International Journal of Languages' Education and Teaching
Volume 5, Issue 4, December 2017, p. 287-307

Received	Reviewed	Published	Doi Number
22.11.2017	12.12.2017	25.12.2017	10.18298/ijlet.2324

Corpus Based Authenicity Analysis of Language Teaching Course Books[1]

Emrah PEKSOY [2] & Özhan HARMAOĞLU [3]

ABSTRACT

In this study, the resemblance of the language learning course books used in Turkey to authentic language spoken by native speakers is explored by using a corpus-based approach. For this, the 10-million-word spoken part of the British National Corpus was selected as the reference corpus. After that, all language learning course books used in high schools in Turkey were scanned and transferred to SketchEngine, an online corpus query tool. Lastly, certain grammar points were extracted first from British National Corpus and then from course books; similaritites and differences were compared. At the end of the study, it was found that the language learning course books have little similarity to authentic language in terms of certain grammatical items and frequency of their collocations. In this way, the points to be revised and changed were explored. In addition, this study emphasized the role of corpus approach as a material development and analysis tool; and tested the functionality of course books for writers and for Ministry of National Education.

Key Words: Authenticity, material development, evaluation, course book corpus.

1. Introduction

There can be many reasons for the language learning problems in Turkey but materials (course books) that show what to teach in which way have an important role in this respect. The course book often defines the curriculum because the Ministry of Education does not have a comprehensive curriculum (Daloglu, 2004). The teaching-learning experience mainly consists of three essential entities: the students, the teacher, and the instructional materials. As Litz (2005) stated, ELT course books (CBs from now on), being one of the most commonly recognized and used forms of instructional materials and the subject of this study, are the fundamental component ts of many ESL/EFL classrooms and programs. Thus, CBs can be accepted as a primary resource for use in the teaching-learning process.

The CB practically fulfills many useful functions. It offers structured content in a standardized design ready for application (Crewe, 2011). It gives students a record of what they are going to study or learn and teachers the role of authority in the classroom as the mediator of its content (Haycroft, 1998). In addition, in settings where target language is available only in the classroom, it is one of the main points of source and reference in and out of the class (Cunningsworth 1995). Dubin and Olshtain (1986) give CBs a more critical and important position as curriculum designers stating that "the writers themselves [becoming] the curriculum designers when their textbook is adopted" (p. 170).

[1] This study is based on writer's (Emrah Peksoy) Master's thesis finalized in 2013.
[2] Instructor, Kilis 7 Aralık University, e.peksoy@gmail.com
[3] Instructor, Kilis 7 Aralık University, ozhanharmaoglu@gmail.com

Given the situation stressed above, "coursebooks are central elements in teaching-learning encounters, not only in school settings but frequently also in tertiary-level service English contexts" (McGrath, 2006, p. 171) and should expose students to the real language (authentic) input for learning to be more effective. Therefore, quality of the instruction and the materials used are of great importance. Although approaches to language and methodology, classroom environment, motivation etc. are the most important part of language teaching and learning process, the main concern is always on the materials used, namely course books. After all, in almost all classrooms the course books are the sole determiners of tasks, activities, classroom language and even language methodology. As Sheldon (1988) indicates CBs are in the heart of any ELT Progmme whether we accept this reality or not.

However, many CBs lack showing how language really works in daily life. One of the problems with CBs is that the elements of real communication is often ignored or ill-treated by CBs (Abalı, 2006). Language used in CBs is easily noticable to almost all language learners as Gabrielatos (2002) summarizes: learners may feel confused and get down by the nature of daily real language when they encounter over-explicit messages in CBs. Thus, if course books do not make learners interact with the authentic language, it will be impossible to acquire functional and contextual features of the target language.

The use of authentic materials in foreign language learning is an issue handled for a long time. Henry Sweet, for example, addressed the use of natural vs. artificial texts in his book (Sweet, 1964, p. 177). Within the 20th century, quite a number of language learning/teaching methods, many of which are argued to be quite artificial, emerged as a growing interest in second language learning. Although they were new and modern, they all dictated using carefully structured syllabuses and demanded highly prescribed language behaviors. According to Howatt (1984), these language learning/ teaching methods were merely cult of materials because methods used in the lessons were decided by the materials themselves rather than the teachers. Almost all the course books studied included definitions of abstract grammar rules, sentences and reading extracts for translation and lists of vocabulary.

Authenticity, by its simplest term, is what is called as real, genuine, natural or related to real world. Throughout the history of English language teaching (ELT), authenticity is taken as being synonymous with "genuineness, realness, validity, reliability, and legitimacy of materials or practices" (Tatsuki, 2006, p. 1). Traditionally, authentic materials are, any sort of material which are not generated with with the aim of language teaching and learning (Nunan, 1989). Lee (1995) agrees with Nunan by stating that if a text is written for a real-life communication rather than for the purpose of teaching, then it can be accepted as authentic text. Some other researchers support those views attributing authenticity the aspect of nativeness and thus defining it as the language produced by native speakers for native speakers in any language community (Porter & Roberts, 1981; Little, Devitt & Singleton 1989, Bacon and Finnemann, 1990). Any native speaker or teachers of English can easily determine and distinguish what is simplified and intended for teaching purposes and what is 'real' and uttered in a real language setting. This definition most resembles to what comes to anyone's mind while speaking of authenticity. In his article, Authentic Materials and Authenticity in Foreign Language Learning, Alex Gilmore (2007) compiles eight different definitions about authenticity from various authors. These can be listed as;

1) The language uttered by native speakers for native speakers in a language community,

2) The language created by a real speaker or writer for a real audience in a real environment,

3) The features bestowed on a text by the receiver, in that it is not seen as something inherent in a text itself, but is imparted on it by the reader/listener,

4) A 'personal process of engagement' and social interaction between students and the teacher,

5) The different types of task chosen in language teaching settings,

6) The social context of the classroom,

7) Evaluation and assessment, and

8) Culture, and the power to act or think like a target language group in order to be recognized and accepted by them.

However, some others ascribe a more general meaning and extend the scope. According to Swaffar (1985), in order for a text to be authentic, it shouldn't be necessarily for native speakers. Any piece of language produced by a real speaker for a real audience, expressing a real message is considered to be authentic. What is important here is whether the intended communicative function is achieved or not. (Benson & Voller 1997; Morrow, 1977; Nunan, 1988/9; Porter & Roberts, 1981; Swaffar, 1985) Lee (1995) states that "a text is usually regarded as textually authentic if it is not written for teaching purposes, but for a real life communicative purpose, where the writer has a certain message to pass on to the reader" (p. 324). If communication is to have purpose and be meaningful, it necessitates the input and context to be 'real', in other words 'authentic'. As a result, in order for a learning/teaching experience to be successful, it should be related to real life situations or expose learners to genuine interactions in daily life. The context language is presented in is, thus, of utmost importance to be successful in the attempt.

