Dicle University

The Institute of Educational Sciences

English Language Teaching Department

Master of Arts

THE CORPUS-BASED ANALYSIS OF AUTHENTICITY OF ELT COURSE BOOKS USED IN HIGH SCHOOLS IN TURKEY

Emrah PEKSOY

Supervisor

Asst.Prof.Dr. Süleyman BAŞARAN

Diyarbakır 2013

TAAHHÜTNAME

SOSYAL BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Dicle Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğine göre hazırlamış olduğum "**CORPUS-BASED ANALYSIS OF ELT COURSE BOOKS USED IN HIGH SCHOOLS IN TURKEY**" adlı tezin tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin kağıt ve elektronik kopyalarının Dicle Üniversitesi Eğitim Bilimleri Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım. Lisansüstü Eğitim-Öğretim yönetmeliğinin ilgili maddeleri uyarınca gereğinin yapılmasını arz ederim.

☒ Tezimin/Raporumun tamamı her yerden erişime açılabilir.

☐ Tezim/Raporum sadece Dicle Üniversitesi yerleşkelerinden erişime açılabilir.

☐ Tezimin/Raporumun … yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

20/06/2013

Emrah PEKSOY

YÖNERGEYE UYGUNLUK SAYFASI

"CORPUS-BASED ANALYSIS OF ELT COURSE BOOKS USED IN HIGH SCHOOLS IN TURKEY" adlı Yüksek Lisans tezi, Dicle Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi'ne uygun olarak hazırlanmıştır.

Tezi Hazırlayan

EMRAH PEKSOY

Danışman

YRD.DOÇ.DR.SÜLEYMAN BAŞARAN

KABUL VE ONAY

EMRAH PEKSOY tarafından hazırlanan "**CORPUS-BASED ANALYSIS OF ELT COURSE BOOKS USED IN HIGH SCHOOLS IN TURKEY**" adındaki çalışma, 18.06.2013 tarihinde yapılan savunma sınavı sonucunda jürimiz tarafından Yabancı Diller Eğitimi Anabilim Dalı, İngiliz Dili Eğitimi Bilim Dalında YÜKSEK LİSANS TEZİ olarak oybirliği ile kabul edilmiştir.

Yrd.Doç.Dr. Süleyman Başaran

Doç.Dr. Nilüfer BEKLEYEN

Doç.Dr. Behçet ORAL

Enstitü Müdürü

…./…./20..

ÖZET

TÜRKİYE'DE LİSE SEVİYESİ OKULLARDA KULLANILAN YABANCI DİL ÖĞRETİMİ KİTAPLARININ DERLEM TEMELLİ ÖZGÜNLÜK ANALİZİ

Emrah PEKSOY

Yüksek Lisans Tezi, İngiliz Dili Eğitimi Anabilim Dalı

Danışman: Yrd. Doç. Dr. Süleyman BAŞARAN

HAZİRAN 2013, 146 sayfa

Yabancı dil eğitimi ders materyalleri, hedef dilin aktif olarak konuşulmadığı ülkelerde, sınıf içi ve sınıf dışı dil etkileşiminin en temel ve yegâne kaynaklarıdır. Bunların içinden kaynak ve kılavuz olarak kullanılan ders kitapları hiç kuşkusuz ülkemizde çok yüksek miktarda kullanılmakta ve dil eğitiminin neredeyse temelini oluşturmaktadır. Ülkemizde tüm seviyelerdeki eğitim kurumlarında diğer tüm bilim alanlarında olduğu gibi dil eğitimi müfredatı ve kaynakları Milli Eğitim Bakanlığı tarafından tamamı Türk uzmanlardan oluşturulan bir komisyon tarafından hazırlanmakta ve bu materyaller ülke çapındaki tüm okullarda zorunlu olarak okutulmaktadır. Tüm liselerde yabancı dil dersi alan öğrenciler her yerde - neredeyse - aynı tür eğitime ve aynı ders kitaplarına maruz kalmaktadır. Hal böyleyken, bu kitapların içeriği, etkisi, aktiviteleri sunuş yöntemi ve hedef dili kullanmadaki yetkinliği başarılı bir dil eğitimi için büyük öneme sahiptir. Öğrencilerin hem ders içi hem ders dışı etkileşimde bulunabileceği tek kaynağın ders kitapları olduğunu düşünürsek bu kitapların önemi daha da artmıştır. Dilin yaşayan ve sürekli değişen bir yapıya sahip olduğunu da göz önünde bulundurursak, bu kitapların sürekli bir değişime, yenilenmeye ve dolayısıyla bir içerik analizine tabi tutulması etkili bir dil eğitiminin ilk basamağıdır. Bu bilgilerin ışığı altında, bu çalışmada, ülkemizde kullanılan İngilizce eğitimi ders kitaplarının

İngilizcenin ana dil olarak konuşulduğu ülkelerde kullanılan dile olan yakınlığı ve sonuç olarak iyi veya eksik yanları derlem temelli olarak gösterilmeye çalışılmıştır. Bunun için, British National Corpus (İngiliz Ulusal Derlemi)'un 10 milyon kelimeden oluşan sözlü kısmı özgün İngilizce olarak temel alınmıştır. Liselerde kullanılan İngilizce eğitimi kitaplarının hepsi yardımcı kitaplarıyla beraber bilgisayar ortamına aktarılmış ve bir çevrimiçi derlem analiz programı olan SketcEngine sistemine yüklenerek incelemeye hazır hale getirilmiştir. Son olarak, belirlenen dilbilgisi konuları önce İngiliz Ulusal Derlemi'nde daha sonra ders kitaplarında taranarak benzerlikler ve farklılıklar sayısal olarak bulunmuş ve daha sonra karşılaştırılmıştır. Araştırmanın sonunda, Türkiye'de lise seviyesinde kullanılan yabancı dil İngilizce öğretimi kitaplarının belirlenen konuların ve bu konularla birlikte kullanılan kelimelerin sıklığı açısından gerçek dile benzerliğinin çok az olduğu bulunmuştur. Bu bağlamda, ders kitaplarında geliştirilmesi veya değiştirilmesi gerekli alanlar ortaya çıkarılmıştır. Ayrıca, bu araştırma bir materyal geliştirme ve inceleme yöntemi olarak derlemin önemini vurgulamış ve ders kitabı yazarları ve Milli Eğitim Bakanlığı yetkilileri için kitapların işlevselliği değerlendirilmiştir.

Anahtar Kelimeler: Yabancı dil ders kitapları, Özgünlük, Gerçek hayata uygunluk, Derlem, Derlem Temelli İnceleme

ABSTRACT

CORPUS-BASED ANALYSIS OF ELT COURSE BOOKS USED IN HIGH SCHOOLS IN TURKEY

Emrah PEKSOY

Master of Arts, English Language Teaching Department
Supervisor: Assist. Prof. Dr. Süleyman BAŞARAN

June, 146 Pages

Foreign language teaching materials in countries where the target language is not spoken in daily life are the sole essential sources of input both inside and outside of the classroom. Of those materials, the course books used as reference and source materials, have, without doubt, been utilized frequently and they almost form the cornerstone of language teaching. In Turkey, the language teaching curriculum and course materials as well as other subjects are prepared by a commission of Turkish experts in the Ministry of National Education and are mandated for all kinds of schools throughout the country. Thus, all the language learning students in all high schools are exposed to –roughly – the same amount of teaching time and course books. Therefore, the content, impact, methodology and language competence of these course books are of vital importance for a successful language teaching. Considering the fact that the course books are nearly the only source for students to interact both in the classroom and out of the classroom, their importance is increasing more. Since a language is a living entity which continuously changes, the first step for successful language teaching is to revise, renew and analyze the course books more often. In the light of such information, 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

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; similarities 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.

Keywords: Language learning course books, Authenticity, Corpus, Corpus based analysis

ACKNOWLEDGEMENT

This study would not have been possible without the support of many people. First of all, I really wish to express my deepest gratitude to my supervisor, Asst.Prof.Dr. Süleyman BAŞARAN who was abundantly helpful and offered invaluable assistance, support and guidance. Deepest gratitude are also due to the members of the supervisory committee, Assoc.Prof.Dr. Nilüfer BEKLEYEN and Assoc.Prof.Dr. Behçet ORAL without whose knowledge and assistance this study would not have been successful.

I am also grateful to all my graduate friends, especially to İbrahim ÇAPAR, for giving me constant help and encouragement anytime I needed during this study. In addition, I want to express my deepest gratitude to all kind members of the Corpora List Mail Group, whose names I could not list here. Without their assistance, I would still be trying to find my way in the study.

Lastly, I wish to express my love and gratitude to my beloved wife, Ayşe Nur, and to my little daughter Nil, for their understanding and endless love. Without their support, I couldn't manage to finish this study. The guidance and support received from all the members who contributed to this project and whose names I forgot to mention here, was vital for the success of the project. I am grateful for their constant support and help.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

BNC: British National Corpus

ELT: English Language Teaching

SLA: Second Language Acquisition

MI-Score: Mutual Information Score

N-pmw: The Sum of Items As Per Million Words

TEFL CC: Turkish English as Foreign Language Coursebook Corpus

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1. Statement of the Problem

It is an indisputable fact that today English language is the lingua franca of the world (Wardhaugh, 1986; Graddol, 1997; House, 2003; Seidlhofer, 2005) for every possible way of communication. It is the leading language of business, policy, technology, science, the Internet, and even TV. Students and academics who can communicate in English can easily get access to information in their fields, and researchers share their findings in English. Without English, it becomes rather difficult to travel to another country and communicate with individuals. The same situation applies to Turkish context too. English in Turkey is used as the international language of access. At every level of education (primary, secondary and higher) English is taught as a compulsory course for various hours. Students in primary schools have 76 hours and students in high schools have 152 hours of average English teaching time. As for the universities, university students have compulsory foreign language courses of 2-4 hours per week in two semesters, with a total of 64-128 hours per year.

Many educational institutions have adopted English as a medium of education as it has become a prerequisite for communication at schools. There are a number of private English medium instruction schools in Turkey, and out of the country's a hundred and seventy-two universities, a few of them offer full English medium instruction: Bogazici University in Istanbul and the Middle East Technical University in Ankara, both of which are esteemed schools.

English language is learned by many for more practical usages such as finding a job (Ellis, 1997; Crystal, 1997). According to Konig (1990) most Turks study English for the international job opportunities it opens as well as the bit of social prestige it brings (p. 163). English is now one of the job requirements for upper level, better paid jobs in Turkey. Therefore, at every educational level English language skills are taught

to make students prepare for the work life after graduation (Ayman, 1995). Dogancay-Aktuna (1998) analyzed job advertisements that showed up in two most selling newspapers in Turkey and found that English is hunted as a top job requirement. In addition, 20 percent of the framed advertisements were printed in English in Turkish-medium dailies.

Dogancay-Aktuna (1998, p. 37) summed up the current situation of English in Turkey as follows:

In Turkey English carries the instrumental function of being the most studied foreign language and the most popular medium of education after Turkish.

As can be seen, though all students in Turkey are said to be studying English as a curricular requirement, research indicates that ELT in Turkey is a problematic area, if not a failure in general (Çakır, 2007; Öztürk & Tılfarlıoğlu, 2007). There is substantial difference in the quality and quantity of instruction students receive across various types of schools (Doğançay-Aktuna, 2005). According to a study conducted by Cetinkaya in 2005, even students who had studied English in the education system since early elementary school were not proficient in the language. She claims that this is because of a teaching approach that focuses too much on accuracy and linguistic structure and not enough on practical communication skills. Another problem with English in Turkey is that even competent students are not reaching a level in which they can converse in English with confidence. (Zok, 2010)

There can be many reasons for the language learning problems in Turkey but materials (course books) that show what to teach in which way are particularly important in Turkey, as 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. ELT course books (CBs), being one of the most commonly recognized and used forms of instructional materials and the subject of this study, are " … the essential constituents to many ESL/EFL classrooms and programs …" (Litz 2005, p. 5). Thus, CBs can be accepted as a primary resource for use in the teaching-learning process.

The CB practically fulfills a number of 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) gives 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).

The significant role that CBs play therefore makes them the focus of attention in that they display the theoretical and practical ideas adopted in a particular setting. By analyzing CBs, "Beliefs on the nature of learning can … be inferred" (Nunan 1991, p. 210). In addition, CBs represent the methodological beliefs of its writer/s (Harmer, 2001).

Given the situation mentioned above, "Course books are a central element in teaching-learning encounters, not only in school settings but frequently also in tertiary-level service English contexts" (McGrath, 2006) and should expose students to the real language (authentic) input for learning to be more effective.

As a result, 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, that is course books. After all, in almost all classrooms the course books are the sole determiners of tasks, activities, classroom language and even language methodology. "[W]hether we like it or not, these represent for both students and teachers the visible heart of any ELT programme" (Sheldon,1988).

However, many CBs lack showing how language really works in daily life. One of the problems with course books is that the elements of real communication is often ignored by course books or ill-treated (Abalı, 2006). Language used in course books is easily noticeable to almost all language learners as Gabrielatos (2002) summarizes, "if

learners expect over-explicit messages, they may be confused and discouraged by the elliptical nature of everyday language" (p. 46). Natural language is noticeably different from a course book language. Naturally occurring language carries both certain grammatical/structural features of spoken language and social roles of the participants (Thanasoulas, 2005). 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.

It is controversial for teachers, too; many would agree with Swales (1980) that "textbooks, especially course books, represent a 'problem', and in extreme cases are examples of educational failure." ELT course books are solutions to some of the problems in class, but are frequently seen by teachers as 'necessary evils' (Sheldon, 1988). ELT books are commonly viewed as poor bridge between what is educationally wanted on the one hand and financially affordable on the other. In simple terms, they often do not seem to provide good value for use in class. Brumfit (1980, p. 30) is rather harsh when criticizing CBs saying that "masses of rubbish is skillfully marketed".

