results indicated that the corpora did have a drastically salient significance level for AWL. While K1 and K2 displayed insignificant statistical results (*K1: .556*; *K2: $p = .87$* , $p > 0.5$), AWL displayed a statistically significant result (*AWL: $p = .000 < 0.5$*). Descriptive statistics suggest that, on average, the exam corpus contained more low-frequency words than the textbook corpus as the textbook corpus demonstrated a higher usage of higher frequency words in mean (K1 and K2) and that the use of academic words was significantly low in the textbook corpus than in the exam corpus.

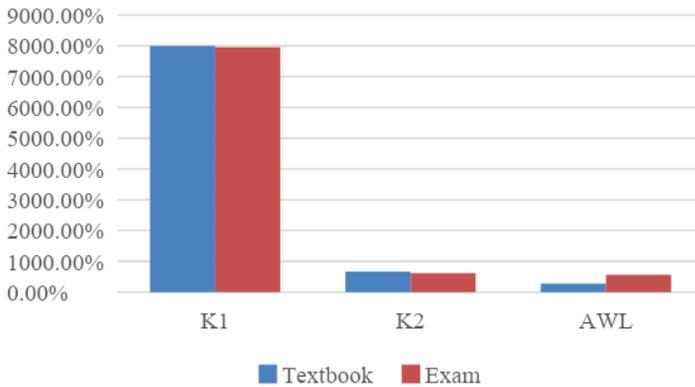


Figure 1. Lexical sophistication overlap

Unlike lexical sophistication findings, lexical diversity findings displayed greater differences in the mean between the two corpora in TTR, LogTTR, and MTL. The assumptions of equal variance and normality were met. It is evident that, regardless of TTR type, the exam corpus always scored a higher mean value (*MTTR*: .2335, *SDTTR*: .016959; *MRootTTR*: 18.096, *SDRootTTR*: 1.50964; *MLogTTR*: .8372, *SDLogTTR*: .010753; *MMTL*: 59.8613, *SDMTL*: 4.90247) than the textbook corpus (*MTTR*: .1212, *SDTTR*: .006937; *MRootTTR*: 17.1479, *SDRootTTR*: .793944; *MLogTTR*: .7864, *SDLogTTR*: .002871; *MMTL*: 55.2500, *SDMTL*: 3.97819). These numbers indicate that the exam corpus was lexically more diverse than the textbook corpus on average. The mismatch of lexical diversity was proven by independent t-tests results ($p < .05$). These results were statistically significant except for Root TTR (*TTR*: .000; *RootTTR*: .105; *LogTTR*: .000; *MTL*: .042, $p < .05$) and supported the claim that the exam corpus was lexically more diverse than the textbook corpus. Except Root TTR ($p = .105 > .05$), all other variables prove a notable variation for the corpora. Using Cohen's *d* (Cohen, 2013), the effect size of the differences between the two corpora regarding lexical diversity can be further explained. The effect sizes for the lexical diversity indices that were found are as follows; TTR: 8.6%, RootTTR: 0.78%, LogTTR: 6.45%, and MTL: 1.03%. In other words, the previously mentioned percentage indicates the amplitude of the gap of lexical diversity between the two corpora. Figure 2 shows the lexical diversity overlap.