Krashen's affective filter hypothesis also supports authenticity in language classrooms. Affective factors such as motivation, anxiety and self-confidence effects learners' readiness to get the language (Schulz, 1991). Krashen (1989) differentiates between two kinds of affective filter; high and low affective filter. If learners' affective filters are low, language learning is promoted. Materials which lower the affective filter are defined to be "on topics of real interest" (p. 29), which is nothing but authentic materials. In fact, the content and design of CBs are the main source of conflict in authenticity debate. A lot of ELT researchers and practitioners agree on the fact that they are on off-shores of real life (Brown and Eskenzai, 2004) and thus serve as barriers to real world. To conclude, the idea of authenticity has many reasons to be in the center of attention in teaching and learning language attempts.

How do we achieve authenticity in CBs? Corpora, particularly those of the spoken language, are an incomparable resource (Mishan, 2005). One advantage of the corpus is that it is 'pure' language and loaded with contextual features. Authenticity is the major distinguishing feature of corpus data which Sinclair (1996) describes as follows: "All the material is gathered from the genuine communications of people going about their normal business" unlike data gathered "in experimental conditions or in artificial conditions of various kinds". Mishan (2005) asserts that corpus data is obviously authentic since it consists of numerious authentic texts on a single platform. Indeed, this is one of the main reasons of using a corpus for learning purposes. In addition, competent speakers of English find corpus to be more natural than made-up texts. Since it is a database storing large collections of language, then a language corpus provides the largest single resource of authentic language available to the language learner.

Corpus linguistics has become influential in several areas of language teaching. More specifically, corpus - based analyses have been particularly relevant to EFL CBs writers and teachers. That is to say, in the past decade numerous researchers have repeatedly argued that many decisions regarding foreign language teaching have been based on solely the intuition of EFL teachers or course book writers (Barbieri and Eckhardt, 2007; Biber and Reppen, 2002; Gavioli and Aston, 2001; Sealey and Thompson, 2004;). Following the rise of corpus linguistics and its thorough analyses of exact data, these decisions are increasingly being made based on the empirically provable results and conclusions of corpus-based analyses.

Computers allow study of texts counting millions of words, thus giving birth to corpus studies. As Barbieri and Eckhardt (2007) argue that there has been almost no effort done in order to apply corpus-based findings to language education in a way that shows recent SLA principles and theories so far. However, corpus linguistic research may provide invaluable insight into foreign language teaching theories and practices. As the context of this study, the content of language course books as a teaching practice should of course benefit from corpus based researches. Lawson (in Barbieri and Eckhardt 2007, p. 322) suggests corpus linguistics can fill the gap between textbook grammar presentation and real language use.

Many scholars have compared EFL course book content of a certain language structure with the language occurring in authentic, everyday situations (Gilmore, 2004; Biber and Reppen, 2002; Barbieri and Eckhardt, 2007). However, these studies indicated that course book structures in many ways do not match up with their naturally-occurring counterparts in written and spoken discourse. Course books are often likely to "neglect important and frequent features of the language spoken by the real language users," and instead present "a patchy, confusing, and often inadequate treatment of common features of the grammar of the spoken language" (Barbieri and Eckhardt 2007, 321).

There are many course book evaluation studies in Turkey. Dülger (2016) tried to find out teacher views on ELT course books used in Turkey. He gathered data by giving participiants a checklist composed of 38 items. Şimşek & Dündar (2017) conducted a broad study by investigating 54 graduate theses evaluating EFL coursebooks in the years 2001-2013. Results showed that studies dealing with EFL coursebooks evaluations used surveys similar to Likerd scales adapted from perivious checklist. So it will not sound unrealistic to claim that no corpus-based course book analysis is conducted to look for ideal course books at least in the 54 graduate theses mentioned in the study above. As a newly flourishing field of linguistic studies, corpus studies have just begun to secure their grounds in language teaching and learning circles. Although there are some scholars conducting research using corpus in Turkey, most of them are not related to language pedagogy. For example, Efe & Özer (2015) used corpus based discourse analysis in order to investigate the vision and mission statements of universities in Turkey. Another study conducted by Ağcam & Bulut (2016) investigated if monolingual adult speakers of Turkish and bilingual adult speakers or Turkish and Arabic vary in terms of their spoken production in Turkish. For this aim, participants were presented two videos and were asked to narrate the videos they watched in a written form. Two sets of corpora gathered from the participants were then analysed by computer software. In summary, this study aims to fill this gap in literature in material evaluation studies in Turkey.

With authenticity in mind, the role of course books in language education was sought after within relevant literature. As the designer of the curriculum and the only contact of students with the language, the value of course books are unappreciated. If not chosen carefully and if content is away

from real life, they may mislead learners. Corpora can present the information on the frequency of a particular linguistic feature in real language. Therefore, corpus findings can be accepted as an ideal starting point for evaluating course book content and the order in which the content is organized.

2. Methodology

2.1. Research Design and Data Collection

This research is a descriptive study, exploiting quantitative analyses, which aims to compile information about similarities and differences between a reference corpus (British National Corpus) and a course book corpus which was created by collecting language data present in all general state high school ELT course books. By using the acquired data from the comparison of both corpora, the authenticity level of ELT course books (TEFL CC from now on) was determined compared to spoken part of British National Corpus. The reason why we decided on the spoken part of BNC was that communicative competence is the most salient feature of language education in Turkey. This study aims to find sufficient and insufficient parts of ELT course books used in Turkey related to spoken part of the BNC.

Firstly, a general representation of some specific grammar points (only tenses in this situation) in the spoken part of the BNC was determined regarding their frequency. Later, the same method was administered to the items extracted from the corpus of ELT course books used in high schools in Turkey. Finally, the frequency counts and their Log-likelihood values were compared using SketchEngine Corpus Query Tool. The likeness of ELT corpus to native speaker BNC corpus is to determine the authenticity level of the input students receive in the classroom. Identified by convenience sampling strategy, the materials to be analyzed for authenticity analysis in the study were language teaching course books, which consist of 4 books and are used in all general high schools. However, the course books also have some supplements like workbooks, handouts and listening transcripts in teacher's book. Since all these act as an input for lerarners, they were also included in the study for analysis.