In fact, CBs seem to be in a vicious circle, wherein CBs produced imitate their predecessors and do not adopt changes from research findings (Sheldon, 1988). Although different kinds of approaches and methods propose different teaching and learning situations, the starting point should be on the high quality authentic materials for learners.

The use of authentic materials in foreign language learning is an issue handled for a long time. Henry Sweet, for example, who is one of the first linguists in history, addressed the use of natural vs. artificial texts in his book 'The Practical Study of Languages' (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 began to flourish. Although they were new and modern, they all dictated using carefully structured syllabuses and demanded highly prescribed language behaviors. As Howatt (1984) puts it, they were just "cult of materials" and "the

authority of the approach resided in the materials themselves, not in the lessons given by the teacher using them" (p. 267). Almost all the course books studied included definitions of abstract grammar rules, sentences and reading extracts for translation and lists of vocabulary. Students got in vigorous efforts of translating sentences like:

The cat of my aunt is more treacherous than the dog of your uncle.

My sons have bought the mirrors of the Duke.

The horse of the father was kind. (Sweet, 1964 p.72)

All these examples are made-up and represent hypothetical situations, which are believed to make almost no contribution to acquisition by many researchers (Gajic, 2010; Sheldon, 1988; Swales, 1980). In 1950s, this out-of-date notion and methodology began to be questioned and rejected by teachers, students and linguists. Memorizing vocabulary, translation, conjugations made students hate language classes or remember language learning as difficult and impossible to cope with and the experience as dreary. Structural theories of language lacked to represent and account for the essential characteristics of language uniqueness and creativity (Chomsky, 1957).

According to Howatt (1984) "situational approach ….had run its course" (p. 280) and it was useless to make generalizations based on hypothetical situations. Considering these negative ideas and assumptions on language teaching methodology, linguists foresaw the need for a more modern and natural approach applicable to current needs of students. Any language methodology emphasizing communicative function of the language found much more advocates and had firm philosophical bases, which structural methods lacked. After all, "...a means of communication can only be learned by using it for this purpose" (Mishan, 2005, p. 2) and therefore, communication based methods were (are) considered to be the most natural ways of learning languages.

The most important function of the language is, without doubt, communication between individuals. This function is not covered in traditional approaches to language teaching. Chomsky's (1965) distinction between competence and performance in Aspects of Theory of Syntax started the era of Communicative methodologies. Communicative competence involved not only the knowledge of language itself but also

the knowledge of culture and successful communication of individuals, that is, expected outcome of interaction. The language was only acquired with an attempt to communicate via language. Texts were used not for their linguistic forms but for their content and meaning. Until the beginning of new millennium communication based ELT course books were indispensable in language classrooms. These ideas resulted in the approach which dominates EFL circles even today – Communicative Language Teaching – and opened the way for introduction of authentic texts, which is of utmost importance to consider and discuss in CLT.

1.2. Purpose of the Study

In this study, the authenticity of ELT course books used in high schools in Turkey is analyzed based on the spoken section of the British National Corpus. As expressed in detail in the section above, the course books are the main element of learning and teaching activity. Therefore, their linguistic, methodological and lexical content should provide learners with the necessary skills and input for interaction in daily life. That is, learners should be exposed to real language (authentic) input which is most likely to be encountered while using the target language. By analyzing the spoken part of one of the largest English corpora in the world, the main features of real language is determined and how and how much of these features are represented in target course books is shown in detail. This way, underused and overused structures and vocabulary in course books are determined. Similarities and differences between the two corpora are revealed and one of the possible reasons - if any - of not being able to communicate effectively in English is highlighted.

1.3. Research Questions

The study was conducted during a period of one year to find answers to the following research questions:

1. What is the degree of authenticity of language course books used in Turkey compared to spoken part of BNC?
2. How much do specific grammar points in course books resemble to the grammar of BNC spoken corpus?

3. How much do vocabulary choice in course books resemble to the vocabulary of BNC spoken corpus?

4. Are Turkish students of language exposed to the input they may really need in communication situation?

1.4. Significance of the study

This study presents a detailed quantitative analysis and comparison of ELT course books used in Turkey with spoken part of the BNC. There have been a number of researchers which take course books in scope and many of which analyzed the course book authenticity from teachers' or students' perspective. However, none of them looked deeper into the textual quality of them. Unlike many other course book analyses, this study questions the content value of course books and their relative resemblance to authentic English. This study is not restricted to a certain school, class or number of students. It investigates all course books used in general state high schools which are the building blocks of secondary education in general in Turkey. Thus, it allows us to make solid assumptions based on the findings and shows us a clear picture of the problem situation.

In addition, this study will act as a critique of language learning syllabus used in Turkey and as a suggestion to Ministry of National Education to review, revise and do necessary changes in the course books and syllabus since the findings will reveal a clear picture of what kind of language learners take as input.

Lastly, the study tries to show one of the possible reasons of language learning/teaching problems in Turkey in course book level despite many years and hours of language instruction. It is based on the assumption that the input learners receive from ELT course books is not authentic and enough to develop communicative competence.

1.5. Limitations of the study

One of the main limitations of this study is that since it is a corpus based analysis, authenticity type considered is just text authenticity. Other types of

authenticity like learner authenticity, task authenticity are disregarded because they are the subject of a long term study and require different analyses methods.

Another drawback of the study is that it is only limited to the high school level language course books. Due to limitation of time, all school level course books are not included in the study. In addition, high school course books represent course book language more than other levels of course books in that they provide richer context and longer conversations and more activities.

Some schools, private colleges and individual teachers use different course books and provide extracurricular activities and texts in the classroom. In addition, students eager to learn English provide more individual effort after school. They deal with language input in their free time, too. Therefore, the authenticity analysis is restricted only to the input provided by English language course books. Classroom environments, student-teacher interaction, tasks undertaken are other sources of authenticity and they are not analyzed in this study.

1.6. Definition of Key Terms

The following terms in the study are used in the meanings suggested below:

Authenticity: The language produced by native speakers for native speakers in a particular language community.

Corpus/Corpora: Large collections of written/spoken text - produced by native speakers - which are collected in computers.

BNC: British National Corpus

Concordancer: The computer software that helps process, analyze and compare corpus texts. (SketchEngine in this case)

WEB Concordancer: The internet applications that help process analyze and compare corpus texts (SketchEngine)

Turkish English as Foreign Language Course book Corpus (TEFL CC): The corpus created by collecting language materials presented by ELT course books used in Turkey.

CHAPTER 2

LITERATURE REVIEW

In this chapter, a summary of previous studies related to this study and theoretical background of the topic in literature were presented.

2.1. Authentic Materials

There is considerable amount of meanings linked with authenticity, and thus it is still not certain what authenticity really means or in what meanings it should be adopted by teachers. Even a little search on the literature, as presented below, will create an ambiguity on teachers' minds. 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).

Traditionally, authentic materials are, "any material which has not been specifically produced for the purposes of language teaching" (Nunan, 1989, as cited in Adams, 1995, p.4). Lee agrees with Nunan by stating that, "a text is usually regarded as textually authentic if it is not written for teaching purposes, but for a real-life communicative purpose. . ." (1995, p. 324). 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 actually most resembles to what comes to anyone's mind while speaking of authenticity.

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. (Morrow, 1977; Porter & Roberts, 1981; Swaffar, 1985; Nunan, 1988/9; Benson & Voller 1997). 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).

On the other hand, researchers like Widdowson (1983) and Breen (1985) add human factor into the debate. According to them, authenticity cannot be achieved in any kind of text even if it was uttered by and for native speakers. Authenticity isn't something that is present in the text. In fact, it is the interaction between the user and the text. Widdowson (1978; c.f. 1998) refers to texts designed for proficient speakers as possessing "genuineness" – a characteristic of the text or the material itself – and he asserts that this is different from "authenticity". Accordingly, the claim here is that language content itself can truly be inherently "genuine" but that authenticity itself is a social unit. In other words, authenticity is achieved through the communication of users, situations and the texts.

"Genuineness is a characteristic of the passage itself and is an absolute quality. Authenticity is a characteristic of the relationship between the passage and the reader and it has to do with appropriate response" (Widdowson, 1978). Likewise, Breen (1985) suggests that authenticity is not an entity happening within the text itself only, but it is also present in the tasks students are engaging on and in the social setting in the classroom. This suggests that authenticity can only be achieved when there is agreement between the material writer's intention and the learner's interpretation (Lee, 1995).

Van Lier (1996) makes a similar definition like Widdowson and Breen. However, he calls the effect of teacher and student interaction into action. Authenticity is achieved with the interaction between students and teachers and is a "personal process of engagement (p.128)." It cannot be attained if needs and expectations of students and teachers are different.

Authenticity lies not only in the 'genuineness' of text, but also in the activities and tasks done for communication purposes in the classroom. The input and its quality are important for language proficiency but it is not enough. Learner production is considered to be another important stage of language development. And it can be achieved with well-designed and carefully planned tasks giving learners opportunities for production. Therefore, authenticity can be said to be the result of the types of tasks chosen (Benson & Voller 1997; Bachman, 1991; van Lier, 1996; Lewkowicz, 2000) because it is the tasks that will create a real life situation in the classroom.

In addition, according to Rings (1986) text authenticity is interdependent on two other authenticity types; the situation and the speaker. Only then the language "content and structure will be authentic for that text type (p. 205). Taylor (1994) supports Rings by stating that "authenticity... is a feature of a text in a particular context. Therefore, a text can only be truly authentic in the context for which it was originally written." To sum up, authenticity cannot be achieved if the text/speech is separated from content.

As can be deduced from various definitions mentioned above, authenticity can mean anything based on your stand point. As Widdowson (1983) puts, it is "… a term which creates confusion because of a basic ambiguity" (p. 30). In this situation, we can refer to Breen's classification of authenticity to get a clearer picture of related literature:

1. Authenticity of the texts which we may use as input data for our students.

2. Authenticity of the learner's own interpretations of such texts.

3. Authenticity of tasks conducive to language learning.

4. Authenticity of the actual social situation of the language classroom. (Breen, 1985, p. 60)

All those different researchers with all their own definitions of authenticity share a common ground: authentic material is not simplified for foreign language/second language learning purposes.

2.2. Theoretical Background of Authenticity

After the communicative approach to language teaching and learning has gained ground for a few decades, the notion that language teaching and learning materials should be 'authentic' is discussed more than ever (Chaves, 1998; Hedge, 2000; Nunan, 1988:99; Harmer, 2001:205, Mishan, 2005). According to Herron and Seay (1991), the notion was that "live texts" achieve learning more than their "pedagogically contrived counterparts" (p. 488).

On the other hand, Mishan (2005) believes that humanistic and material focused approaches have also authenticity on their agenda, as well as communicative approaches. She states that:

> *Sifting through the history books reveals many precedents for authenticity in language learning, and these can be seen to fall in three groups: 'communicative approaches' in which communication is both the objective of language learning and the means through which the language is taught, 'materials-focused' approaches, in which learning is centered principally round the text, and 'humanistic approaches' which address the 'whole' learner and emphasize the value of individual development (p.1).*

Communicative Language Teaching is "the teaching of communication via language, not the teaching of language via communication" (Allwright, 1979, p. 167). 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.

The rationale of using authentic materials is voiced by Blaz (2002), stating that the "national standards for foreign language education center around five goals: Communication, Cultures, Connections, Comparisons, and Communities", that is, the national standards of the target language country (p.1). These five goals can be clearly linked to usage of authentic materials. Hadley (2001) has the same idea with different words. He asserts that communicative language teaching movement along with

other approaches "…emphasizes this need for contextualization and authenticity" (p. 140).

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.

One of the earliest researchers on the notion of authenticity, Widdowson (1996) reasons on the use of authenticity stating that "if real communicative behavior is what learners have eventually to learn, then that is what they have to be taught" (p.67). If the aim of a language teacher or program is to get students to encounter with English in real world (Widdowson, 1979; Rivas, 1999), so it will be common sense to prepare the classroom and activities for the real world accordingly (Bacon,1989; Hadley, 2001; Rogers & Medley, 1988). In this vein, Otte (2006) states that, "to develop proficiency in the target language, language learners must be provided with expanded opportunities to both perceive authentic language as it is used as a fundamental means of communication among native speakers…, and to practice using authentic language themselves in order to be better prepared to deal with authentic language in the real world" (p.56). Language teachers should use authentic readings in their classrooms if their students will face them in daily life (Dunlop, 1981). As a result, learners should be faced with "immediate and direct contact with input data which reflect genuine communication in the target language" (Breen, 1985, p. 63).

Last but not least, the content and design of course books is 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.

2.3. The Role of Authentic Materials in Language Teaching and Learning

2.3.1. Arguments in Favor of Authentic Materials

The researchers and teachers in FL/SL teaching have increasingly accepted the use of authentic materials in the classroom. The arguments for the use of authentic texts in language learning may all be reduced to one essential point: that their use enhances language acquisition.

Authenticity enhances proficiency – that is the key point all language researchers - no matter how they define it - agree upon. Over a century ago, Sweet (1964, p. 22) criticized creating texts to clarify grammar points and foresaw the difficulties the textbook writers will face. “If we try to make our texts embody certain definite grammatical categories, the texts cease to be natural: they become trivial, tedious or long-winded, or else they become more or less monstrosities” (Sweet, 1964, p. 192). Krashen supported his ideas and emphasized that texts only need to be comprehensible and get student attention (Krashen, 1989, p. 19-20).

Larsen-Freeman and Long (1991) showed the potential harms of linguistically simplified texts by examining the case studies carried out between 1980 and 1987 and resolved that “[i]nput [linguistic] modifications are not necessary [...] the very process of removing unknown structures and lexical items from the input in order to achieve an improved level of understanding simultaneously renders the modified samples useless as a source of new acquirable language items” (p. 143-4). Krashen (1989) supports their ideas by citing from Blau that simplification could actually impair comprehension (p. 28). They are, therefore, felt to be preferable to simplified texts in that they provide richer and more naturalistic input.