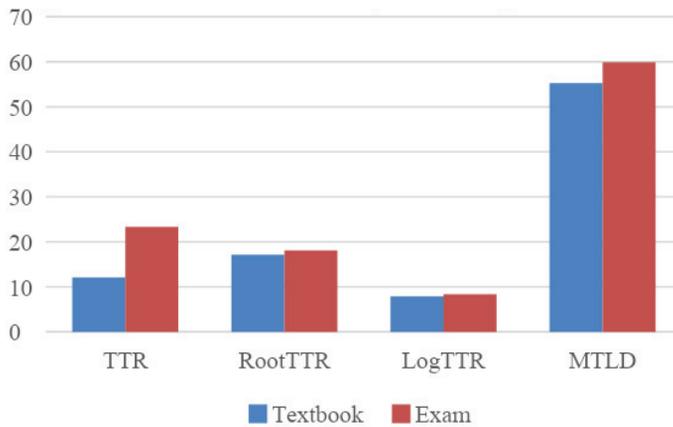


Figure 2. Lexical diversity overlap

Syntactic Complexity

Corresponding to the previous findings in the lexical section, syntactic complexity indices indicate significant differences regarding MLT, MLC, CT/T, and CN/T. The means of exams were higher (*MMLT*: 15.47, *SDMLT*: 3.39884; *MMLC*: 9.80, *SDMLC*: 1.79345; *MCT/T*: .4631, *SDCT/T*: .12678; *MCN/T*: 1.84, *SDCN/T*: .61127) than the textbooks means (*MMLT*: 10.40, *SDMLT*: 2.67762; *MMLC*: 7.97, *SDMLC*: 1.26800; *MCT/T*: .2609, *SDCT/T*: .13291; *MCN/T* 1.01, *SDCN/T*: .43015). (See Figure 3 for the differences). On the surface, it seems as if the exams were syntactically more complex than the textbook corpus. The results of the independent T-test further proved this point by displaying a significance level of ($p = .000 < 0.5$). Departing from our lexical findings, results for all four indices examined in this study performed a significance level ($p = .000 < 0.5$). These numbers suggest that the exam corpus was notably more complex than the textbook corpus regarding syntactic complexity. The implications of this finding are discussed in the next section.

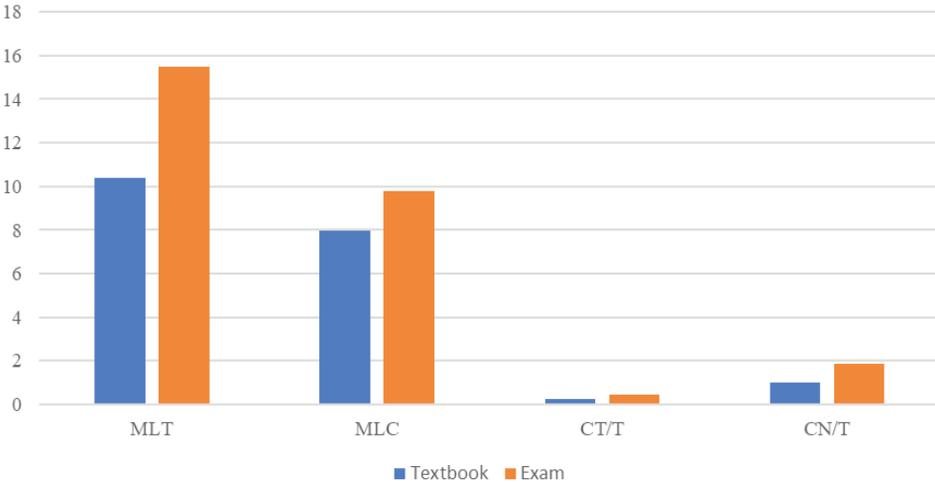
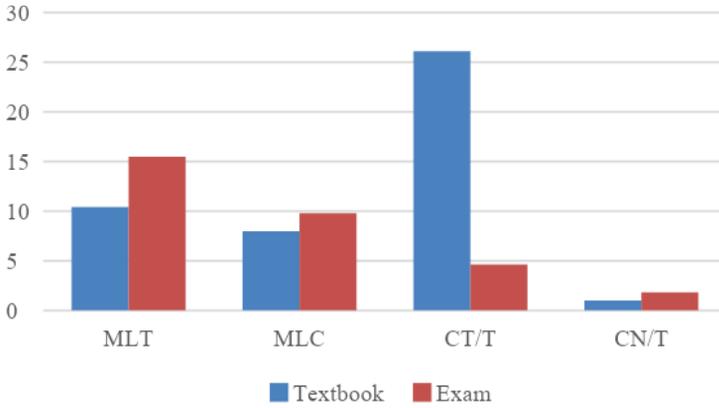


Figure 3. Syntactic complexity overlap

Discussion and Conclusion

The present research paper explored the lexical sophistication, lexical diversity, and syntactic complexity differences between the English high school textbooks and the English university entrance exams in Turkey.