2.2. Reference Corpus

As a reference corpus to compare the findings from course books and as the source of authentic discourse in real life, the 10-million-word spoken part of the British National Corpus was chosen. There were a few corpora for this purpose. However, the first and best-known corpus is perhaps the BNC, which is designed to represent as wide a range of modern British English as possible so as to "make it possible to say something about language in general" (Burnard 2002, 56). The BNC consists of approximately 100 million words of written texts (90%) and spoken transcripts (10%) in modern British English. In addition to part-of-speech (POS) information, the BNC is annotated with rich metadata (i.e. contextual information). Because of its generality, as well as the use of internationally agreed standards for its encoding, the BNC corpus is a useful resource for a very wide variety of research purposes, in fields as distinct as lexicography, artificial intelligence, speech recognition and synthesis, literary studies and, of course, linguistics.

2.3. TEFL CC Creation

The corpus creation step was a long, time consuming and tedious process. All 12 books needed to be computerized to make a smooth comparison. Therefore, all pages were scanned as pdf files one by one into computer and then were converted into text files to access them easily and to make searches fast. In addition, almost all corpus query softwares process only text files, not others. Later, the most difficult task started. The course books included fill-in-the-blank, matching, multiple choice and cloze test activities as well as dialogues, readings, sample sentences etc. Not to exclude any data that acts as an input for students, all exercises which are in sentence format were manually completed and blanks were filled. Extra spaces, extra punctuations, spelling mistakes, numbers, etc. which are not language items were deleted. They were organized accordingly into four high school levels.

After all ELT course books were arranged and processed into text files, the issue of part-of-speech-tagging emerged. As the BNC is POS tagged and complex part of speech queries can be made, the same thing had to be done to the course books. In corpus linguistics, part-of-speech tagging (POS tagging or POST) is the method of marking up a word in a text as corresponding to a particular part of speech, (i.e. noun, verb, adjective etc.) based on both its definition, as well as its context. SketchEngine's corpus creation feature enabled the researcher to upload the corpus of course book and to make it tagged using the PENN Treebank Tagset.

The other authenticity types, as Breen (1985) observes, teacher talk, discussions in the classroom, mutual interaction in target language, additional materials and activities brought into the class by the teacher were not taken into consideration. The nature of the study, being a corpus based one, necessitated high utilization of computer and related software. It would be impossible to extract data by hand from 10-million-word spoken part of the BNC and 100-thousand-word course book corpus. For data extraction from both corpora, SketchEngine Corpus Query System was employed. It is the most widely used software programme (Kilgarrif et al., 2014) since it allows self-created corpus and uploading it on its server. In addition, it also provides access to British National Corpus via its efficient and easy to use interface, as well as access to corpora of many different kinds and languages.

At the end of data extraction process, the frequency counts of each searched item must be compared with another. However, comparing raw frequencies may not always present reliable results. For this purpose, the best method for corpus comparison studies is believed to be the log-likelihood value (Rayson, 2002). Log likelihood is a test of significance just as chi-square. In this study, it is used to determine how significant the frequency counts of defined language items extracted from both corpora are. This study, being the comparison of small course book corpus with a large reference corpus, thus, exploits the log-likelihood test scores. To compute the log-likelihood value of an item, the online wizard needs its frequency in corpus, the number of running words in corpus, its frequency in the reference corpus, and the number of running words in the reference corpus, and finally cross-tabulates these.

Raw frequency count can be misleading. Comparing raw frequency counts of two different sized corpora presents illogical data as the sizes of corpora are different. For instance, a word having 50 instances in 2000-word text and 50 instances again in a 15,000-word academic essay has different ratio of use in both of them. Although the frequency is the same in each situation, its relative value within each text is different. Therefore, frequency counts must be normalized to per thousand or per million words (pmw values); or their relative frequency against the size of corpora must be computed.

This study investigates only one element of both corpora; grammatical comparison. Other language elements like speech acts, discourse features, phraseology etc. are not dealt with in this study since it will be unrealistic to analyse other language items in a single study. This study is believed to be an example of corpus use in material development and evaluation. As for the language items planned to be analyzed, there are no pre-defined criteria as none of the features in language can be considered as the sole representative of authentic language. Therefore, the researcher embraced a more inclusive approach to select the items to be analyzed. At this point, the researcher's intuition steps in. By just looking through the content pages of each book, what language points are most favored and repeated in each book can be easily seen. Being at the core of this study, ELT course books used in Turkey seem to revolve around Tenses in general.

After defining reference corpus, selecting course books and creating corpus, POS-tagging of the corpus and uploading it to a useful interface, and deciding on items to be analyzed, now it was all ready to start data extraction from both corpora and to start comparing the results. Data collection process and tools used are performed in three steps:

1. First, the spoken part of British National Corpus was analyzed regarding tenses so that a general characteristic of authentic language was determined.

2. Next, the similar analyses were conducted on TEFL CC so that course book representations of those items are determined.

3. Finally, frequency counts and Log-likelihood values extracted from both corpora were compared, and similarities and differences are shown in detail for the grammar item in question.

3. Results

Searching multi-word items and grammatical relations in a corpus is always problematic, which is not the case with single words. Almost all tensed verbs contain one or more extra words. Thus, to search the related items in a tagged corpus like BNC required thorough knowledge of POS tagging system employed by BNC, in this case CLAWS POS Tagging system. To give an example, a present progressive tense structure has a few elements such as am, is, are and adding –ing to the end of the verb. One can make Sketch Engine look for all instances of V+ing structures; however, it doesn't differentiate between adjective or reduction with a progressive aspect. The concordance item had to look for a language item like this for progressives: am, 'm, is, 's and are, 're are each followed by V+ing.

How do you tell a machine to search this? Thus, the researcher had to define a precise formula for Sketch Engine to look for the intended structure. See Appendix 1 for a small part of data extraction formula devised to search the items in BNC-Spoken and TEFL CC.

3.1. Verb Tenses in Positive Form

By using the formula, frequencies of each verb tense were extracted from both corpora. Later they were normalized to per million words since it would be impossible to make comparison with the TEFL CC. Normalized scores are computed by dividing the tense frequencies to the total number of words in the corpus and then multiplied by 1 million. It roughly means that what the frequency

would be if we had a corpus of one million words. It is especially useful while working with large corpora such as BNC. It is shown as *N-pmw* on the tables. For the tense distribution within the corpus, first positive sentences were taken into consideration. Raw frequencies and normalized scores are shown in Table 1.

By looking at the table below, a distinct superiority of simple present tense in BNC-Spoken can easily be seen at the first look. In addition, tense with simple aspects are far more common than their counterparts that require more complex structures such as perfect Continuous. Even there is not an occurrence of Future Perfect Continuous Tense in the 10-million-word corpus.