Based on several studies including Miller (2005), Otte (2006) and Thanajaro (2000), the listening comprehension skills and motivation of learners are found to be increasing when authentic listening texts were included in teaching practice. Likewise, Herron and Seay (1991) conducted a similar study on intermediate level students. Two groups of students were chosen for the study and one group listened to authentic radio tapes in addition to their regular classroom curriculum. The other group of students who

were not given extra listening practice was less successful in listening comprehension improvement than those who were given authentic listening activities.

In addition, authenticity enhances reading comprehension skills by introducing new vocabulary (Berardo, 2006). According to Young (1999, p. 361), learners may be misguided by the simplification that everything in a text is important and needs to be memorized. It eliminates the most important elements in communication and thus limits learners' access to language, let alone assisting in comprehension. He conducted a study with 127 second year Spanish language students at a state university on their reading comprehension. Students who read the original – authentic – version of a text and students who read the simplified version of the same text had different recall scores. As can be deduced, authentic text had more scores than its simplified version. Besides, although the simplified texts are significantly more understandable than the original ones, they did not increase the level of learning of specific linguistic areas (Leow, 1993, as cited in Devitt, 1997).

What is more noteworthy is that simplified texts have more grammatical difficulty than authentic ones as Crossley et al. (2007) observe using computational methods in a study to explore the differences between simplified and authentic reading texts. According to Ur (1996), students usually have difficulty understanding texts in daily life because classroom reading materials are not suited to the language of the real world. She wants "…learners to be able to cope with the same kinds of reading that are encountered by native speakers of the target language" (p.150). What is more, Hadley (2001) concludes that the "use of real or simulated travel documents, hotel registration forms, biographical data sheets, train and plane schedules, authentic restaurant menus, labels, signs, newspapers, and magazines will acquaint students more directly with real language than will any set of contrived classroom materials used alone" (p.97). Thus, it would be better to get more of authentic reading materials in classroom.

Furthermore, authenticity in learning materials enhances communicative competence, which is the ultimate goal of almost all language learners. Little et al. (1989) defines an authentic text as the one "created to fulfill some social purpose in the language community in which it was produced (p. 27). From this perspective, Guariento

and Morley (2001) define the teachers' role as the "simulator of the real world" to prepare them for the world outside and "…one way of doing this has been to use authentic materials…" (p. 347). According to Wilkins (1976), authentic texts act as a bridge to fill the gap between theoretical knowledge in the classroom and students' ability to get in touch with the real world. Gilmore in his paper published in 2007 announced his forthcoming study that compared the authentic versus textbook materials on various levels including their effect on communicative competence. The group receiving authentic input made considerable improvement over control group. He concluded that "[t]his result was attributed to the fact that the authentic input allowed learners to focus on a wider range of features than is normally possible … and that this noticing had beneficial effects on learners' development of communicative competence" (p.111). According to Schiffrin (1996), contrived materials fail to meet students' communicative needs and thus, authentic materials have "the potential to be exploited in different ways and on different levels to develop learners' communicative competence (Gilmore, 2007, p. 103).

Authentic materials are not suitable only for advanced level or adult learners. Learners at beginning levels can also benefit from authentic materials. Allen et al. (1988) studied 1500 high school students' abilities to read text materials of different genres and of different difficulty levels (simplified to authentic) after one to five years of foreign language instruction at three different levels of language difficulty. The researchers found that all learners were able to cope with all of the authentic texts they were asked to read, even at the beginning level; "regardless of level, all subjects were at the very least able to capture some meaning from all of the texts" (p. 168). Allen et al.'s study also showed that even beginner learners were able to deal with authentic texts that are 250 to 300 words long without "experiencing debilitating frustration" (p. 170).

Maxim (2002) conducted a similar study on beginners to study the effect of reading authentic texts on beginning level college students. The results of the study showed that students in the treatment group were able to read an authentic popular novel beginning in the 4th week of instruction and to perform at least as well on exams as students who followed the standard syllabus for the entire semester. In addition,

students' restricted language knowledge did not prevent "their ability to read authentic texts" (p.29).

In the same vein, Swaffar (1981) argues that "the sooner students are exposed to authentic language; the more rapidly they will learn that comprehension is not a function of understanding every word" (p. 188). According to Herron and Seay (1991), introducing students to authentic materials at earlier stages will allow them to experience success in advance in language and thus, this will block negative affective factors possible to occur in their following years. In another word, they will promote positive feelings towards language. Bacon (1989), McNeil (1994) and Miller (2005) support Herron and Seay (1991) and suggest that contact with authentic materials should start in the earliest stages of language learning. Duquette et al. (1987, as cited in Bacon, 1990) observed linguistic progress on the language level of kindergarten children after being taught with authentic materials.

The above-mentioned studies and comments from various researchers illustrate the exploit of authenticity and authentic materials in terms of its linguistic benefit. However, authenticity also proposes affective benefits.

Authenticity enhances autonomy. Activities based on authentic texts play a key role in enhancing positive attitudes to learning, in promoting the development of a wide range of skills, and in enabling students to work independently of the teacher. In other words, they can play a key role in the promotion of learner autonomy (McGarry, 1995, p. 3). According to Fernandez-Toro and Jones (1996, p. 200), learners at high proficiency levels benefit more from authentic material in autonomous modes.

In addition, that authenticity boosts motivation is a popular subject in the literature (Wipf, 1984; Swaffar, 1985; Little, Devitt & Singleton, 1989; Morrison, 1989; Bacon & Finnemann, 1990; King, 1990; Little & Singleton, 1991; McGarry, 1995; Peacock, 1997). Making learners feel that they are able to cope with an authentic text can be considered as an inherently motivating force towards achievement. By means of authentic material, students are quite capable of drawing inferences from the material rather than relying on the instructor's interpretation or personal experience" (p. 26). Many more researchers such as Gilmore (2007) and Sherman (2003) argue for the

motivation which authentic materials may provide and which is a key factor influencing successful language learning (Samimy&Tabuse, 1992; Masgoret & Gardner, 2003; Krashen, 1981).

It also builds self-confidence of students. Research on students' attitudes towards authentic foreign language videos revealed positive results (Wen, 1989; Baltova, 1994). Terrel's views (1993) are parallel to Wen and Baltova, adding that students have more confidence with language after exposure to authentic materials. Various other studies (e.g. Kim, 2000; Otte, 2006; Peacock, 1997; Thanajaro, 2000) have also observed an increase in students' motivation and self-satisfaction after exposure. Authentic materials are believed to be more interesting than simplified ones because they aim to communicate a message rather than to emphasize a language topic (Swaffar, 1985; Hutchinson & Waters, 1987; Little, Devitt & Singleton, 1989; King, 1990; Little & Singleton 1991).

To sum up, authentic materials in language classrooms have positive effects both on linguistic and affective aspects of the learning.

2.3.2. Arguments against Authenticity

Despite this fame of authenticity, there are still many researchers and teachers who emphasize the value of simplified texts, especially for lower level of learners (Johnson, 1981; Shook, 1997; Young, 1999).

Perhaps the most influential hypothesis supporting the use of simplified texts is Krashen's (1981, 1985) theory of comprehensible input. Simply put, this theory states that learners' proficiency increases by exposing learners to input that is a little beyond their current language ability (i +1 system). As long as the input is at a level for learners to understand, the learner will be exposed to the necessary language features. As a result, unabridged, real life interactions in an authentic context will not be i + 1 level for most learners, especially for lower level of learners.

Morrrow (1977) gives a more sharp statement by saying that real authenticity is "unattainable. We cannot recreate absolute authenticity in the texts we use." (p. 14-15). By using real language situations in made-up classroom settings, we are actually

destroying its reality. Because according to Hutchinson and Waters (1987) “A text can only be truly authentic [...] in the context for which it was originally written.” (p. 159)

Furthermore, returning to Breen’s (1985, p. 60) categorization of authenticity, it has four dimensions. Authentic language learning cannot be attained just by using real life texts. In addition to text authenticity, we talk about task, learner, and test authenticity. As a result, it can be assumed that just using authentic texts does not necessarily mean that tasks and language learning will be authentic (Arnold 1991, p. 238). According Taylor (1994), authenticating texts, classroom and tasks are not necessary since “the classroom itself is a real place” (p. 1). He disregards the nativeness of authenticity and embraces the definition that as long as a real communication takes place within real audience, there is authenticity achieved.

Clark (1983) argues that media do not influence learning under any circumstance; thus, the issue of authentic versus non-authentic makes no difference (p.224). Mihwa (1994) found that the reading comprehension level of lower-level ESL learners was not affected by the type whether the text is authentic or simplified. Davies (1984) also prefers simplified texts to authentic ones. Further, Kienbaum et al. (1986) found no considerable variation in the language performance of children using authentic materials compared with those in a more traditional classroom context.

Richards (2001, as cited in Kilickaya, p.253) mentions that authentic materials are often unsystematic in terms of language, vocabulary length and content and therefore cause difficulty for the teacher in lower-level classes. According to Rogers and Medley (1988) authentic materials are usually accepted as too difficult to be understood. Hadley (2001) warns that unedited authentic materials are "random in respect to vocabulary, structure, functions, content, situation and length, much of it impractical for classroom teachers to integrate successfully into the curriculum” (p.128). According to Gilmore (2004) “There is a danger in authentic texts […] distracting peripheral information […] will confuse students and obstruct acquisition of the target language” (p.366). In addition, they may be culturally biased, making them impossible to understand outside the community they are produced in (Martinez, 2002).

Ur (1996) and Dunkel (1986) also caution that presenting the students with difficult material can damage morale and motivation. According to some researchers, beginner level learners get too little of authentic materials. According to Guariento and Morley (2001), "At lower levels, however, … the use of authentic texts may not only prevent the learners from responding in meaningful ways but can also lead them to feel frustrated, confused, and, more importantly, demotivated" (p. 348). As learners at lower levels do not have necessary basic background of the target language, they will probably feel demotivated and discouraged (Kilickaya, 2004; McNeil, 1994). Kim (2000) asserts that comprehensible input at lower levels cannot be achieved with authentic language. Besides, Schmidt (1994) prefers using simplified texts because authentic communication may fright learners with a mixture of known and unknown vocabulary and structures.

McNeil (1994) states that "[i]t is often difficult for the teacher to find an appropriate pedagogical function for authentic materials" (p.314), and that it causes students not to see the benefit of using them. In addition, teachers may have difficulty accessing to authentic materials, purchasing them and designing suitable tasks for the class. For instance, in a study conducted by Al-Musallam (2009), teacher complained about the lack of time, heavy teaching load and obligation to follow the curriculum, and these resulted in less use of authentic materials in language classrooms although most of them favor the spending more time on authentic materials.

To conclude, there are too many arguments on the issue of authenticity after communicative approach to language learning began to flourish. Views vary from strong criticism to encouragement. However, even the strongest critics do not reject authentic materials completely (Walz, 1989), but warn of dangers if not used wisely. This can be overcome by a careful planning of tasks that will eliminate the disadvantages mentioned above. According to Guariento and Morley (2001), "[a]s long as students are developing effective compensatory strategies for extracting the information they need from difficult authentic texts, total understanding is not generally held to be important; rather, the emphasis has been to encourage students to make the most of their partial comprehension" (p.348).

2.4. Selection of Authentic Materials

The authenticity issue has a very important position not only among material writers but also among language teachers as the practitioner of the content. Teachers should find, design, change and implement the learning materials instead of directly taking them. Therefore, selection of authentic materials is an essential factor to keep in mind.

"Authentic materials enable learners to interact with the real language and content rather than the form" (Berardo, 2006, p.62). Learners feel that they are learning a target language as it is used in the real world outside. Nuttall (as cited in Berardo, 2006) mentions of three main criteria when choosing texts to be used in the classroom; suitability of content, exploitability and readability. However, Berardo (2006) adds a fourth criterion: suitability of content, exploitability, readability and presentation. Suitability of content includes relevance to students' needs, being interesting and closeness to real life. Lee (1995) argues that materials should be learner-centered and promote learners' interest. Exploitability, on the other hand, deals with whether the text can be exploited for teaching purposes and whether it serves to the purpose. Just because a material is in the target language does not make it suitable to use in the classroom. Readability, as the name suggests, refers to structural and lexical difficulty of the material chosen. Not all authentic materials are at the same level with students. Finally, authenticity is also important when presenting it to students. A more eye-catching text will get more attention from students and motivate them.

According to Breen (1985), following questions must be answered in order to build a bridge between authentic texts and learner:

"—What is an authentic text?

—For whom is it authentic?

—For what authentic purposes?

—What is authentic to the social situation of the classroom?" (p.61)

Lee (1995) suggests four guiding principles in authentic text selection.

"Is the material textually authentic?

Is the material compatible with the course objectives?

Is the material suitable for the teaching approach we adopt?

Is the material suitable for the tasks/activities designed?" (p.326)

Guariento and Morley (2001) claim that at lower levels authentic texts should be chosen according to their lexical and syntactic simplicity and content familiarity. Brown and Eskenzai (2004) basically share the same viewpoint, arguing that the main criteria for selecting authentic text should be the reader's current vocabulary knowledge and the desired vocabulary knowledge, that is lexical density. Rivers (1987) states that "[a]lthough length, linguistic complexity, and interest for the student all play significant roles in the selection of materials, the single most important criterion for selection is content" (p. 50). According to Mishan (2005) learners' needs are primary determinants in the choice of authentic texts. In this sense, Little et al. (1989) state that "the quality of a given psychological interaction relates to the extent to which the interactant sees the material being processed as having personal significance" (as cited in Mishan, 2005, p.28).

2.5. Studies Related to Attitudes towards Authenticity

There is little empirical research on the attitudes of students and teachers but some exist to give us a picture - but blurry - on the issue. Hillyard, Reppen, and Va´squez (2007) found out that a group of students had reported great satisfaction of being exposed to authentic texts. However, the result was not acquired using data collection methods (questionnaires, surveys etc.), but it was based on class discussion. A similar observation was recorded by Berardo (2006) with advanced learners of engineering students. The use of authentic reading materials was discussed with students and s/he noted high motivation and a sense of achievement on students' side, benefiting from exposure to authentic language.