Descriptive statistics suggest that lexical sophistication levels (for AWL) between the corpora demonstrate a considerable variation. Although the coverage of K1 and K2 were not significantly different between the two corpora, the coverage of the AWL was found to be significantly different. This indicates that the exam corpus contains more academic words than the textbook

corpus. Furthermore, because lexical sophistication level in AWL is lower for the textbook corpus, the learners who conduct English lessons with these textbooks are less likely to encounter low-frequency words AWL words than the AWL lexical items available in the exam corpus. This would indicate that these students would be less likely to encounter words that render them near-native-like. The exam corpus, on the other hand, proves to be lexically more sophisticated regarding AWL and contain less high-frequency AWL words in its inventory. Although K1 and K2 levels showed similar results, one should still note the slight variation between the corpora, especially when there needs to be a one-to-one correspondence between the exam and textbook materials. Frequency words also indicate that the decrease in the overlap correlates with the increase in the gap between the two corpora in terms of lexical alignment.

Results for the lexical diversity levels of the corpora tell a similar story. The differences in TTR, RootTTR, LogTTR, and MTLT among the corpora suggest that a statistically significant mismatch is present between the two corpora. More practical interpretation is averagely speaking, in every 100 words, the textbook corpus introduces ten new (different) words. This increases the lexical diversity gap between the two corpora, leading to poor input in the textbook corpus compared to the exam corpus. The statistical findings for lexical sophistication and diversity levels give the stakeholders (e.g., students, test and textbook-writers, English language teachers) a better insight and reinforce the recurring claim that the textbooks do not prepare students for the upcoming high stakes exams in terms of lexis.

The findings in lexical sophistication and diversity match with the findings of Yu's study (2018). Yu suggests that Turkish learners of English, in their academic writings, have the highest "coverage of the high-frequency words, namely the first and second 1,000 words" (Yu, 2018, p. 167). Furthermore, Yu's study, comparing Turkish speakers' written output to five other NNS groups, proves that Turkish learners of English demonstrate very poor lexical sophistication and diversity performances. These findings correspond to the current findings in this study, suggesting a cause-and-effect relationship of the materials used and tested. That is, if the materials used in classroom are more compelling regarding lexical sophistication and diversity, when they are tested in nationwide English exams, they are more likely to be acquired (see positive backwash effect, Heaton, 1989). Therefore, to improve the performance of Turkish learners of English, "vocabulary lists of academic, substitutional, and discipline-based words should be provided" (Yu, 2018, p. 168) in textbook materials.

Syntactic complexity findings are, perhaps, the most dramatic results in this study. Descriptive statistics results for syntactic complexity indices (MLC, MLT, CT/T and CN/T) always demonstrate a higher mean in the exam corpus. This means that on average, exam takers are likely to spend more time reading the

sentences (MLC). Due to higher means of MLT (and T-unit's nature which is "one main clause with all subordinate clauses attached to it" (Hunt 1965, p. 20) in the exam corpus, exam takers are more likely to be under a cognitive load to process the syntactic packaging compared to the textbook corpus. As with MLT, CT/T also significantly affects the exam takers processing times significantly as CT/Ts pack more complex T-units. Complementarily, higher means of CN/T indicates a heavier syntactic load for the exam takers, to decode the complex nominals. The difference between the two corpora was statistically significant for all indices. Namely, if students are to prepare for the high stakes exams using the government imposed books, then the chances of students' success (unless they have access to external educational materials and teachers who are aware of this mismatch, or this mismatch has been addressed by the exam and textbook preparation teams) is very low because of the mismatch between MLC, MLT, CT/T, and CN/T levels.