Table 1. Positive Tense Frequeny Distribution in BNC-Spoken

BNC-Spoken	Overall Frequency	
Positive Tenses	f(x)	N-pmw
Simple Present Tense	438.022	37027,66
Present Continuous Tense	45.795	3871,23
Simple Past Tense	166.769	14097,62
Past Continuous Tense	15.404	1302,16
Future Tense	50.873	4300,49
Future Continuous Tense	1.233	104,23
Future Perfect Tense	253	21,39
Future Perfect Continuous Tense	0	0,00
Present Perfect Tense	67.453	5702,06
Present Perfect Continuous Tense	2.155	182,17
Past Perfect Tense	7.425	627,66
Past Perfect Continuous Tense	295	24,94

Another thing to note is that perfect aspect in total is much more common than continuous aspect. Although perfect aspect is not present in many languages including Turkish, natural language occurrences provide us with enough support to make more use of them.

Another interesting observation to be deduced from the frequencies is that Present Continuous is used less compared to Simple Past and Present Perfect. Although most course books, reference guides, self study materials and even teaching practitioners introduce Present Continuous at the early stages and at the beginning of language education and reserve a huge amount of teaching time on it, the corpus findings prove the opposite. Present Perfect Tense and Past Tense, which are more common in BNC-Spoken but are given less importance in course books, should be presented in course materials more often.

Lastly, Future Perfect Tense and Past Perfect Continuous Tense had their places with 21, 39 and 24, 94 pmw values. Though very little compared to other tenses, they have their unique uses in the language.

Present Tense occupies the 56% of all tensed positive sentences while Present Continuous and Future Tenses have a rate 6% for each. Simple Past Tense and Present Perfect have the next largest shares with 21% and 8%. To sum up, Simple Present Tense is the most used tense in authentic language. As opposed to common belief, Present Continuous does not have the second place in distribution. Instead, Simple Past and Present Perfect are the most commonly used tenses after Present Tense.

The same procedure was followed with the TEFL CC. First the raw frequencies were extracted and then they were normalized to per million words. As the size of the course book corpus is quite small compared to BNC-Spoken, the normalized scores will be very helpful while comparing both corpora. The overall frequencies and normalized scores of positive tensed verbs are given in Table 2. A similar observation as in BNC-Spoken was made concerning Simple Present Tense within TEFL CC. It is also the most frequent tense followed by Simple Past Tense in TEFL CC. The simple aspect in all tenses is the most common feature of tensed verbs in our small corpus. In the third place comes Present Continuous Tense with the score of 4480 per million words. It is more frequent than Future Tense with 3966 pmw and Present Perfect Tense with 3208 pmw. Future Continuous, Future Perfect and Past Perfect Continuous Tense have too few occurrences in TEFL CC to be considered significant. Future Perfect Continuous Tense and Present Perfect Continuous Tense have 0 % frequency.

Table 2. Positive Tense Frequency Distribution in TEFL CC

	Overall Frequency	
TEFL CC Positive	f(x)	N-pmw
Simple Present Tense	4.804	30104,59
Present Continuous Tense	715	4480,60
Simple Past Tense	2.771	17364,66
Past Continuous Tense	224	1403,71
Future Tense	633	3966,74
Future Continuous Tense	6	37,60
Future Perfect Tense	28	175,46
Future Perfect Continuous Tense	0	0,00
Present Perfect Tense	512	3208,48
Present Perfect Continuous Tense	0	0,00
Past Perfect Tense	127	795,85
Past Perfect Continuous Tense	3	18,80

As for the percentages, Simple Present Tense is 50%, Simple Past Tense 29%, Present Continuous Tense 7%, Future Tense 6% and Present Perfect 5%. The other tenses have too little shares on over all frequencies.

To compare the results from BNC-Spoken with the data extracted from course TEFL CC, first the normalized frequencies were compared and then their log-likelihood values were computed. They present us with a clear picture of similarities and differences between two corpora regarding the positive tense usages. Table 3. shows the frequencies normalized to per million words.

Taking BNC-Spoken as our reference corpus to compare, characteristics of tense usage in authentic language is determined and compared with TEFL CC. By looking at the normalized frequencies, it is clearly seen that Simple Present Tense is underused by TEFL CC with 30104 pmw against 37027 pmw in BNC-Spoken. The 7,000 difference is a huge amount. However, a clear overuse of Present Continuous Tense is observed in TEFL CC with 4480 pmw, a feature most language course books have. The same overuse can also be observed in some other positive tenses. Simple Past Tense Past

Continuous, Tense Future Perfect Tense and Past Perfect Tense frequencies per million words are much more common than their BNC-Spoken counterparts.

Table 3. Positive Tenses Comparison in Both Corpora

Positive – BNC-Spoken vs. TEFL CC	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
Simple Present	37027,66	30104,59	218,04	-
Present Continuous	3871,23	4480,60	14,36	+
Simple Past	14097,62	17364,66	110,83	+
Past Continuous	1302,16	1403,71	1,22	+
Future Tense	4300,49	3966,74	4,19	-
Future Continuous	104,23	37,60	8,94	-
Future Perfect	21,39	175,46	66,40	+
Future Perfect Continuous	0,00	0,00	-	0
Present Perfect	5702,06	3208,48	204,68	-
Present Perfect Continuous	182,17	0,00	-	-
Past Perfect	627,66	795,85	6,53	+
Past Perfect Continuous	24,94	18,80	0,26	-

Another interesting difference is on the use of Present Perfect Tense. BNC-Spoken has much more occurrences of Present Perfect Tense with 5702 pmw while TEFL CC has 3208 pmw. It can be deduced that learners in Turkey are not exposed to enough positive Present Perfect structures in course books they use. Roughly same amount of Present Perfect Continuous Tense, Future Tense and Future Continuous Tense are also underrepresented in TEFL CC.

Comparing normalized frequencies may give us a general picture on the use of positive tense distribution. On the other hand, this distribution values may not always be reliable. Therefore, a further analysis of significance should be done. For this study, the researcher exploited log-likelihood value comparison, which is widely a used technique for corpus comparison.

Table 3 also shows the Log-likelihood values of two corpora and their expected frequencies within themselves. Last column indicates the overuse or underuse of selected item with symbols - and +. A similar observation can be made from the table. However, we can do some other sensitive observations with Log-likelihood values.

For instance, it was assumed that Past Continuous Tense was overrepresented in TEFL CC with 1403 pmw. The log-likelihood table presents the same result that it is more frequent than in BNC-Spoken, but the significance level tells the opposite. With 1,22 Log-likelihood value, Past Continuous Tense is overused in course books but it is not significant enough to take into consideration. The same situation is applicable to Past Perfect overuse in TEFL CC.

Under-usages of some tenses in TEFL CC have also different significance levels. Past Perfect Continuous and Future Tense have lower log-likelihood values and the difference is not significant.

Simple Past, Present Continuous and Future Perfect Tenses in positive are overused; on the other hand, Simple Present, Future Continuous and Present Perfect Tenses in positive are underused in TEFL CC compared to BNC-Spoken, and thus, they need to be revised.