Kim (2000) tried to analyze students' language level and attitudes after exposure to authentic input. Twenty-six Korean university students from two groups were

randomly selected and interviewed. The experimental group was instructed with both graded and ungraded input, whereas the control group was instructed with only graded input. Each subject was interviewed at the end of the study. The results showed that both groups reported considerably low levels of confidence in their understanding of authentic input. However, the majority of students in the experimental group reported that their attitudes toward authentic input were changed positively, and their English proficiency improved over the treatment period. Results of the study showed that the majority of students in the experimental group reported that their attitudes toward authentic input were changed positively, and they believed that their English proficiency improved over the semester.

Additionally, Peacock (1997) tested whether authentic materials increase the classroom motivation of learners, "a claim often made but rarely… tested" (p.144). Two beginner-level EFL classes used authentic and artificial materials alternately. Results indicate that on-task behavior, observed and self-reported motivation increased significantly when authentic materials were used. Bacon and Finnemann (1990) studied first-year Spanish students' willingness to face authentic input. The results indicated that exposure to authentic input had a positive effect on comprehension and satisfaction but a negative effect on frustration.

Gilmore (2007) examined the Japanese students' communicative competence after exposure to authentic materials over a 10-month quantitative/qualitative classroom-based study. Ninety-two 2nd year English major students, of similar proficiency levels, were assigned to either a control or experimental group for the period of the trial. The control group received input from two textbooks commonly used in Japanese universities, while the experimental group received input from authentic materials (films, documentaries, 'reality shows', TV comedies, web-based sources, songs, novels and newspaper articles). Putting communicative competence question aside, one-to-one interviews with students showed that students displayed their marked preference for authentic materials over textbook materials, despite the challenges this posed.

Gonzalez (1990) investigated the impact of using authentic materials as textbook supplements on FL students' attitudes, motivation, culture and language achievement.

Four classes totaling 43 students participated in her study. The results showed that there are statistically significant differences between the experimental and control groups in achievement, and that students, on the whole, responded favorably to the use of authentic materials.

Al-Musallam (2009) examined EFL learners' and teachers' beliefs and attitudes on the use of authentic materials. The analysis of the results indicated that learners and teachers had positive attitudes toward the use of authentic materials in their reading classes. In fact, they indicated that an ideal reading class should use a combination of both authentic texts and textbooks.

Lee (1995) conducted a similar study. A three-week supplementary English programme at Hong Kong Polytechnic University, aiming to improve general English proficiency was carried out exploiting textually authentic materials which are selected according to students' needs and interests. After the course, most of the students gave positive affective and cognitive responses. Materials were interesting, useful, readable, accessible, they said.

2.6. Previous Studies on Using Authentic Materials in the Turkish Context

There are several studies, mostly of Master of Arts dissertations, on the use of authentic materials in language classrooms. Torun (2008) investigated the role of contextualized language instruction through authentic animated stories in teaching EFL to young learners. 31 sixth grade students aged 11-12 participated in the study. Five authentic animated stories were used. On the whole, the outcome of the study revealed that students kept their initial positive attitudes towards learning English with contextualized language instruction through authentic animated stories. The study also showed that students' anxieties related to learning English declined through such instruction.

Ozgen (2008) conducted similar research to find out the positive effects of captioned authentic video on the listening comprehension of intermediate level university students. This study was conducted over 8 weeks in which the two groups of students watched the same chosen episodes of a sitcom, and took the

same pre and posttests, and questionnaires. The students in the control group watched the episodes without captions while experimental group watched the episodes with captions. The results illustrated that the group with captioning scored significantly better on the listening comprehension test and perceptions on authentic materials were all positive.

Aktan (2010) tried to illustrate the use of authentic French songs in language classrooms. In his descriptive study, he showed the possible benefits of the songs as a lesson and self-improvement material. However, his study was not empirical. He only confirmed the use of authentic songs by extensively referring to related literature.

Boran (1999) examined EFL instructors' attitudes to authentic materials and using television programmes as authentic video activities so that the similarities and discrepancies between the students' attitudes and the instructors' attitudes emerged. The students' attitudes to the video activities, in which television programmes were used, were found to be positive. The instructors' attitudes to such activities and other authentic materials were also positive.

Kilickaya (2004) tried to answer the questions of when and how authentic materials should be used in EFL classrooms, and how cultural content may be included in the curriculum.

2.7. Authentic Material in Course books

The idea of authenticity that a text has to be true to its original cultural context calls up the question of inclusion of such texts in ELT course books. Early course books were infamous for being full of realia. This use of realia has been blamed of making a 'touristic' rather than a cultural approach (Shanahan 1997: 165). However, culture should not be conveyed "in an anecdotal non-reflective manner" (Kramsch 1998: 82) since it weakens authenticity.

According to Mishan (2005), there is a significant number of research to advocate the use of "linguistically rich, culturally faithful and potentially emotive input" (p. 11) given by authentic texts. Moreover, there is little support as to why the course books are the way they are now. The content is given in specific order in course books

and the reason is unknown. After all, fixed acquisition order has little support from research evidences (p.11). Yet, most course book subjects are ordered the same way, having no scientific background to do so.

The arguments for and against the use of course books in language teaching and learning literature will be presented below. Most prominent studies conducted on the topic were briefly explained.

2.7.1. Arguments in Favor of Course Books

According to McGrath (2006) "[c]ourse books are a central element in teaching-learning encounters, not only in school settings but frequently also in tertiary-level service English contexts. They will tend to dictate what is taught, in what order and, to some extent, how as well as what learners learn" (p.171).

Course books are of fundamental importance in most language programs in all types of educational situations. They are the first and most basic tools playing significant roles in getting information in a lesson. The characteristics of textbooks have, therefore, decisive traits upon teaching and learning situations.

Harmer (1991, p. 257) claims that course books have clear advantages for both teachers and students. Quality course books often offer materials in a lively and interesting fashion; they provide the language items in a definite and concise order; they supplement the students with aim of the course and how much they have covered the curriculum. In addition, course books give learners a chance to learn at their own pace and time outside the classroom. They also have the advantage for teachers of having course materials and activities within an arm's reach, without having to devise original materials for every class hour.

Richards (2001) argues that without course books, a course may have no impact, for they provide structure and a syllabus. It is the course book that provides the syllabus (Mishan, 2005). Moreover, the use of a course book in a program can assure that students in different classes or in different cities will get the same content. In other words, they assure the standardization in education.

On the advantages of course books, Ur (1996) names two benefits in relation to meeting the curriculum objectives: a) they provide a clear framework which the teacher and the students know where they are going and what is coming next, b) they serve as a syllabus which includes a carefully planned and balanced selection of language content.

Cunningsworth (1995) argues that they are an effective resource for autonomous learning and for presentation material, a source of ideas and activities, a reference guide for students, a syllabus showing learning objectives, a resource for self-directed learning and support for teachers at the beginning of their careers. Since many teachers are busy and preparing extra materials is a time-consuming duty, course books are very important in the language teaching and can be seen as helpers of the teachers, especially for the novice ones (Garinger, 2002).

In addition, Ansary and Babaii (2002) consign textbooks some additional worth underlining that course book is a cheap way of providing learning materials and that a learner without a course book is out of focus and teacher-dependent. That is an issue almost all teachers are trying to defer in communicative language teaching/learning age. Richards (2001) argues that they preserve quality and that they are usually attractive.

Chall, Conrad, and Harris-Sharples (in Xu, 2004) call attention to a very remarkable point proper to the extensive use of ELT course books claiming that by the time most students complete high school, they will have been exposed to over 32,000 pages in course books and that almost all of their time in reading instruction and at least three-fourths of their time in content classes will be spent with only a course book. Bearing this extensive exposure to course book language in mind, its design, content and quality are essential features to consider in language learning attempts.

Richards (2001:1) summarizes the issues mentioned above as:

> *In some situations, they (textbooks) serve as the basis for much of the language input learners receive and the language practice that occurs in the classroom. They may provide the basis for the content of the lessons, the balance of skills taught and the kinds of language practice the students take part in. In other situations, the textbook may serve primarily to supplement the teacher's instruction. For learners, the textbook*

may provide the major source of contact they have with the language apart from input provided by the teacher. In the case of inexperienced teachers textbook may have also serve as a form of teacher training – they provide ideas on how to plan and teach lessons as well as formats that teachers can use.

2.7.2. Arguments against Course Books

The course books are not without critiques, though. While many language teaching professionals present the benefits of using course books in foreign language teaching and learning, some researchers object to the idea of relying on course books since published materials may not always offer types of texts and activities that a teacher is looking for a given class (Block, 1991).

On the inefficiency of course book use in the classroom, many teachers and researchers complain about the artificiality of the language used in the course books. The language presented by course books to language learners are, without doubt, poor representation of real language. The language the potential learners will face while language encounter is informal, colloquial and far away from bookish style of course books. Therefore, syllabus design should be changed according to "what language is and what people use it for" (Mccarthy & Carter, 1994, p. 201). Holmes (1988) surveying four ESL-textbooks finds that modal items are not represented properly. Römer (2004) also analyzes if-clause usage of German EFL text books and she finds considerable differences between what is in textbooks and in real life.

Therefore, various problems concerning using course books in language classes have been addressed by a number of researchers. Richards (2001) claims that textbooks sometimes present inauthentic language since texts, dialogues and other aspects of content tend to be produced to illustrate specific language points and are, therefore, not often representative of real language.

In his study on the discourse features of seven dialogues published in course books, Gilmore (2004) also shows the inauthentic language use in course books by proving that the course book dialogues differ noticeably from their authentic equivalents. Teaching materials often instruct students on the grammar of the target language without any link with the native language. However, students do draw

conclusions on their own trying to connect the new concept with something they already know. In this way, they are in danger of making erroneous conclusions and acquiring insufficient knowledge (Gajic, 2010).

According to Mishan (2005), there is a significant number of research to advocate the use of "linguistically rich, culturally faithful and potentially emotive input" (p. 11) given by authentic texts. By evaluating both advantages and disadvantages of course books, it can be deduced that the course books are one of the central elements in language classrooms and thus, should be in line with the real life outside and expose learners to real language.

2.8. Corpus Linguistics

Ironically, as the need for learning foreign languages for communicative purposes increased, the authenticity of the languages in terms of materials tended to decline (Mishan, 2005). The importance of authenticity and course books in language classrooms are dealt with above referring to comprehensive literature on the topic. As a result, we can conclude that authenticity is an issue that must be inherent in course books. At present, ELT course books are making more use of authentic texts and find their ways in the market. Under the subheading corpus linguistics, we will highlight how an enhancement of English language teaching materials can be attained on the basis of native speaker corpus data.

The corpus is a computer database storing large collections of language, which is available for analysis by linguists and other researchers. Computer corpora have been around since the 1960s (Kennedy, 1998) but advances in storage capacity and computer technology have brought out their potential more fully in the last 20 years. According to McEnery and Wilson (1996) it is the "study of language based on examples of "real life" language use" and "a methodology rather than an aspect of language requiring explanation or description" (p.1).

A corpus is defined as a principled collection of naturally occurring texts which are stored on a computer to permit investigation using special software. A corpus is principled (Biber, Conrad and Reppen, 1998) because texts are selected for inclusion

according to pre-defined research purposes for qualitative and quantitative analysis. It is not just a random collection of language. For example, a researcher who wants to investigate metaphors used in university lectures will try to gather a representative sample of lectures across a number of disciplines, rather than attempting to collect lectures that contain a lot of figurative language of a specific discipline.

Although there are many - though the same in essence - definitions of a linguistic corpus, Meyer's (2002) definition can be accepted as the most inclusive, except that he did not mention the real life bound of a corpus: "a collection of texts or parts of texts upon which some general linguistic analysis can be conducted" (p. xi). After all, why do we need a language corpus if not for a language analysis? Authors of corpus studies are absorbed in analyzing large amounts of authentic texts with the help of computational tools for text analyses (Biber et al., 1999; Flowerdew, 2001; Granger, 1999; Guillot, 2002).

2.8.1. Authenticity of the Corpus

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. Mishan (2005) states that "[c]onsisting of thousands of 'authentic' texts on a single platform, it would seem self-evident that corpus data is 'authentic'. 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.

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".

2.8.2. Corpus-based Research

This study is different from many course book analyses in that it adopts corpus-based analysis. Biber, Conrad & Reppen (1998) defined the characteristics of a corpus based analysis. First of all, it is empirical in that it analyzes the actual patterns of use in natural texts. Next, it uses a principled collection of texts (corpus) for analysis. Another important aspect is that it mostly uses computers and special software for analysis. Lastly, a corpus-based study depends both on quantitative and qualitative techniques. Judged altogether, these features result in a highly reliable analysis (Biber et al., 1998).

The most important component of a corpus approach is computers, making it possible to analyze millions of words and sentences at a glance, which would be impossible to process by hand. In addition, they provide consistent and reliable analyses, which do not change over the time or according to researcher. This eliminates emotive factors in scientific research.

In the same vein, Granger (1999) identifies three basic benefits of a corpus study, which are frequency, variation and lexico-grammar. Frequency shows items which are more frequent than others; variation presents what is more likely to be faced in daily life; and lexico-grammar explores how a certain word behaves in natural context.

2.8.3. Corpus Based Course book Analyses

In many subfields of applied linguistics, corpora have made their presence uniquely felt. They are now common in lexicography, translation studies, forensic linguistics and even critical literary appreciation. Krishnamurthy (2002:4 as cited in Ranalli, p.3) states that “all of the current EFL dictionaries make some claim to the use of corpora in their compilation”.

Corpus linguistics has also become influential in several areas of language teaching. More specifically, corpus - based analyses have been particularly relevant to EFL course book 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

(Gavioli and Aston, 2001; Sealey and Thompson, 2004; Biber and Reppen, 2002; Barbieri and Eckhardt, 2007). Following the rise of corpus linguistics and its thorough analyses of exact data, these decisions are increasingly being made on the basis of the empirically provable results and conclusions of corpus-based analyses.