The pedagogical implications of this study are as follows: because there is a remarkable differentiation of lexical sophistication, lexical diversity and syntactic complexity levels, the students who have used these textbooks and taken these exams may have been forced to develop a more distorted idea of L2 (in this case, English). This distorted idea (also known as negative backwash effect) reinforces that languages can be split into smaller units and that no matter how hard they study for the English university entrance exam using government-based textbooks, they run the risk of not being able to succeed in the high-stakes English university exams. Another important point to explain is that students who use these textbooks are likely to struggle with exam fatigue due to heavy syntactic processing even from the very beginning of the exam. Moreover, this study can be beneficial for the major stakeholders of English language teaching in Turkey, namely, the textbook and exam-writers, the English language teachers, and the students. These stakeholders, with the findings at hand, can communicate and reconcile this apparent gap of lexical knowledge expected from students in the high stakes exams. The textbook and exam writers also need to work collaboratively to account for these to provide a more reliable exam experience for everyone, on equal grounds. The discussion of equal grounds can also be expanded to include the inequalities across socio-economically advantaged and disadvantaged students. Most students who come from a disadvantaged background may not have access to lexically and syntactically more compelling textbooks and may be more likely to fail in the university entrance exam while the advantaged students are ever so subtly favored and made to succeed as they already have access to more compelling language learning materials. This may not be the case for everyone in Turkey, but it might disclose an important—mostly overlooked—inequality that affects the lives of many young students who just wish to be successful but cannot figure out why they keep failing.

Although this study attempts to bridge the gap in the literature of Turkish corpus linguistics, it has several limitations. First, the study has relatively small corpora and only discovers the current situation of the corpora that are in use; Second, the study includes only four syntactic complexity indices out of fourteen. Future studies should consider these limitations and conduct a study that can utilize larger corpora and evaluate the overlap and mismatch of lexical sophistication/diversity and syntactic complexity alignment levels of the corpora.

References

- Allen, H. W. (2008). Textbook materials and foreign language teaching: Perspectives from the classroom. *The NECTFL Review*, 62, 5–28.
- Bardel, C., Gudmundson, A., & Lindqvist, C. (2012). Aspects of lexical sophistication in advanced learners' oral production: Vocabulary acquisition and use in L2 French and Italian. *Studies in Second Language Acquisition*, 34(2), 269–290. <https://doi.org/10.1017/S0272263112000058>
- Beauchamp, D., & Constantinou, F. (2020). Using corpus linguistic tools to identify instances of low linguistic accessibility in tests. *Research Matters: A Cambridge Assessment publication*, 29, 10–16.
- Biber, D. (1989). A typology of English texts. *Linguistics*, 27, 3–43.
- Bulté, B., & Roothoof, H. (2020). Investigating the interrelationship between rated L2 proficiency and linguistic complexity in L2 speech, *System*, 91, 1–16.
- Casanave, C. P. (1994). Language development in students' journals. *Journal of second language writing*, 3(3), 179–201.
- Choi, I. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, 25(1), 39–62.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
- Crossley, S. A., & Salsbury, T. (2010). Using lexical indices to predict produced and not produced words in second language learners. *The Mental Lexicon*, 5(1), 115–147.
- Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly*, 45(1), 182–193.
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10(1), 5–43.
- Daller, H., Van Hout, R., & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied linguistics*, 24(2), 197–222.
- Dean, A. C. (2017). Complex Dynamic Systems and Interlanguage Variability: Investigating Topic, Syntactic Complexity, and Accuracy in NS-NNS Written Interaction. *Working Papers in TESOL & Applied Linguistics*, 17(1), 56–97.
- Du, W. (2019). Analysis on the development of lexical complexity in Chinese science students' English writing. *Noble International Journal of Social Sciences Research*, 4(7), 116–120.
- Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition*, 26(1), 59–84.