3.2. Verb Tenses in Negative Form

Negative tensed sentences have similar distribution to their positive counterparts in BNC-Spoken. However, there are some differences to be considered. Table 4. shows the frequencies and normalized scores of negative tensed sentences.

Table 4. Negative Tenses In BNC-Spoken

BNC-Spoken	Overall Frequency	
Negative	f(x)	N-pmw
Simple Present Tense	51.71	4371,411
Present Continuous Tense	5.25	443,886
Simple Past Tense	14.37	1215,00
Past Continuous Tense	740	62,55
Future Tense	6.22	526,56
Future Continuous Tense	146	12,341
Future Perfect Tense	27	2,282
Future Perfect Continuous Tense	0	0
Present Perfect Tense	10.50	888,28
Present Perfect Continuous Tense	82	6,93
Past Perfect Tense	1.37	115,81
Past Perfect Continuous Tense	21	1,77

Normalized scores indicate that all negative sentences have the similar amount of usage like positive ones. On the other hand, percent distribution shows a clearer picture. The overall relative distribution of Simple Present Tense and Present Continuous Tense in the corpus remained unchanged in negative sentences with 56% and 6% in corpus. However, there is an observable decline on the use of Simple Past Tense. Its distribution reduced from 21% to 16% of all tensed sentences. However, Present Perfect sentences in negative form increased from 8% to 12%. Future tense negative sentences ratio increased, too. Negated sentences have a different usage, mostly in spoken registers. It is the contracted use of them. In most daily interactions contracted negatives are used. It is one of the main characteristics of authentic language. Table 5. shows the normalized frequencies of negatives and contracted negatives extracted from BNC-Spoken corpus.

Table 5. Contracted and Full Negatives in BNC Spoken

BNC-Spoken	Contracted Negative	Full Negative
Contracted Negative	N-pmw	N-pmw
Simple Present Tense	4154,15	217,25
Present Continuous Tense	26,71	417,17
Simple Past Tense	1185,67	29,33
Past Continuous Tense	56,04	6,51
Future Tense	471,61	54,95
Future Continuous Tense	11,24	1,10
Future Perfect Tense	2,11	0,17
Future Perfect Continuous Tense	0	0,00
Present Perfect Tense	796,13	92,14
Present Perfect Continuous Tense	4,56	2,37
Past Perfect Tense	104,56	133
Past Perfect Continuous Tense	1,69	1

As can be deduced from the table, contracted negative usage is far more common than using full length negative sentences. In Simple Present Tense, more than 90% of negatives are in contracted form. The same situation applies to other tensed sentences, too. In fact, the contracted negatives can be regarded as one of the main elements of spoken authentic language.

That contracted negatives are more common than others has an exception as can be observed from Table 5. Present Continuous Tense has 26 pmw instances of contracted negatives while it has 417 pmw instances of full negatives, which is not the case with other negated tenses. Although the contracted negatives are used more in almost all tenses, its number is very low in Present Continuous Tense.

These findings from BNC-Spoken on negated tense forms now can be compared with the numbers from TEFL CC. The differences between two corpora are more notable than positive tensed sentences. The total negated tenses seem to be underrepresented in TEFL CC in almost all tenses. BNC-Spoken negated tenses are twice as much as their course book equivalents. While Present Continuous Tense has 443 pmw score in BNC-Spoken, it has 150 pmw in TEFL CC, which creates a huge discrepancy between them. The situation applies to Present Perfect Tense, too. Course books contain only half of the natural language representation of Present Perfect Tense with 444 pmw value. However, in two tenses, Future Perfect and Past Perfect Tenses, negative forms seem to be equally distributed in TEFL CC. The results on the distribution of all negated tenses are shown in Table 6.

Table 6. Negative N-pmw Values in Both Corpora

Negatives all	BNC-Spoken N-pmw	TEFL CC N-pmw
Simple Present Tense	4371,41	3114,48
Present Continuous Tense	443,88	150,39
Simple Past Tense	1215,0	764,52
Past Continuous Tense	62,55	25,06
Future Tense	526,56	432,39
Future Continuous Tense	12,34	6,26
Future Perfect Tense	2,28	6,26
Future Perfect Continuous Tense	0	0
Present Perfect Tense	888,28	444,92
Present Perfect Continuous Tense	6,9317	0
Past Perfect Tense	115,81	119,06
Past Perfect Continuous Tense	1,77	0

The distribution of contracted negatives in BNC-Spoken was presented above. It can be claimed that contracted negatives are more frequent than full negatives and are characteristics of natural language. The findings from BNC-Spoken and TEFL CC are compared in Table 7.

Table 7. Negatives in BNC-Spoken and TEFL CC Compared

Negatives	TEFL CC Contracted Negative N-pmw	BNC-Spoken Contracted Negative N-pmw	TEFL CC Full Negative N-pmw	BNC-Spoken Full Negative N-pmw	Log-likelihood	Overuse/ Underuse
Simple Present Tense	2688,35	4154,15	426,12	217,25	63,40	-
Present Continuous Tense	62,66	26,71	87,73	417,17	41,31	-
Simple Past Tense	751,98	1185,67	12,53	29,33	30,38	-
Past Continuous Tense	18,79	56,04	6,26	6,51	4,60	-
Future Tense	413,59	471,61	18,79	54,95	2,83	-
Future Continuous Tense	6,26	11,24	0	1,10	0,58	-
Future Perfect Tense	6,26	2,11	0	0,17	0,73	+
Future Perfect Continuous Tense	0	0	0	0,00	-	0
Present Perfect Tense	426,12	796,13	18,79	92,14	42,85	-
Present Perfect Continuous Tense	0	4,56	0	2,37	-	-
Past Perfect Tense	119,06	104,56	0	11,24	0,01	+
Past Perfect Continuous Tense	0	1,69	0	0,08	-	-

By looking at the results, we can realize that contracted negatives are also low in number in most tenses compared to BNC-Spoken. The inadequate frequencies in TEFL CC are in Simple Present, Simple Past, Past Continuous, Future Continuous and Present Perfect tenses. Other tenses are adequately represented in TEFL CC except Present Continuous Tense. Contracted negative form in

Present Continuous Tense is marked with 62 pmw against 26 pmw in BNC-Spoken, which is notable considering the low frequency of Present Continuous contracted negative in authentic language.

Table 7. also presents the comparison of full negative forms. Again, all tenses in TEFL CC are quite low in frequency except Simple Present Tense. It overuses the full negatives with 426 pmw against 217 pmw in BNC-Spoken while Present Continuous Tense negatives, which is more marked in BNC-Spoken, are clearly underused. Six other tenses do not have any representation of full negatives at all TEFL CC.