Educators noticed that students learning a foreign or second language should be exposed to such educational context that reflects real-life situations. To have a better understanding of everyday language, linguists started collecting data outside of classrooms. This authentic data was studied by means of computers. Computers allowed study of texts counting millions of words, thus giving birth to corpus studies.

As Barbieri and Eckhardt (2007) argue, "to date, little or no effort has been made to apply corpus-based findings to language teaching in a way that reflects current SLA principles and theories." 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).

Linguistic studies contained grammatical and lexical analyses. Since it is believed that one of the leading principles regarding the content of teaching materials should be the frequency of items in corpus (Sardinha, 1999), frequency counts were considered in the analyses. J. Flowerdew (2001) stressed that frequency data could tell

us which items should be selected to teach while concordancing information could tell us how these items can be used. This data might be used to determine what to teach in which order and to evaluate syllabus.

Regarding frequency data, Biber and Reppen (2002) investigated the difference between the information presented in ESL-EFL teaching materials and the information regarding actual language use using corpus comparison technique. They tried to explore a) the grammatical features to be included, b) the order of grammatical topics, and c) the vocabulary used to illustrate these topics. Their findings showed a serious inconsistency between books and real-world language. The results for regarding the use of progressive and simple aspect showed that progressive aspect is used more than other in course books compared to authentic texts. It also indicated that 12 most common lexical verbs were ignored by course books. As the result of the study, Biber and Reppen (2002) concluded that:

> *Given its importance in acquisition, we would argue that frequency should also play a key role in the development of materials and in the choices that teachers make in language classrooms. With the recent availability of comprehensive frequency-based grammatical descriptions, such integration of pedagogy and research has become feasible (p. 207-208).*

Some other researches also take frequency data into consideration. Römer (2004a; 2004b; 2005), Gilmore (2004), Anping (2005), Hyland (1994), Gabrielatos (1994), Nitta and Gardner (2005) and several other scholars investigated course books regarding one or two aspects such as authenticity, grammar, vocabulary choice, discourse features etc. Compared to language teaching/learning literature they are quite a few in number and not enough to deduce implications for language teaching practice in general.

For instance, Römer (2004a) tried to find out similarities and differences regarding modal auxiliaries between the language confronted in EFL course books and authentic English as it is used in natural situations. She concluded that "the way the topic of 'modal auxiliaries' is treated in English lessons … differed considerably from the use of those verbs in contemporary spoken British English" (p.197). She observed

that corpus-driven approaches to language learning and teaching can be very helpful for teachers and publishers.

In a similar study, Römer (2004b) investigated the use of conditional clauses in course books and British National Corpus. She noticed that if-clauses taken from course books have usual classification of 'type 1', 'type 2' and 'type 3'; however, real life examples have many different types and irregularities as opposed to the ones described in course books. At the end of the study, she refers to Glisan and Drescher (1993:32), who state that "authentic language must continue to be examined if we are to use real language as the basis for our teaching" (cited in Römer, p.162).

In 2006, Römer, again, made a book-length detailed analysis of progressive patterns in course books compared to BNC. The results of the study showed that corpus-based research has proved to be of help to material writers and teachers. Table 2.1 summarizes the studies conducted on course book materials by various researchers.

Table 2.1 Corpus Based Course Book Analyses in the Literature

Research area	Author	Focus	Level	No of vol.
	Römer (2004a)	modal auxiliaries	EFL	6
	Römer (2004b)	if clauses	EFL	12
	Römer (2006)	progressive	EFL	12
	Gilmore (2004)	discourse features	EFL, EGP	7
	Anping (2005)	vocabulary;grammar	EFL	50
	Hyland(1994)	modals	EAP	22
Authenticity	Gabrielatos (1994)	possessive	EFL, EGP	1
	NittaGardner(2005)	grammatical tasks	EFL, EGP	9
	Boxer&Pickering(1995)	speech acts	ELT	7
Grammar	Vellenga(2004)	speech acts	ESL, EFL	8
	Miura(1997)	oral communication	ELT	16
Pragmatics	Cane(1998)	conversation skills	ELT	3
	Chujo(2004)	vocabulary levels	EGP, ESP	7
Speaking	Ranalli (2003)	learning strategies	EFL, EGP	3
	Reda (2003)	vocabulary	EFL, EGP	6
	Gabrielatos (1994)	collocations	EFL	3
	Hill(1996)	verb form clustering	EFL	?
	Biber et al.(2004)	lexical bundles	EAP	?
	Koprowski (2005)	lexical phrases	EFL, EGP	3
Vocabulary	Meunier&Gouverneur (2007)	phraseology	EFL, EGP	5

As for Chujo's work (2004), it aimed to measure vocabulary change across levels in EGP and ESP textbooks. The texts included in the textbooks were scanned and part-of-speech tagged. Lemmatized wordlists were then computed and compared to wordlist from the British National Corpus. Chujo also compiled specialized vocabulary wordlists from the course books and from tests.

Anping (2005) tried to explore "how the language content selected and exercises/tasks designed in the current EFL textbooks in China have revealed the modern ideology of ELT education, especially the ideas of making 'real-life' language input and providing guidance for inquiry and explorative learning" (p.1). By exploiting vocabulary lists, key words lists, lexical bundles and concordances of particular words and patterns, she compared the course books. According to Anping (2005), "ELT course books, with language items as their major components, have become a special genre for corpus analysis" (p.1).

Gabrielatos (2006) also studied 'if-sentences' used in a number of ELT course books, with reference to a random sample of 1000 if-sentences from the written subcorpus of the BNC. The study focused on frequency, modality and special cases of if-sentences in course books and their comparison with the ones taken from BNC. He illustrated three basic short comings of ELT course books: "a) It provides learners with an incomplete, and in some cases distorted, picture of if-conditionals…, b) It tends to overwhelm learners with long lists of 'special cases' or 'exceptions… and c) It potentially limits the learners' language production by restricting their repertoire to a small number of pre-fabricated combinations" (p.2).

As can be seen, there are quite a few studies carried on corpus-based course book research, a field of study which is located at the intersection of language description, pedagogical grammar, and pedagogical materials evaluation (e.g. Owen, 1993; Hunston & Francis, 1998; Harwood, 2005; Römer, 2004, 2005).

2.8.4. Corpus Based Course book Analyses in Turkish Context

There are many course book evaluation studies in Turkey as mentioned in section 2.6 above. However, no corpus-based course book analysis, to my best

knowledge, is conducted to look for ideal course books. As a newly flourishing field of linguistic studies, corpus studies have just began 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, and thus, this study is aimed at filling this gap in literature in material evaluation studies in Turkey.

2.9. Conclusion

The review of literature has illustrated that there is not a general agreement among scholars on the use of authentic materials in language classrooms. All three language teaching approaches - materials based, humanistic and communication based - seem to favor its use one way or the other. Actually, it is an issue debated as early as the beginning of language studies as Sweet (1964) observed. Finally, the unreal representation of real life in course books is another trigger for authenticity dispute. Therefore, some scholars have studied the impact of authentic materials on language performance and students' motivation and observed the advance in performance and motivation of students as a result of exposure to authentic language. On the other hand, some studies have shown that authentic materials have no impact on students' language performance. However, even those who voted against the use of authenticity in class do not totally reject it. Most of them stressed the difficulty of its implementation in the class, defining extra authenticity types.

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. However, there is little effort to use this authentic corpus data in course books. Although publishers praise their books on the closeness to real life, they do not seem to be using this precious authentic corpus data.

In the next chapter, the data collection and analysis methods adapted to determine authenticity levels of course books were presented. The software, tools and analyzed items were explained in detail.

CHAPTER 3

METHODOLOGY

3.1. Research Design

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. Table 3.1 below summarizes the main elements of methodology of the study.

Table 3.1 Overall Research Design

Research Design	Quantitative
Sampling Strategy	Convenience sampling
Materials	All ELT course books used in high schools in Turkey with supplements (12 in total)
Data Collection Tools	• Corpus of ELT Course Books used in Turkey (TEFL CC) British National Corpus XML Edition
Data Analysis Tools	• British National Corpus XML Edition • Sketch Engine Corpus Query Tool • Paul Ryson's Log-likelihood calculator
Analyses	• Frequency Analyses • Frequency Comparison • MI Score comparison • T-score comparison • Log-likelihood Calculation
Time and Duration	1 year

This study is a quantitative study which aims to find sufficient and insufficient parts of ELT course books used in Turkey related to spoken part of the BNC and was conducted during 1 year, from November 2011 to June 2012, to find answers to the following research questions:

1. What is the degree of authenticity of language course books used in Turkey compared to spoken part of BNC?
2. How much do specific grammar points in course books resemble to the grammar of BNC spoken corpus?
3. How much do vocabulary choice in course books resemble to the vocabulary of BNC spoken corpus?
4. Are Turkish students of language exposed to the input they may really need in communicative situations?

Quantitative methods were used to collect data regarding the research questions above. Firstly, a general representation of some specific grammar points (only tenses and modals in this situation) in the spoken part of the BNC was determined regarding their frequency, and with which verbs they go together. The collocation strength of each item was calculated using MI (Mutual Information) and T-score. Later, the same method mentioned above was administered to the items extracted from the corpus of ELT course books (TEFL CC) used in high schools in Turkey. Finally, the frequency counts, their Log-likelihood values and MI and T-Scores 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 New Bridge to Success, language teaching course books, which consist of 4 books and are used in all general high schools. Each of them is used with different level of students during four years of secondary education. To give all students in the country an equal level of education, all high schools in Turkey have a standardized curriculum for each course. The curriculum for each course also defines which materials or course books to be used in teaching process. Accordingly, all course books are in line with the curriculum in terms of the order of lessons, time to spend on a unit, time to finish the course, etc. As a result, to determine whether students are exposed to naturalistic language input at course book level in classrooms, 4 ELT course books with supplements, used in all state and private high schools were chosen. 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 learners, they were also included in the study for analysis.

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. 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. As a result, a high level of computer literacy was a must to carefully conduct the study. Before commencement of the study, the researcher had familiarized himself with the usage of raw corpus data, corpus creation steps, corpus query tools, different part-of-speech tagsets defined by different authors, query language, statistical methods to compare the frequencies and to compute collocation strengths, and what those number may mean.

In addition, all collected date had to be organized in an efficient manner, since it would be impossible to work with all those numbers, statistical scores and words. For this reason, the researcher used tools such as Microsoft Excel, an electronic spreadsheet program, as it features calculation, graphing tools, pivot tables, and a macro programming language.

3.2. Data Collection and Analysis Techniques

In this section, the tools (web interfaces and software) used to compile corpus, to make word and POS tag searches, to extract word and collocation strength lists, and to compare frequencies are explained.

3.2.1. Sketch Engine

For data extraction from both corpora, SketchEngine Corpus Query System was employed. It is believed to be the best program 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.

The Sketch Engine is a web-based program which takes as its input a corpus of any language and a corresponding grammar patterns with an appropriate level of linguistic mark-up. The Sketch Engine has a number of language-analysis functions and the important ones are:

The Concordancer: A program which displays all occurrences from the corpus for a given query. The program is very powerful with a wide variety of query types and many different ways of displaying and organizing the results. For instance, one can search a particular word alone or make a very complex POS search such as 'the verb be followed by an adjective before a preposition'. It is very helpful for querying grammatical categories and functions. In addition, one can find the 1, 2, 3, 4, 5 left or right collocates of a search word, and get their statistical relations in numbers. It computes collocational strength of specified words using many different statistical techniques like T-score, MI, MI3, log likelihood, logDice etc. To get the frequency counts, collocations and collocations strength analyses, this feature was extensively used during the study for both corpora.

The Word Sketch: This program provides a corpus-based summary of a word's grammatical and collocational behavior.

The Create Corpus Program: This program is used if one wants to create a corpus in the interface and use it on-line like other corpora. The TEFL CC was created by using this function. It automatically analyzes the text, annotates it and processes it ready for use.

3.2.2. Loglikelihood Calculator

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. The researcher may wonder whether the frequency differences between two corpora are just a random, chance happening. The statistical term for this is "significance". If results are significant, we are reasonably certain (usually 95% certain, sometimes 99% certain) that these results are not due to chance. Very often, researchers have to test whether the results are significant.

It is important to know what words are overused or underused in 'analysis corpus' when compared with 'reference corpus'. If one word appears, say, 5% of the time in the small wordlist and 6% of the time in the reference corpus, the occurrence will not be considered significant, though it may well be the most frequent word. Therefore, a more consistent statistical significance test should be used. 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. However, Williams (1976) notes that the log-likelihood is preferable to Pearson's chi-squared in general. In this study, it is used to determine how significant the frequency counts of defined language items extracted from both corpora are. Log Likelihood test, which gives a better estimate of significance, especially when contrasting small corpora against a reference corpus. There are two main types of corpus comparison: a) comparison of a sample corpus with a larger standard corpus (e.g. Scott, 2000) b) comparison of two equal sized corpora (e.g. Granger, 1998). This study, being the comparison of small course book corpus with a large reference corpus, thus, exploits the log-likelihood test scores.

Rayson & Garside (2001) summarizes the advantages of the log-likelihood ratio over the other measures as follows:

1. LL values are directly comparable

2. LL is not as expensive to compute as Fisher's Exact test, and gives similar results for large sample sizes

3. LL has been shown to be better 'in general' than the chi-squared test

4. The chi-squared statistic is an approximation to the LL for large samples

The mathematics behind log likelihood is quite complex, but fortunately there are many softwares and web sites that compute the value in seconds. Web-based log-likelihood wizard is available on-line and free-of-charge, provided by Paul Rayson. In addition, the site also allows downloading an Excel file, prepared for computing log-likelihood scores. To compute the log-likelihood value of an item, the 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.

3.3. Data Collection Steps

In the study, quantitative analysis techniques were employed. The relative frequencies, collocation strengths, cluster lists and log-likelihood results were compared to show the similarities and differences between two corpora. The tools for data extraction are explained above. The data analysis will be a one-to-one comparison of scores.