- Ertmer, P. A., Bai, H., Dong, C., Khalil, M., Hee Park, S., & Wang, L. (2002). Online professional development: Building administrators' capacity for technology leadership. *Journal of Computing in Teacher Education*, 19(1), 5–11.
- Fletcher, P. (1985). *A child's learning of English*. Blackwell.
- Frase, L. T., Faletti, J., Ginther, A., & Grant, L. (1999). *Computer analysis of the TOEFL test of written English*. Educational Testing Service.
- Gençoğlu, C. (2017, October). *Republic of Turkey Ministry of National Education*. COMCEC, Ankara, Turkey.
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9(2), 123–145.
- Hatipoğlu, Ç. (2016). The impact of the university entrance exam on EFL education in Turkey: Pre-service English language teachers' perspective. *Procedia-Social and Behavioral Sciences*, 232, 136–144.
- Hunt, K. W. (1965). Grammatical structures written at three grade levels. *NCTE Research Report No. 3*, 2–176.
- Hyltenstam, K. (1988). Lexical characteristics of near-native second-language learners of Swedish. *Journal of Multilingual & Multicultural Development*, 9(1–2), 67–84.
- Kalyuga, S. (2006). Rapid assessment of learners' proficiency: A cognitive load approach. *Educational Psychology*, 26(6), 735–749. <https://doi.org/10.1080/01443410500342674>
- Kim, J. Y. (2014). Predicting L2 Writing Proficiency Using Linguistic Complexity Measures: A Corpus-Based Study. *English Teaching*, 69(4), 27–51.
- Kirkgoz, Y. (2007). English language teaching in Turkey: Policy changes and their implementations. *RELC Journal*, 38(2), 216–228.
- Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* [Georgia State University]. http://scholarworks.gsu.edu/alesl_diss/35/
- Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102(2), 333–349.
- Kyle, K. (2019). Measuring Lexical Richness. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 454–475). Routledge.
- Kwary, D., Artha, A., & Amalia, Y. (2018). Lexical word-class distributions in research articles of four subject areas. *Studies about Languages*, 33, 108–118.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27(4), 590–619.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics*, 16(3), 307–322.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36–62.
- Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16–27.
- Milli Eğitim Bakanlığı. (2018). Ortaöğretim İngilizce Dersi Öğretim Programı. Retrieved from: <http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=342>
- Mirshojaee, S. B., & Sahragard, R. (2015). Reading comprehension passages of Iranian general English books and MA reading comprehension tests: A corpus analysis. *Journal of Modern Research in English Language Studies*, 2(2), 77–98.
- McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior research methods*, 42(2), 381–392.

- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57–86.
- Miller, D. P. (1981). The depth/breadth trade-off in hierarchical computer menus. In *Proceedings of the Human Factors Society 25th Annual Meeting* (pp. 296–300). HFES.
- Nelson, N. W., & Van Meter, A. M. (2007). Measuring written language ability in narrative samples. *Reading & Writing Quarterly*, 23(3), 287–309.
- Norrby, C., & Håkansson, G. (2007). The interaction of complexity and grammatical processability: The case of Swedish as a foreign language. *International Review of Applied Linguistics in Language Teaching*, 45(1), 45–68.
- Nur, S., & Islam, M. (2018). The (Dis)Connection between Secondary English Education Assessment Policy and Practice: Insights from Bangladesh. *International Journal of English Language Education*, 6(1), 100–132.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied linguistics*, 24(4), 492–518.
- Park, S.-Y. (2012). A corpus-based study of syntactic complexity measures as development indices of college-level L2 learners' proficiency in writing. *Korean Journal of Applied Linguistics*, 28(3), 139–160.
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.
- Sheldon, L. E. (1998). Evaluating ELT textbooks and materials. *ELT Journal*, 42(4), 237–246.
- Skalicky, S., Duran, N., & Crossley, S. A. (2020). Please, please, just tell me: The linguistic features of humorous deception. Retrieved from: osf.io/qdjmn
- Tai, S., & Chen, H.-J. (2015). Are teachers test-oriented? A comparative corpus-based analysis of the English entrance exam and junior high school English textbooks. In F. Helm, L. Bradley, M. Guarda, & S. Thouésny (Eds.), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 518–522). Research-publishing.net. <http://dx.doi.org/10.14705/rpnet.2015.000386>
- Thomas, D. (2005). Type-Token Ratios in one teacher's classroom talk: An investigation of lexical complexity. University of Birmingham.
- Torruella, J., & Capsada, R. (2013). Lexical statistics and typological structures: A measure of lexical richness. *Procedia-Social and Behavioral Sciences*, 95, 447–454.
- Underwood, P. (2010). A comparative analysis of MEXT English reading textbooks and Japan's National Center Test. *RELC Journal*, 41(2), 165–182.
- Vermeer, A. (2004). Vocabulary size in Dutch L1 and L2 children. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 173–189). John Benjamins Publishing Company.
- Wang, M. D. (1970). The role of syntactic complexity as a determiner of comprehensibility. *Journal of Verbal Learning and Verbal Behavior*, 9(4), 398–404.
- Wang, S., & Slater, T. (2016). Syntactic complexity of EFL Chinese students' writing. *English Language and Literature Studies*, 6(1), 81–86.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency. Accuracy and complexity*. University of Hawaii Press.
- Yu, X. (2018). *Analyses and comparisons of three lexical features in native and nonnative academic English writing* [University of Central Florida]. <https://stars.library.ucf.edu/etd/6061>

Tan Arda Gedik, Yağmur Su Kolsal

Eine korpusbasierte Analyse englischer Lehrbücher für die Oberschule und englischer Hochschulaufnahmeprüfungen in der Türkei

Zusammenfassung

Die vorliegende Studie untersucht die Diskrepanz zwischen dem Inhalt von englischen Lehrbüchern, die man in den Oberschulen (9. bis 12. Klasse) verwendet, und Englischkenntnissen, die während Aufnahmeprüfungen an türkischen Universitäten (2010–2019) geprüft werden. Unter Verwendung von korpuslinguistischen Werkzeugen wie AntWordProfiler, TAALED bzw. L2 Syntactic Complexity Analyzer (L2SCA) werden anhand des Untersuchungsmaterials die lexikalische Vielfalt und syntaktische Komplexität analysiert. Aus dem Vergleich der offiziellen Lehrbücher und zusätzlichen Materialien des Ministeriums für Nationale Bildung mit den offiziellen Hochschulaufnahmeprüfungen lässt sich schließen, dass: (i) es treten Unterschiede im lexikalischen Niveau zwischen den beiden Korpora auf – das lexikalische Niveau des Prüfungskorpus war höher als das des Lehrbuchkorpus, (ii) zwischen den beiden Korpora besteht ein statistisch signifikanter Unterschied in Bezug auf die lexikalische Vielfalt – das Prüfungskorpus hat ein wesentlich höheres Niveau der lexikalischen Vielfalt als das Lehrbuchkorpus, (iii) es gibt statistisch signifikante Unterschiede zwischen den beiden Korpora hinsichtlich der syntaktischen Komplexität – das Niveau der syntaktischen Komplexität im Prüfungskorpus war höher als das im Lehrbuchkorpus. Die angeführten Schlussfolgerungen deuten darauf hin, dass türkische Oberschüler, die aus offiziellen Lehrbüchern Englisch lernen, bei landesweiten Prüfungen mit dem seltener gebrauchten und anspruchsvolleren Wortschatz auf höherem Niveau der syntaktischen Komplexität umgehen müssen. Dies wiederum führt zu einem negativen Backwash-Effekt, der ihre Einstellung zur Fremdsprache verzerrt und weitere Bedenken hinsichtlich Abweichungen zwischen den offiziellen Sprachlehrmaterialien und landesweiten Prüfungen aufkommen lässt.

Schlüsselwörter: Korpuslinguistik, lexikalische Vielfalt, syntaktische Komplexität