Table 7 shows the log-likelihood values of both corpora in negative tensed forms. There are only Future Perfect and Past Perfect Tenses overused in TEFL CC; however, the significance values are quite low. Past Continuous, Future Tense, Future Continuous, Present Perfect Continuous and Past Perfect Continuous tenses are underused in TEFL CC with low significance levels. Nevertheless, Simple Present, Present Continuous, Simple past and Present Perfect tenses have high log-likelihood values after comparison. Simple Present, Present Continuous, Simple past and Present Perfect tenses in negative forms are underrepresented in TEFL CC and need to be revised. On the other hand, other tenses have different, positive or negative significance values, but they are not so high to be regarded as significant.

3.3. Verb Tenses in Total

The positives, negatives and contracted negatives of all tensed verbs are shown on separate tables for each corpus and the findings are compared in the section above. In this section, all positive and negative tensed structures are counted and compared. Table 8. shows the overall frequencies and normalized scores of all tenses in BNC-Spoken.

Table 8. Tenses in Total in BNC-Spoken

BNC-Spoken Tenses Total	Overall Frequency f(x)	N-pmw
Simple Present	489.734	41399,07
Present Continuous	51.046	4315,11
Simple Past	181.142	15312,62
Past Continuous	16.144	1364,71
Future Tense	57.102	4827,04
Future Continuous	1.379	116,57
Future Perfect	280	23,66
Future Perfect Continuous	0	0
Present Perfect	77.961	6590,33
Present Perfect Continuous	2.237	189,10
Past Perfect	8.795	743,47
Past Perfect Continuous	316	26,712

The distribution and percentages of all tensed verbs are shown in Figure 1. The distribution looks very similar to positive tensed verbs. Simple Present Tense has the highest share in pie chart below. Simple Past Tense comes next; and Present Perfect, Present Continuous and Future Tense follow them in order.

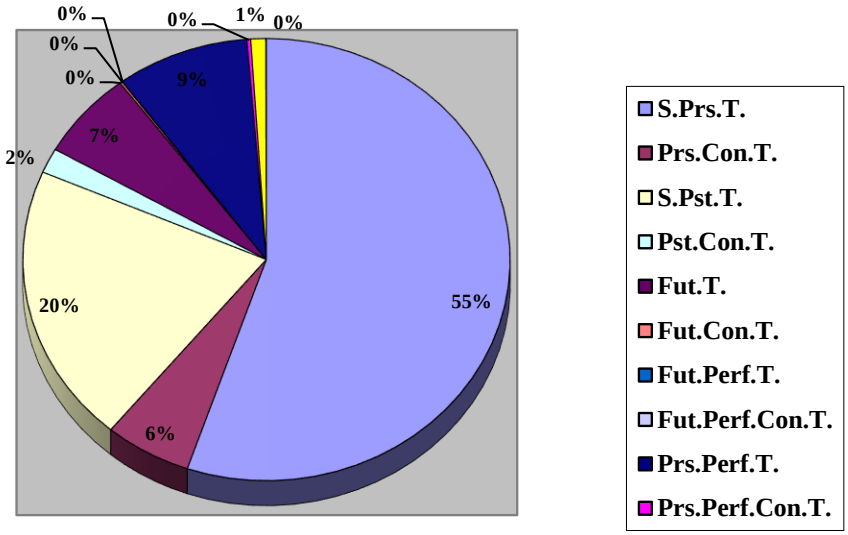

Figure 1. Distribution of all tensed sentences in BNC-Spoken

All frequencies and normalized scores are shown above regarding the distribution of verbs among verb tenses. Overall raw frequencies of verb tenses and their tense and aspect distribution are shown below in Table 9.

Table 9. Tense-Aspect Frequency Distribution in BNC-Spoken

Aspect/Tense	Present	Past	Future	Total
Simple	489.734	181.142	57.102	727.978
Progressive	51.046	16.144	1.379	68.569
Perfect	77.961	8.795	280	87.036
Perfect Continuous	2.237	316	0	2.553
	620.978	206.397	58.761	886.136

Even the raw frequencies present us with the general distribution of tensed verbs and their aspects. Present Tenses are by far the most common tenses, followed by past and future tenses. It seems that people in their daily life talk mostly with present tense forms while they get too little of future sentences compared to present ones. Normalized to per million scores are shown in Table 10. Present Tense has the score of 52,493 per million words, while past has 17,447 and future has 4967 per million words in the corpus analyzed. That means 70% of all-natural tensed verbs in authentic interactions are in present tense form.

Table 10. Tense and Aspect N-Pmw Values In BNC-Spoken

Tense	N-pmw
Present	52493,62
Past	17447,52
Future	4967,29

Aspect	N-pmw
Simple	61538,74
Progressive	5796,397
Perfect	7357,48
Perfect Continuous	215,81

The simple aspects are the most frequent ones in all tenses. Contrary to general belief, progressive aspects are not as common as their perfect counterparts. Progressives and perfect countunious aspects are the least used ones among four of them. Simple aspect has 82% share while perfect aspect is on the second rank with 10%. Progressive aspect follows them with 8% and 5796 N-pmw.

In conclusion, the raw frequencies, distribution and normalized scores of tensed verbs indicate the general features of tense usage in authentic interactions. In addition, the use of negative forms in all tenses is highlighted. The analyses present us with the general use of tenses and give us a hint on their usage in language learning materials. The overall distribution of verb tenses is compared to the findings from TEFL CC. Table 11. shows the normalized scores of both corpora on distribution of all tensed structures.

Table 11. Tenses in Total N-Pmw Values Compared

Tenses Total	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
Simple Present	41399,07	33219,07	273,38	-
Present Continuous	4315,11	4630,99	3,55	+
Simple Past	15312,62	18129,18	76,92	+
Past Continuous	1364,71	1428,77	0,47	+
Future Tense	4827,04	4399,13	6,16	-
Future Continuous	116,57	43,86	9,42	-
Future Perfect	23,66	181,73	65,60	+
Future Perfect Continuous	0	0	-	0
Present Perfect	6590,33	3653,40	246,67	-
Present Perfect Continuous	189,10	0	-	-
Past Perfect	743,47	914,91	5,79	+
Past Perfect Continuous	26,71	18,79	0,41	-

Positive, negative and contracted negatives presented important differences when taken separately. However, based on the results in Table 11, we can assume that disparity among two corpora is not much as positive and negative. Course books seem to underuse Simple Present Tense with 3321 pmw against 41399 pmw in BNC-Spoken. Future Tense, Future Continuous and Present Perfect Tense are other tense forms which are used considerably lower than authentic language. Present Perfect Continuous Tense with a high score of 189 pmw is not even represented in TEFL CC.