3.3.1. The Reference Corpus Selection

In order to compare the data findings from the course book, a fine reference corpus representing the natural English in daily life had to be chosen. There may be thousands of corpora created for different purposes. However, not all of them are suitable to the nature of this study. A corpus, highlighting the general characteristics of language, proportionally balanced with different genres, covering both written and spoken language is what is called as reference corpus.

There were a few corpora for this purpose. American National Corpus (ANC), Longman Spoken American Corpus, Corpus of Contemporary American English (COCA) and COBUILD Bank of English are some of the most popular reference

corpora in the world. However, when the size, efficiency of use, easy access, price, downloading the data option, statistical analysis features, employability with different corpus query softwares and interfaces were taken into consideration, the candidates thinned to only one: BNC. 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. Written texts were selected using three criteria: "domain", "time" and "medium". Domain refers to the content type (i.e. subject field) of the text; time refers to the period of text production, while medium refers to the type of text publication such as books, periodicals or unpublished manuscripts. The spoken data in the BNC was collected on the basis of two criteria: "demographic" and "context-governed". The demographic component is composed of informal encounters recorded by 124 volunteer respondents selected by age group, sex, social class and geographical region, while the context-governed component consists of more formal encounters such as meetings, lectures and radio broadcasts recorded in four broad context categories.

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.

3.3.2. Course Book Corpus 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 software 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. At the end of this process, the raw TEFL CC was completed.

3.3.3. Course Book Corpus POS tagging

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. A simplified form of this is commonly taught to school-age children when they are asked to identify words as nouns, verbs, adjectives, adverbs, etc.

There are many automatic POS taggers present on the market or free of charge. After a detailed search on the issue, the researcher decided to make use of SketchEngine Corpus Query System, incorporating word sketches, one-page, automatic, corpus-derived summary of a word's grammatical and collocational behavior. Its corpus creation feature enabled the researcher to upload the corpus of course book and to make it tagged using the PENN Treebank Tagset. The following part-of-speech tags are used in the corpus:

1. CC Coordinating conjunction
2. CD Cardinal number
3. DT Determiner
4. EX Existential there
5. FW Foreign word
6. IN Preposition or subordinating conjunction
7. JJ Adjective
8. JJR Adjective, comparative
9. JJS Adjective, superlative
10. LS List item marker
11. MD Modal

12. NN Noun, singular or mass
13. NNS Noun, plural
14. NP Proper noun, singular
15. NPS Proper noun, plural
16. PDT Predeterminer
17. POS Possessive ending
18. PP Personal pronoun
19. PP$ Possessive pronoun
20. RB Adverb
21. RBR Adverb, comparative
22. RBS Adverb, superlative
23. RP Particle
24. SYM Symbol
25. TO to
26. UH Interjection
27. VB Verb, base form
28. VBD Verb, past tense
29. VBG Verb, gerund or present participle
30. VBN Verb, past participle
31. VBP Verb, non-3rd person singular present
32. VBZ Verb, 3rd person singular present
33. WDT Wh-determiner
34. WP Wh-pronoun
35. WP$ Possessive wh-pronoun
36. WRB Wh-adverb

3.3.4. Selection of Items to Be Analyzed

This study investigates two features of both corpora; grammatical and lexical comparison. Other language elements like speech acts, discourse features, phraseology etc were not dealt with in this study.

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 and Modals as the grammar points covered on each chapter.

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. Initially, a few sample queries were conducted to make sure the program is running smoothly. Data collection process and tools used are performed in three steps:

1. First, the spoken part of British National Corpus was analyzed regarding to tenses and modals and their verb usages 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, Log-likelihood values and collocation strengths data extracted from both corpora were compared, and similarities and differences are shown in detail for each grammar and vocabulary item.

3.4. Data Analysis Types

In this section, data analysis tools and methods used for the study are introduced and data analysis steps are explained.

3.4.1. Log-Likelihood Value Analysis

The easiest approach to working with raw language corpus data is simply counting the occurrences of the search item. A frequency list is simply a list of all types in a corpus with the number of occurrences of each type.

Language with its infinity is generally impossible to conquer. Therefore, a word frequency list can be regarded as the representation of language through numbers so that researchers can get the visualization of language and generate likely insights into language. In addition, frequency lists prove that ‘linguistics is the scientific study of language’ since they may guarantee objectivity and exactness in language studies. That

is to say, they allow the linguists to get on the field. For this reason, as most corpus based studies do, this study produces word frequency lists of both corpora and compares them.

However, 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.

3.4.2. Mutual Information and T-Score

To measure the statistical significance of two words or language items regarding their co-occurrence, the researcher adopted the use of two mostly used co-occurrence measure scores: Mutual Information (MI) and T-score. Both scores are used because each has certain weaknesses that the other complements, i.e. the scores are interdependent to judge the association strength of two words.

Mutual information scores are the most commonly used measure of collocation, because they are easy to interpret: they simply measure the strength of relationship between a target word and a possible collocate.

Very generally, an MI-score shows the strength of a collocation. In statistical terms, this is a measure of the strength of association between words x and y. Mutual Information score is a measure of how strongly two words seem to associate in a corpus, based on the independent relative frequency of two words. The score shows the extent to which observed frequency of co-occurrence differs from what we would expect. Hindle (1994) explains that "[w]e use mutual information to identify relations that occur more often than chance" (p.123).

A score around 0 indicates that there is no relationship between words. Then, any score greater than 0 confirms that the chance of two words are occurring together is

increasing. Therefore, the higher is the score, the stronger is the collocation strength. Collocate pairs which have high frequency tend to have high MI-score. However, MI score does not work well with very low frequencies - the T-score provides a way of getting away from this problem as it also take frequencies into account.

The T-score is a second statistic used with collocations. It is not a measure of the strength of association but the confidence with which we can claim that there is an association. With T-score we can be sure that the collocation pairs are not happening to be together by chance. MI is more likely to give high scores to totally fixed phrases, whereas T-score will yield significant collocates that occur relatively frequently. In most cases, T-score is the most reliable measurement, since it also takes the corpus size into account, while MI score doesn't.

The main differences between MI-score and T-score are:

- MI measures the strength of a collocation; T-score, on the other hand, measures the certainty of collocation.
- MI-score is not dependent on the size of corpus; T-score, however, is computed using the corpus size.
- MI scores can be compared across different sizes of corpora, but T-score cannot be compared across corpora.
- MI generally gives information about lexical behavior of a word, whereas T-score informs about grammatical behavior of a word.

If two words occur together a lot, they form a collocation; however, if two words co-occur a lot in a corpus, it does not mean that they are collocations. Therefore, they need to be verified by MI and T-score results.

As mentioned above, both of them have their strengths and weaknesses, and thus, need to be carefully interpreted. Deducing results just by looking at one of them will not produce reliable interpretations. In short, they close the deficits of each other, and thus both of them are exploited in this study. If a collocate appears in the top of

frequency count, MI and T-score lists, it can be clearly assumed as a reliable and a rock-solid collocate.

CHAPTER 4

RESULTS

This chapter presents the findings acquired from the statistical analyses and the comparison of them mentioned in Chapter 3. Language items defined were compared to ones extracted from TEFL CC, and were evaluated. The BNC representation of Tenses and Modals was shown and analyzed by looking at the frequency counts, relative frequencies and their immediate collocates. Thus, a clear picture of these specific items was illustrated in authentic interactions.

4.1. Data Extraction and Findings

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 if it is an 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? So, the researcher had to define a precise formula for Sketch Engine to look for the intended structure. See Appendix 1 for data extraction formula devised to search the items in BNC-Spoken and TEFL CC.

4.2. Verb Tenses in Positive Form

By using the formula in Appendix 1, 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 4.1

Table 4.1 Positive Tense Frequeny Distribution in BNC-Spoken

BNC-Spoken Positive Tenses	Overall Frequency 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

By looking at the table above, 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.

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 4.2.

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 4.2 Positive Tense Frequency Distribution in TEFL CC

TEFL CC Positive	Overall Frequency 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 4.3 shows the frequencies normalized to per million words.

Table 4.3 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	-

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.

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 4.3 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.

4.3. 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 4.4 shows the frequencies and normalized scores of negative tensed sentences.

Table 4.4 Negative Tenses In BNC-Spoken

BNC-Spoken Negative	Overall Frequency 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. Table 4.6 shows the negated tense distribution of all tenses in percentages.

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 4.5 shows the normalized frequencies of negatives and contracted negatives extracted from BNC-Spoken corpus.

Table 4.5 Contracted and Full Negative in Bnc Spoken

BNC-Spoken Contracted Negative	Contracted Negative N-pmw	Full Negative 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. The percent distribution of all negated sentences within themselves is illustrated in Table 4.6.

Table 4.6 Negatives Distribution in BNC-Spoken

Negative	Contracted Negative %	Negative %	Negative all %
Simple Present	61	26	56
Present Continuous	0	50	6
Simple Past	17	4	16
Past Continuous	1	1	1
Future Tense	7	7	7
Future Continuous	0	0	0
Future Perfect	0	1	0
Future Perfect Continuous	0	0	0
Present Perfect	12	11	12

Present Perfect Continuous	0	0	0
Past Perfect	2	1	2
Past Perfect Continuous	0	0	0

That contracted negatives are more common than others has an exception as can be observed from Tables 4.5 and 4.6. 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.

Table 4.6 shows its overall distribution within all contracted negatives. Although Simple Present Tense and Simple Past Tense has a percentage of 61 and 17 contracted forms, Present Continuous Tense has %0 similar to Past Perfect Continuous, Present Perfect Continuous, Future Perfect, Future Perfect Continuous and Future Continuous.

On the other hand, it has the highest percentage in full negative sentences with % 50. It is interesting because it is more common than Simple Present and Simple Past full negatives. Actually, it is the most common full negative structure and comprises half of the all tensed full negatives.

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 4.7

Table 4.7 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 is 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 4.8

Table 4.8 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 4.8 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 are more marked in BNC-Spoken, are clearly underused. Six other tenses do not have any representation of full negatives at all TEFL CC.

Table 4.8 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.

4.4. 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 4.9 shows the overall frequencies and normalized scores of all tenses in BNC-Spoken.

Table 4.9 Tenses in Total in BNC-Spoken

BNC-Spoken Tenses Total	Overall Frequency f(x)	N-pmw
Simple Present	489.73	41399,07
Present Continuous	51.04	4315,11
Simple Past	181.14	15312,62
Past Continuous	16.14	1364,71
Future Tense	57.10	4827,04
Future Continuous	1.379	116,57
Future Perfect	280	23,66
Future Perfect Continuous	0	0
Present Perfect	77.96	6590,33
Present Perfect Continuous	2.23	189,10
Past Perfect	8.79	743,47
Past Perfect Continuous	316	26,712

The distribution and percentages of all tensed verbs are shown in Figure 4.1. 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 4.1

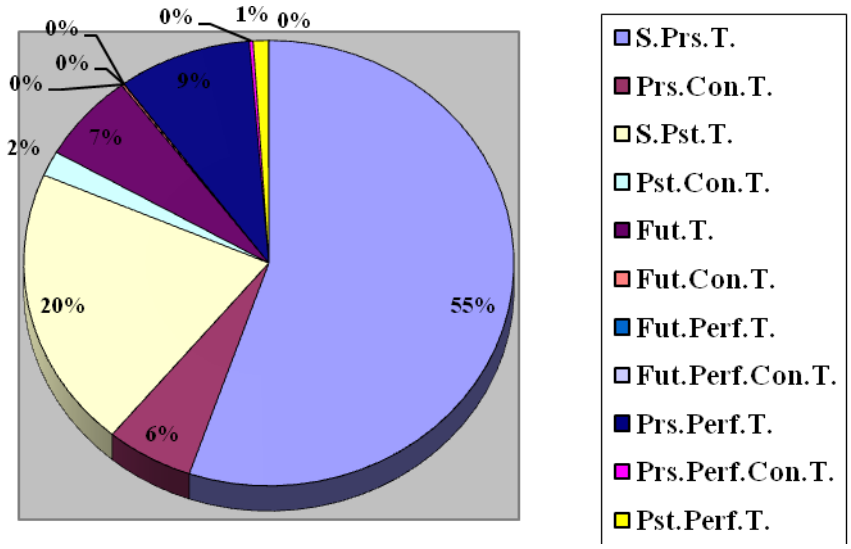

Figure 4.1 Distribution of all tensed sentences in BNC-Spoken

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 4.10.

Table 4.10 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 4.11. 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 4.11 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 follow 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 are compared to the findings from TEFL CC. Table 4.12 shows the normalized scores of both corpora on distribution of all tensed structures

Table 4.12 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 4.12, 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 4.12 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.

Tense/aspect distribution scores may give us a more general idea on the use of verb tenses. Table 4.13 indicates the Tense distribution of both corpora. TEFL CC seems to be underrepresented in terms of Present and Future Tenses while it is overrepresented in Past Tenses.

Table 4.13 Tense N-Pmw-Values Compared

Tense	BNC-Spoken N-pmw	TEFL CC N-pmw
Present	52493,62	41503,47
Past	17447,52	20491,67
Future	4967,29	4618,46

In terms of aspect, TEFL CC is far from typical scores. Simple aspect in BNC-Spoken with 61538 pmw score is higher than its course book equivalent with 55747. The same thing is applicable to Perfect and Perfect Continuous aspects, too. However, Progressive aspect is overused with a score of 6103 pmw in TEFL CC. The results and comparison are shown in Table 4.14.

Table 4.14 Aspect N-Pmw Values Compared

Aspect	BNC-Spoken N-pmw	TEFL CC N-pmw
Simple	61538,74	55747,38
Progressive	5796,39	6103,63
Perfect	7357,48	4743,79
Perfect Continuous	215,81	18,79

Simple Present, Future Tense, Future Continuous, Present Perfect and Present Perfect Continuous tenses are clearly underrepresented in TEFL CC compared to BNC-Spoken. Present Continuous, Simple Past, Past Perfect and Future Perfect tenses, conversely, are underrepresented in TEFL CC. Past Continuous and Past Perfect Continuous tenses have low log-likelihood significance values.

4.5. Verb Usage in Tenses

It is not always reliable to determine the general characteristics of a language item just by looking at its frequency and distribution within a corpus. The researcher needs more data to accept it as a de facto feature of authentic language. As explained in Chapter 3, the environment an item occurs in presents us with a very valuable data regarding its use. Therefore, collocation analyses of all verb tenses will show how verbs behave in context; that is, which tenses use which words more than others and their significance will get us more into understanding the target language behavior. As a result, in this section, the verb distribution of each verb tense will be explored and their significance level using T-score and Mutual Information scores will be shown.

4.5.1. Simple Present Tense

Simple Present Tense has the highest percentage among tensed structures. As shown in previous section, it has a score of 41,399 per million words, which is huge amount compared to others. Therefore, it seems impossible to identify all verbs used in Simple Present Tense. The BNC corpus data shows that more than 5,000 words occur less than 10 times in all simple present tense sentences. For this reason, it is most logical to consider the most used verbs again.

By accepting 100 per million words as a critical limit for the most used verbs, the following verbs in Table 4.15 were extracted from the BNC. As can be deduced from the list, the researcher did not consider *do, does, have and has* because they are by far the most used ones and they have functional roles in sentences.

In addition, the table lists the lemmatized list of the verbs used in Simple Present Tense. That is to say, all 3-rd person-singular conjugated verbs were also counted. Comparing the normalized score of top 20 most used verbs in Simple Present Tense, it is clearly visible that most of the verbs in BNC-Spoken are not properly represented in TEFL CC. While the verb know has 4299 pmw score in BNC-Spoken, it has 695 pmw in TEFL CC. This number is relatively small in such a big corpus. There are only a few verbs which are almost evenly distributed in TEFL CC. The verbs *suppose, say, mean and come* are quite common in BNC-Spoken. However, in TEFL CC, they are far below the ideal frequencies. Table 4.15 shows the log-likelihood comparison of Simple Present Tense verbs

Table 4.15 Simple Present Tense-Verb Distribution Comparison

Top 20 Verbs in Present Tense	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
know	4.299,13	695,58	739,29	-
think	2.953,61	1.384,91	164,12	-
mean	1.979,44	256,92	379,13	-
say	1.448,74	407,32	165,83	-
go	1.431,07	1.146,78	9,56	-
get	1.353,89	520,12	106,20	-
see	1.265,97	338,39	152,10	-
want	1.323,37	1.353,57	0,11	+
come	1.057,43	313,32	114,73	-
look	862,83	332,12	67,48	-
put	688,02	68,93	145,78	-
need	577,02	739,45	6,60	+
thank	506,27	31,33	122,80	-
take	486,74	419,86	1,52	-
make	414,72	438,65	0,21	+
give	391,22	194,26	19,25	-
like	473,72	1.021,45	74,33	+
tell	271,35	75,19	31,50	-
seem	251,65	112,79	15,27	-

suppose	236,4410	43,8660	37,54	-

Table 4.15 summarizes and shows the underused and overused verbs in Simple Present Tense. The verbs *want, need and make* seem to be similar to the distribution in BNC-Spoken with small and positive log-likelihood scores. The value of verb *take* is also too low to be considered significant. Apart from those four verbs, other 16 verbs have a huge negative log-likelihood value, and show that although Simple Present Tense is the most used tense in course books, the verb distribution is not similar to natural language.

The most frequent 20 verbs in Simple Present Tense in BNC-Spoken is not represented sufficiently in TEFL CC. Four of the verbs have distribution similar to BNC-Spoken, but others have diverse log-likelihood values compared to BNC-Spoken.

4.5.2. Present Continuous Tense

As for Present Continuous tense comparison, the same technique is applied. The most used verbs in Present Continuous from in BNC-Spoken were extracted and then normalized to per million words for comparison. Table 4.16 indicates normalized score of the V+ing forms in both corpora.

Table 4.16 Present Continuous Tense-Verb Distribution Comparison

Top 20 Verbs in Present Continuous Tense	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
going	1559,56	1554,10	0,00	-
doing	268,81	93,99	24,03	-
saying	163,23	6,26	43,26	-
getting	153,76	106,53	2,57	-
talking	135,76	125,33	0,13	-
coming	126,71	56,39	7,79	-
being	122,82	31,33	15,39	-
trying	105,49	75,19	1,53	-
looking	104,82	81,46	0,89	-
having	80,72	175,46	13,00	+
working	61,20	62,66	0,01	+
taking	54,69	50,13	0,06	-
happening	46,15	12,53	5,46	-

making	49,79	43,86	0,12	-
thinking	39,81	31,33	0,31	-
putting	31,86	25,06	0,25	-
asking	29,75	6,26	4,34	-
using	30,43	12,53	2,14	-
listening	24,59	25,06	0,00	+
paying	27,21	0	-	-

As opposed to great differences of verb usages in Simple Present Tense, Present Continuous variation is not remarkable in both corpora. Almost all verbs seem to be equally used in both corpora. The normalized scores are very close to each other in most of the verbs, and those who aren't do not show great divergences. Log-likelihood comparison will present a more accurate picture of the subject.

There are only three verbs which have extraordinary values. These are *having, doing and saying*. *Doing and saying* are used in small frequencies in TEFL CC compared to BNC-Spoken, with 24,03 and 43,26 log-likelihood values. They have the highest divergence score from the reference corpus. On the other hand, the verb *having* is overly used in TEFL CC with 13,00 log-likelihood value, which is the highest positive score on the table.

Present Continuous Tense verbs in TEFL CC seem to be more or less reflecting the everyday language in verb choice. Event the discrepancies are not too divergent to pose a problem in over all use.

4.5.3. Simple Past Tense

Having the second place in total tense frequencies, Simple Past Tense needs a closer look. Interestingly, there are many more overused items in TEFL CC compared to other tenses. Although verbs are generally underused compared to authentic language, many verbs used in Past Tense forms in TEFL CC are visibly higher. This can be observed in normalized scores, too. The verb *want* is more common in course books with 357 pmw score against 294 pmw in BNC-Spoken. Table 4.17 shows the normalized frequencies of Past Tense verbs.

Table 4.17 Simple Past Tense Verb Distribution Comparison

Top 20 Verbs in Simple Past Tense	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
say	2.147,15	965,05	129,57	-
get	1.074,17	651,72	30,54	-
go	811,86	526,39	18,10	-
think	666,04	131,59	101,55	-
come	423,76	350,92	2,10	-
want	294,17	357,19	1,99	+
take	250,81	325,86	3,22	+
tell	197,72	388,52	22,48	+
make	185,04	263,19	4,58	+
start	179,04	432,39	40,06	+
see	196,79	388,52	22,76	+
know	246,24	125,33	11,45	-
give	132,88	181,73	2,53	+
look	130,77	181,73	2,78	+
buy	120,29	62,66	5,29	-
ask	116,48	657,98	185,53	+
happen	106,68	169,19	4,88	+
find	103,13	181,73	7,65	+
leave	96,79	181,73	9,27	+
feel	86,81	156,66	7,10	+

The Log-likelihood values after the comparing the corpora present more reliable insights on Past Tense distribution between corpora. With 129,57 log-likelihood score the verb *say* is the most underused item. It is followed by *think* with 101,05 score. Two of the most frequent verbs in English language – *get and go*- are also underrepresented in TEFL CC with 30,54 and 18,10 log-likelihood values in order. The last significant underused item is know with 11,45 log-likelihood value.

Simple Past Tense verbs in course books are generally overused. 13 verbs out of 20 most used Past Tense verbs have positive log-likelihood values. The most prominent of them is the verb *ask* with 185,53 log-likelihood value. It may be because there is distinct unit in course books on Reported Speech, in which the verb *ask* is used a lot. The verbs *start, see, tell, leave and find* are other overused verbs on the table. There are

a few more overused items with small Log-likelihood values; therefore, they are not taken into consideration.

Analyzing the log-likelihood comparisons of both corpora, it is deduced that half of the most frequent 20 verbs in Past Tense seem to be represented in TEFL CC; however, the other half has significantly high divergent log-likelihood scores. As a result, Simple Past Tense verb forms have to be revised to comply with BNC-Spoken.

4.5.4. Past Continuous Tense

Course book representation of Past Continuous Tense is quite low. 6 verbs are not even present in TEFL CC. *Saying, thinking, telling, taking, asking and happening* have no samples in the whole corpus. With 17 occurrences, the verb *going* is the most frequent verb in the corpus. There is not enough evidence in TEFL CC to make comparisons. Table 4.18 indicates the log-likelihood values of both corpora in Past Continuous forms.

Table 4.18 Past Continuous Tense Verb Distribution Comparison

Top 20 Verbs in Past Continuous Tense	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
going	329,43	106,53	32,43	-
saying	87,15	0,00	-	-
talking	61,96	31,33	2,92	-
looking	42,44	31,33	0,50	-
coming	39,56	18,80	2,14	-
getting	37,96	25,07	0,79	-
thinking	36,10	0,00	-	-
trying	30,94	31,33	0,00	+
working	31,11	31,33	0,00	+
telling	20,88	0,00	-	-
sitting	14,54	37,60	3,96	+
making	14,29	18,80	0,20	+
playing	13,27	62,67	14,90	+
taking	12,51	0,00	-	-
asking	10,48	0,00	-	-
walking	9,97	37,60	6,95	+
running	10,06	50,13	12,58	+
happening	9,30	0,00	-	-
watching	9,13	50,13	13,78	+
living	8,20	18,80	1,57	+

By looking at the log-likelihood comparison values, the verbs *watching, running and playing* are much more frequent than their BNC-Spoken counterparts. However, the verbs *happening, asking, taking, telling, thinking and saying* have no occurrence in TEFL CC. The only underused item in corpus is the verb going with 32,43 minus log-likelihood value. Other verbs seem to be equally distributed in both corpora with small log-likelihood scores.

There is not enough evidence on TEFL CC regarding the use of Past Continuous Tense. However, with this little data at hand, we can suppose that only %50 of all Past Continuous forms are reflected in TEFL CC.

4.5.5. Future Tense

Future Tense is one of the most underused tenses in TEFL CC. Almost all 20 frequent words have less relative frequency compared to BNC-Spoken Future verbs. Table 4.19 indicates the normalized frequencies and log-likelihood values of both corpora regarding most frequent 20 Future Tense forms.

Table 4.19 Future Tense Verb Distribution Comparison

Top 20 Verbs in Future Tense	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
get	181,57	81,46	10,99	-
do	163,15	68,93	11,00	-
go	160,52	144,13	0,27	-
see	100,67	43,86	6,43	-
give	91,46	31,33	8,39	-
come	89,69	75,19	0,39	-
take	86,81	43,86	4,11	-
tell	65,09	37,59	2,16	-
put	53,42	12,53	7,18	-
say	55,62	6,26	11,29	-
find	49,02	37,59	0,46	-
make	42,94	43,86	0,00	+
try	33,64	12,53	2,76	-
need	32,96	18,79	1,14	-
leave	27,72	12,53	1,66	-
know	31,78	0,00	-	-
ask	23,24	0,00	-	-
look	21,97	12,53	0,76	-
let	26,29	6,26	3,49	-
keep	19,52	12,53	0,45	-

Considering only the normalized scores or only the log-likelihood values is not always reliable. Therefore, both of them need to be taken into consideration while comparing corpora. Eight verbs are distributed evenly within both of them as their log-likelihood values are quite low and close to 0. It doesn't matter whether scores are positive or negative as long as they are close to 0.

Two verbs – *know and ask* – have no occurrence at all in TEFL CC. The rest of the future verbs are all underrepresented with negative and high log-likelihood scores. 8 items out of 20 most frequent verbs comply with the authentic BNC-Spoken items. The other remaining verbs are underused in TEFL CC with high log-likelihood values.

4.5.6. Future Continuous Tense

Future Continuous Tense sentences taken from TEFL CC are quite few in numbers. There are only 6 sentences in Future Continuous forms. Therefore, their distribution among top 20-most-frequent verbs will be low in accordance. Most of the verbs – except *taking and working* – are not used in TEFL CC. Table 4.20 shows normalized scores and log-likelihood values. As can be deduced from the table, since there are no occurrences in TEFL CC, the results cannot be computed and compared against BNC-Spoken.

Table 4.20 Future Continuous Tense Verb Distribution Comparison

Top 20 Verbs in Future Continuous Tense	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
going	11,91	0,00	-	-
doing	8,19	0,00	-	-
looking	6,25	0,00	-	-
coming	6,76	0,00	-	-
talking	5,24	0,00	-	-
getting	5,83	0,00	-	-
having	3,12	6,26	0,38	+
making	2,53	0,00	-	-
taking	2,53	6,26	0,61	+
working	2,45	12,53	3,22	+
saying	2,02	0,00	-	-
paying	2,45	0,00	-	-
giving	1,77	0,00	-	-

seeing	1,52	0,00	-	-
using	1,77	0,00	-	-
putting	1,52	0,00	-	-
sitting	1,43	0,00	-	-
asking	1,18	0,00	-	-
telling	1,09	0,00	-	-
thinking	1,09	0,00	-	-

Future Continuous tense verb forms are relatively underused in TEFL CC. With only three verbs and 6 occurrences, the course books do not reflect authentic language Future Continuous usages.

4.5.7. Future Perfect Tense

With 28 distinct occurrences in TEFL CC, Future perfect Tense is said to be not given enough importance. In addition, there is not a single occurrence of top 20 most frequent Future Perfect verbs extracted from BNC-Spoken; thus, a comparison cannot be made. With 259 raw frequencies, BNC-spoken seems to place much more emphasis on it than TEFL CC.

Future Perfect Tense is, just like, Future Continuous Tense, has very little occurrence in TEFL CC compared to our reference corpus, BNC-Spoken. 28 future perfect forms are not even within the most frequent 20 verbs.