An interesting observation is that Future Perfect Tense with 182 pmw score is overly used in TEFL CC compared to BNC-Spoken. The score is significant because it proves the assumption that language materials are prepared based on the writers' intuition, not on research findings. Other tenses in BNC-Spoken appear to be distributed evenly in TEFL CC. Although there is not one-to-one correspondence, the differences are not significant.

Table 11. shows the log-likelihood values of all tenses in both corpora. The values are in compliance with the normalized score of all tenses. Roughly the same deductions can be made from the values, but with little differences. In normalized scores Simple Past Tense is overrepresented in TEFL CC with 18129 pmw score against 15312 pmw in BNC-Spoken. It does not seem significant at the first look, but log-likelihood score tells the opposite. With a 76,92 log-likelihood value, Simple Past Tense is over represented in TEFL CC. Past Continuous and Past Perfect Continuous tenses have too low significance results on the table. Therefore, they are not considered as determining factor in differences and similarities.

4. Discussion and Implications

This study reveals that there is a discrepancy between the target language corpus and coursebook corpus in terms of the grammatical item analyzed and acts as a sample methodology to be utilized by curriculum designers and course book writers. Unlike previous studies conducted on material evaluation field where teacher and student perceptions on course books are analyzed, this study provides an empirical research methodology for everybody involved in material development. By using corpus based analysis of coursebooks, material writers could produce more up-to-date, authentic and content-rich textbooks so that students get more chance to be exposed to real life input rather than mechanically produced language items. Language course books are developed by Turkish writers and distributed to all schools in Turkey. Since material developers are not native speakers, their intuitive knowledge regarding language can be misleading or simply be wrong. Furthermore, they can ignore the elliptical nature of language. With this method, they can revise, test and correct their material to produce more authentic course books.

According to Barbieri & Eckhardt (2007) "corpus-based analysis is an ideal tool to re-evaluate the order of presentation of linguistic features in textbooks and to make principled decisions about what to prioritize in textbook presentations" (p. 322). The main purpose of this study was, from a specific grammatical perspective, to investigate the authenticity levels of ELT course books used in all high schools in Turkey compared to the spoken part of the BNC (British National Corpus), which acts as the source of authenticity in this context. As the designer of the curriculum and the only contact of students with the language, the value of course books are unappreciated. If not chosen carefully and if content is away from real life, they may mislead learners.

The grammatical items and their normalized frequencies in BNC-spoken and their comparison with TEFL CC indicate that, on the whole, the language course books used in high schools in Turkey do not reflect authentic language use in the target language. The little similarity scores may give the material writers an insight to revise the books and to give much more place to the items which this study tried to explore. Since course books are the main elements of teaching and the main source of input for students, their closeness to real life language use is very important to get the right picture of the target language. If the topics were distributed unsystematically, students might be confused on the nature of

the language and find themselves learning irrelevant grammar topics and endless lists of vocabulary. Therefore, the grammar items explored and their weight in course books should be reconsidered in order to have more authentic course materials.

The overall results indicate that in general the TEFL CC items do not reflect the real-life usages. As the tables in previous section indicate, majority of the items has little or no representation in the whole TEFL CC. These discrepancies in question between two corpora present us with invaluable data and some suggestions for an improvement of language teaching materials can be made based on the findings. Assuming that the course book corpus and the findings it gives us indicates the kind of language education and type of English prioritised in classroom settings in Turkey, some changes and improvements regarding Tenses can be suggested in order to achieve more natural and native-like language teaching. First of all, I would suggest changing the order of Tenses introduced in course books. The Tenses are introduced in a way which differs in all 4 course books and they seem to follow a pattern like:

Simple Present, Present Continuous, Simple Past, Future, Past Continuous Present Perfect Tenses

Other tenses are sprinkled in course books with little or too much stress, and these are covered again in each course book but this time in different order. As can be seen, the order of Tenses introduced in course books does not seem to be following a logical or scientific order. However, the frequencies of each tense in TEFL CC are completely different from the order above. This suggests that course books do not even follow their own pattern of the frequencies Tenses are presented. This observation proves that course books are written just by intuition. Since there are two different orders in course books, it would be wiser to suggest BNC-Spoken order which is like:

Simple Present, Simple Past, Present Perfect, Future, Present Continuous, Past Continuous, Past Perfect, Present Perfect Continuous, Future Continuous, Past Perfect Continuous and Future Perfect

In order to eliminate possible disadvantages of giving too much/little place and importance in language teaching materials, an order and a roughly-similar number of items extracted from BNC-Spoken corpus findings will help achieve a higher degree of authenticity and help learners focus on more important and frequent items in real life situations. Thus, they will have the more oppurtunities and possibility to be prepared for real life communication situations. Therefore, the material writers should review their course books and give more emphasis in course books to the items which have high scores in BNC-spoken.

References

Abalı, E. (2006). *Analysis Of Contextual Genuıneness Of The Dıalogues In The Coursebooks Headway, Cuttıng Edge And Opportunıtıes.* M.A. Thesis, Gazi University, Ankara.

Adams, T. (1995). What Makes Matearials Authentic? *ERIC DOCUMENT 391389*. Retrieved from http://www.eric.ed.gov/PDFS/ED391389.pdf

Ağçam, R., & Bulut, A. (2016). *A Corpus-based Study on Turkish Spoken Productions of Bilingual Adults*.

Bacon, S. M., & Finnemann, M. D. (1990). A study of the attitudes, motives, and strategies of university foreign-language students and their disposition to authentic oral and written input. *Modern Language Journal, 74*(4), 459-473.

Barbieri, F., & Eckhardt, S. E. B. (2007). Applying corpus-based findings to form-focused instruction: The case of reported speech. *Language Teaching Research, 11*(3), 319-346. doi: 10.1177/1362168807077563

Benson, P., & Voller, P. (1997). *Autonomy and independence in language learning*. Harlow: Longman.

Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition, 24*(2), 199-208.

Breen, M. (1985). Authenticity in the language classroom. *Applied Linguistics, 6*(1), 60-70.

Brown, J., & Eskenzai, M. (2004, June). *Retrieval of authentic documents for reader-specific lexical practice.* Paper presented at the InSTIL/ICALL Symposium, Venice, Italy.

Burnard, L. (2002). Where did we go wrong? A retrospective look at the British National Corpus. *Language and Computers*, 42(1), 51-70.

Crewe, J. (2011). *How far do 'global' ELT coursebooks realize key principles of Communicative Language teaching (CLT) and enable effective teaching-learning?* Unpublished MA Dissertation, University of Birmingham, Birmingham.

Cunningsworth, A. (1995). *Choosing Your Coursebook*. Oxford: Heinemann.

Daloğlu, A. (2004, June). *Place of culture in English language teaching in Turkey.* Paper presented at the TESOL Convention, Long Beach, CA.

Dubin, F., & Olshtain, E. (1986). *Course design : Developing programs and materials for language learning*. Cambridge: Cambridge University Press.

Dülger, O. (2016). Evaluatıon of EFL coursebooks taught in Turkey based on teachers' views. *Journal of Advances in English Language Teaching*, *4*(1), pp-1.

Efe, I., & Ozer, O. (2015). A corpus-based discourse analysis of the vision and mission statements of universities in Turkey. *Higher Education Research & Development*, *34*(6), 1110-1122.

Gabrielatos, C. (2002). Inference: Procedures and Implications for ELT. In R. P. Millrood (Ed.), *Research methodology: Discourse in teaching a foreign language.* (pp. 30-51). Tambov, Russia: Tambov University Press.

Gavioli, L. (2001). The learner as researcher: Introducing corpus concordancing in the classroom. In G. Aston (Ed.), *Learning with Corpora* (pp. 108-137). Bologna: Athelstan.

Gilmore, A. (2004). A comparison of textbook and authentic interactions. *ELT Journal, 58*(4), 363-374. doi: 10.1093/elt/58.4.363

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching Research*(40), 97-118. doi: 10.1017/S0261444807004144

Haycroft, J. (1998). *An introduction to English language teaching*. New York: Longman.

Howatt, A. P. R. (1984). *A history of English language teaching*. Oxford: Oxford University Press.

Krashen, S. D. (1989). *Language acquisition and language education: extensions and applications*: Prentice Hall International.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., ... & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, *1*(1), 7-36.

Lee, W. Y. (1995). Authenticity revisited: Text authenticity and learner authenticity. *ELT Journal, 49*(4), 323-328.

Little, D., Devitt, S., & Singleton, D. (1989). Learning foreign languages from authentic texts theory and practice.

Litz, D. (2005). Textbook Evaluation and ELT Management: A South Korean Case Study. *Asian EFL Journal*.

McGrath, I. (2006). Teacher's and Learner's Images for Coursebooks. *ELT Journal, 60/2*(April), 171-180. doi: 10.1093/elt/cci104

Mishan, F. (2005). *Designing authenticity into language learning materials*. Bristol: Intellect.

Morrow, K. (1977). 'Authentic texts in ESP', in: S.Holden, ed. *English for specific purposes.* London: Modern English Publications, 13-15.

Nunan, D. (1988). *The learner-centred curriculum : Astudy in second language teaching*. Cambridge: Cambridge University Press.

Porter, D., & Roberts, J. (1981). Authentic listening activities. *ELT Journal, 36*(1), 37-1.

Rayson, P. (2002). *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison.* Unpublished PhD Dissertation, Lancaster University.

Schulz, R. A. (1991). Secondlanguage acquisition theories and teaching practice: How do they fit? *The Modern Language Journal, 75*(1), 17-26. doi: 10.1111/j.1540-4781.1991.tb01078.x

Sealey, A., & Thompson, P. (2004). 'What do you call the dull words?' Primary school children using corpus-based approaches to learn about language. *English in Education, 38*(1), 80-91. doi: 10.1111/j.1754-8845.2004.tb00774.x

Sheldon, L. (1988). Evaluating ELT textbooks and materials. *ELT Journal, 42/4*(October), 237-246.

Sinclair, J. (1996). The search for units of meaning. *Textus, 9*, 75-106.

Swaffar, J. K. (1985). Reading authentic texts in a foreign language: A cognitive model. *The Modern Language Journal, 69*(1), 15-34.

Sweet, H. (1964). *The practical study of languages*. Oxford: Oxford University Press.

Şimşek, M. R., & Dündar, E. (2017). Investigating EFL coursebook research in Turkey: Trends in graduate theses of the 2001-2013 period. *Educational Sciences: Theory & Practice*, *17*(3).

Tatsuki, D. (2006). What is authenticity? *The Language Teacher,* (16(5)), 17-21. Retrieved from http://jalt.org/pansig/2006/HTML/Tatsuki.htm website:

APPENDIX

Appendix 1. Sketch Engine Data Query Formula for Tenses (an extract)

For BNC-Spoken	
Simple Present Tense	[tag="VVB.*"] | [tag="VVZ.*"] | [tag="VDB.*"] | [tag="VDZ.*"] | [tag="VHB.*"] | [tag="VHZ.*"]
Simple Present Tense Negative	[tag="VDB.*"] [tag="XX0.*"] | [tag="VDZ.*"] [tag="XX0.*"]
Present Continuous Tense	([tag="VBB.*"] | [tag="VBZ.*"]) ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VDG.*"] | [tag="VHG.*"])
Present Continuous Tense Negative	([tag="VBB.*"] | [tag="VBZ.*"]) [tag="XX0.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VDG.*"] | [tag="VHG.*"])
Simple Past Tense	[tag="VVD.*"] | [tag="VDD.*"] | [tag="VHD.*"]
Simple Past Tense Negative	[tag="VDD.*"] [tag="XX0.*"]
Past Continuous Tense	[tag="VBD.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VDG.*"] | [tag="VHG.*"])
Past Continuous Tense Negative	[tag="VBD.*"] [tag="XX0.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VDG.*"] | [tag="VHG.*"])
Future Tense	[lemma="will*."]
Future Tense Negative	[lemma="will*."] [tag="XX0.*"]
Future Continuous Tense	[lemma="will*."] [tag="VBI.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VDG.*"] | [tag="VHG.*"])
Future Continuous Tense Negative	[lemma="will*."] [tag="XX0.*"] [tag="VBI.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VDG.*"] | [tag="VHG.*"])
Future Perfect Tense	[lemma="will*."] [tag="VHI.*"] ([tag="VVN.*"] | [tag="VDN.*"] | [tag="VBN.*"] | [tag="VHN.*"])
Future Perfect Tense Negative	[lemma="will*."] [tag="XX0.*"] [tag="VHI.*"] ([tag="VVN.*"] | [tag="VDN.*"] | [tag="VBN.*"] | [tag="VHN.*"])
Future Perfect Continuous Tense	[lemma="will*."] [tag="VHI.*"] [tag="VBN.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VDG.*"] | [tag="VHG.*"])
Future Perfect Continuous Tense Negative	[lemma="will*."] [tag="XX0.*"] [tag="VHI.*"] [tag="VBN.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VDG.*"] | [tag="VHG.*"])
Present Perfect Tense	([tag="VHB.*"] | [tag="VHZ.*"]) ([tag="VVN.*"] | [tag="VDN.*"] | [tag="VBN.*"] | [tag="VHN.*"]) Present Perfect
Present Perfect Tense Negative	([tag="VHB.*"] | [tag="VHZ.*"]) [tag="XX0.*"] Present Perfect Negative