4.5.8. Present Perfect Tense

Present Perfect Tense usage in TEFL CC is not as problematic as Future Tenses since there enough occurrences in both corpora to make deductions and comparisons. Table 4.21 summarizes the frequency distributions and comparison results among 20 most frequent verb forms from BNC-Spoken.

Table 4.21 Present Perfect Tense Verb Distribution Comparison

Top 20 Verbs in Present Perfect Tense	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
got	2.466,27	714,38	273,97	-
had	320,29	81,46	40,25	-
done	260,02	75,19	28,93	-
gone	171,60	18,79	35,21	-
seen	138,29	62,66	8,22	-
come	88,42	50,13	3,11	-
put	75,15	6,26	16,88	-
made	72,69	75,19	0,01	+
said	80,89	12,53	14,24	-
heard	63,99	25,06	4,88	-
taken	51,56	50,13	0,01	-
given	48,43	25,06	2,17	-
happened	43,19	12,53	4,79	-
lost	38,29	12,53	3,72	-
finished	37,02	31,33	0,15	-
changed	34,32	62,66	2,95	+
written	25,95	18,79	0,34	-
found	28,74	12,53	1,83	-
worked	26,62	25,06	0,01	-
told	26,45	12,53	1,44	-

With 273,97 negative log-likelihood value, the verb *get* is clearly the most underused item in TEFL CC. Though being the most used verbs in English, the verb *have* is another item that was not given much emphasis in the course books with 40,25 log-likelihood value. The same thing is true for the verbs *go, do and put*, with 35, 28 and 16 negative log likelihood values. Some other verbs – *see, come and hear, lose and happen* – are also underused but with low negative log-likelihood scores. The verbs which are thought to be reflected in enough numbers in TEFL CC are shown on the table. Two of them are overused compared to BNC-Spoken; however, the difference is not significant enough to take into consideration. According to table, 9 verbs out of 20 comply with BNC-Spoken counterparts in frequency and log-likelihood values.

Some most frequent Present Perfect verbs are not represented evenly in TEFL CC; and therefore, the distribution of Perfect Tense among verb forms should be

revised. 9 verbs are in compliance with authentic language types. The remaining 11 should be customized according to BNC-Spoken examples.

4.5.9. Present Perfect Continuous

Present Perfect Continuous Tense is among the tenses on which TEFL CC does not present enough data to make comparisons. There are 79 Present Perfect Continuous occurrences in TEFL CC; but, they lack using most frequent verbs. Most of the verb forms are used one or two times at most. The verb *work* is used 15 times with 46,43 log-likelihood value compared to BNC-Spoken. This value states that the verb is too much overused with 93 against 7 in BNC-Spoken. Table 4.22 indicates the Present Perfect Continuous verb usage distribution.

Table 4.22 Present Perfect Continuous Tense Verb Distribution Comparison

Top 20 Verbs in Present Perfect Continuous Tense	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
doing	20,79	6,26	2,22	-
going	12,84	6,26	0,66	-
talking	11,07	0,00	-	-
working	7,43	93,99	46,43	+
trying	7,01	25,06	4,33	+
saying	6,59	0,0	-	-
looking	5,15	6,26	0,04	+
getting	3,55	0,00	-	-
having	3,29	0,00	-	-
waiting	3,21	6,26	0,36	+
taking	3,46	18,79	5,11	+
using	3,04	18,79	5,73	+
sitting	2,78	0,00	-	-
coming	2,70	0,00	-	-
happening	2,53	0,00	-	-
telling	2,28	0,00	-	-
running	2,28	6,26	0,73	+
playing	2,36	6,26	0,69	+
thinking	2,28	0,00	-	-
living	2,02	18,79	7,72	+

9 verb forms out of 20 were not even used in TEFL CC in addition to 4 verbs which are significantly overused. Only 5 of the most frequent 20 verbs in Present Perfect Continuous seem to be used in equal share as BNC-Spoken.

The Present Perfect Continuous verb forms in TEFL CC are not in line with their BNC-Spoken counterparts. Therefore, the frequencies and verb distributions of the verb forms need customizing while preparing the course book.

4.5.10. Past Perfect

Past Perfect Tense lacks representing the most frequent 20 verb forms. Out of 127 verb frequency in TEFL CC, only 14 were used with 37 times in course books. That means the verbs do not cluster and course books have many different verb forms with low frequency. Table 4.23 shows the most frequent verb distribution with log-likelihood values.

Table 4.23 Past Perfect Tense Verb Distribution Comparison

Top 20 Verbs in Past Perfect Tense	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
got	109,13	18,79	18,12	-
gone	37,53	37,59	0,00	+
done	31,02	12,53	2,25	-
had	29,16	18,79	0,67	-
seen	15,38	37,59	3,57	+
said	13,60	0,00	-	-
come	11,15	12,53	0,03	+
left	9,12	18,79	1,23	+
made	9,80	18,79	1,02	+
taken	9,04	18,79	1,25	+
finished	8,79	25,06	3,12	+
happened	7,94	0,00	-	-
forgotten	6,67	0,00	-	-
lost	6,34	6,26	0,00	-
given	6,17	6,26	0,00	+
heard	6,42	0,00	-	-
bought	4,90	6,26	0,05	+
died	4,48	0,00	-	-
worked	4,48	0,00	-	-
thought	5,91	6,26	0,00	+

According to table, the verb '*get'* is the most underused item with 18,12 log-likelihood score and 18 normalized frequency counts in TEFL CC against 109 counts in BNC-Spoken. Another underused item is '*do*' with 2,25 negative log-likelihood value. The verbs *see and finish* are overused with 3,57 and 3,12 log-likelihood values. 9 other verbs are overused, but with little significance scores, as can be observed from the table. Apart from these, 6 items were never used in TEFL CC.

Past Perfect forms in TEFL CC are few in number. Although there are about 170 occurrences of Past Perfect verbs, only ten verbs are represented as in authentic language.

4.5.11. Past Perfect Continuous

In TEFL CC, there is not enough data to compare with BNC-Spoken. There are only 3 occurrences of Past Perfect Continuous items and none of them are represented in BNC-Spoken. Since there are no occurrences of Past Perfect Continuous Tense in TEFL CC, a comparison between two corpora is not possible. Past Perfect Continuous items should be used more to be compatible with the BNC-Spoken.

4.6. Tense Collocation Strengths

In the previous section, 20 most frequent R1 collocates of Tense forms were explored and their relative frequency and log-likelihood values after comparison of both corpora were showed in detail. Log-likelihood value, as explained in Chapter 3, is a useful measure to compare the use of a specific word or structure among different corpora. In a way, it uses the raw frequencies of words and overall corpus size; and cross tabulates these with the reference corpus.

However, it is not always reliable just to compare the raw frequencies and their relative values. There may always be a chance factor to take into consideration. We cannot be %100 sure that a specific word is one of the main collocates of another word or they happen to be together just by chance. Therefore, we need other tools to strictly determine their co-occurrence together. As two of the most used association measures in corpus studies, T-score and MI values are best candidates to look for collocation strengths of the verbs explored in previous section. If two words occur together a lot,

they form a collocation; however, if two words co-occur a lot in a corpus, it does not mean that they are collocations. Therefore, they need to be verified by MI and T-score results. Their characteristics and features are explained in detail in Chapter 3.

In this chapter, the verb forms extracted from both corpora will be analyzed and compared with each other; and therefore, their co-occurrence will be found and the order they should be presented in TEFL CC will be identified. Table 4.24 shows the association score of 20 most frequent verbs in all tenses.

4.6.1. Simple Present Tense

All the verbs given in previous section were ordered according to their frequency in BNC-Spoken and their comparison was done on frequency level. The verbs are ordered in frequency order again in Table 4.24. However, their MI and T-scores are incompatible with frequency order. For instance, although the verb 'know' is the most used one in Simple Tense, its T-score and MI values are way below the scores of the verb '*think'*. The same thing is true for the verbs '*want', 'need' and 'thank'*. Although the verb '*thank'* is in the last ten of the most frequent verbs, T-score and MI values carry it to the fourth place in terms of its strength as a collocate. The scores in Table 4.24 indicates that these verbs are stricter collocates of Simple Present Tense than their frequencies show. Another observation is with the verb '*like'* which is more frequent than the last three verbs; however, it is on the last place to be used as a collocate as can be observed in Table 4.24.

In course book side, the TEFL CC seems to have the verb '*think'* on the first place. In addition, the verb '*get'* as a collocate has the same rank with BNC-Spoken. A few other verbs have similar – not exact- usages. The verbs '*suppose, tell, look and see'* are roughly in the same order as collocates. However, the remaining 14 verbs have completely different ranks. The verb '*thank'* is one of the strongest collocates in BNC-Spoken, it is on the last place in TEFL CC.

4.6.2. Present Continuous Tense

The frequency rank and collocation ranks are never in the same order. This can be observed in Present Continuous verb forms, too. There is only one verb – *go* – which

is on the top of the list. All other verbs have different ranks compared to their frequencies. It is impossible to expect a one-to-one match; therefore, items which are roughly on the same rank on the list are left out. There are two significant verbs –*be and have* – which have little T-scores compared to their frequencies. Although they are the most frequent ones in terms of frequency, they are the weakest verbs as collocates and have the least values in the list.

As for TEFL CC, the table presents us with valuable data. Most of the verbs have no similarity to BNC-Spoken items. There are 6 verbs – *get, come, do, happen, say and pay* – which have low collocation strengths compared to BNC-Spoken. On the other hand, 5 items have high collocation ranks. Only two verbs are exactly on the same rank; and the remainings are on different ranks but with not too much significance. In addition, as shown in previous paragraph, although the verbs *be and have* are the weakest collocates in BNC-Spoken, they are on high ranks in TEFL CC.

Table 4.24 Association Measures of the Most Frequent Verb Tenses in Both Corpora

Simple Present					Present Continuous					Simple Past				
	BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC	
	T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score
know	45.150	4.912	8.365	3.825	going	98.836	6.628	15.419	7.211	say	28.246	4.785	12.164	5.659
think	52.782	5.618	13.095	4.476	doing	49.918	3.652	3.643	5.252	get	18.344	4.178	9.671	4.560
mean	39.288	5.837	4.992	4.142	saying	37.619	3.586	0.973	5.217	go	16.727	3.851	8.883	5.714
say	28.725	3.047	5.448	4.761	getting	38.778	4.169	4.085	6.760	think	17.467	4.546	0	0
go	24.503	3.061	9.906	2.968	talking	38.504	6.453	4.428	6.665	come	14.815	4.124	7.208	5.714
get	25.753	3.259	6.147	2.761	coming	36.129	4.487	2.782	5.944	want	14.630	4.741	6.907	5.427
see	22.654	3.122	5.976	2.691	being	-12.151	-0.410	2.126	4.343	take	11.262	3.226	6.867	5.714
want	30.604	4.812	11.482	4.329	trying	34.035	5.648	3.427	6.529	tell	8.020	3.444	7.714	5.624
come	23.592	3.509	5.096	4.761	looking	32.817	4.472	3.566	6.503	make	11.107	2.973	6.156	4.697
look	20.192	3.353	4.920	2.941	having	7.490	0.447	5.243	6.777	start	9.934	4.502	8.137	5.613
put	19.106	3.932	2.548	4.761	working	23.648	3.499	2.716	4.652	see	8.759	2.601	7.471	5.714
need	25.438	4.266	9.274	4.365	taking	21.325	2.966	2.787	6.102	know	6.715	2.122	3.670	5.714
thank	33.627	7.798	0.875	0.830	happening	22.755	5.307	1.405	7.217	give	8.153	2.881	5.191	5.714
take	18.160	2.706	6.015	4.761	making	19.765	2.595	2.384	5.217	look	8.984	3.123	5.093	5.662
make	15.415	2.188	5.610	4.719	thinking	18.162	2.807	2.218	6.954	buy	4.494	3.305	2.911	5.077
give	15.247	2.697	3.336	4.761	putting	16.769	3.508	1.975	6.343	ask	9.998	4.292	9.991	5.620
like	9.398	1.620	7.734	2.436	asking	16.642	3.675	0.973	5.217	happen	8.485	4.673	5.079	5.469
tell	11.025	2.565	2.548	4.761	using	14.495	2.307	1.326	3.995	find	8.136	3.154	5.137	5.095
seem	15.636	3.583	2.344	4.539	listening	15.977	5.755	1.955	5.480	leave	5.782	2.856	5.148	4.506
suppose	10.142	4.584	2.534	4.568	paying	15.015	3.824	0	0	feel	8.981	3.974	4.593	5.589

Table 4.24 Association Measures of the Most Frequent Verb Tenses in Both Corpora

Past Continuous					Future Tense					Future Continuous				
	BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC	
	T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score
going	48.573	6.154	3.800	3.674	get	44.153	4.173	2.396	2.045	going	11.618	5.534	0	0
saying	30.632	4.442	0	0	do	38.002	2.728	0.054	0.026	doing	9.132	3.779	0	0
talking	26.895	6.972	2.170	5.084	go	40.834	3.882	3.156	1.966	looking	8.415	5.525	0	0
looking	21.602	4.802	2.114	4.191	see	32.662	3.589	1.803	1.922	coming	8.729	5.376	0	0
coming	20.660	4.443	1.525	3.062	give	30.855	3.953	1.889	2.686	talking	7.808	6.889	0	0
getting	19.694	3.824	1.650	2.514	come	30.241	3.703	2.942	3.145	getting	7.966	4.607	0	0
thinking	19.614	4.299	0	0	take	29.099	3.388	1.787	1.887	having	3.205	1.080	0	0
trying	18.706	5.492	2.163	4.940	tell	27.125	4.293	1.645	2.496	making	4.965	3.419	0	0
working	18.134	4.193	2.049	3.579	put	23.718	4.053	1.016	1.828	taking	5.055	3.698	0.987	6.315
telling	15.022	4.501	0	0	say	19.177	1.912	-0.268	-0.342	working	5.051	4.012	0.989	6.480
sitting	12.799	5.377	2.404	5.758	find	22.053	3.429	2.114	2.870	saying	4.034	2.502	0	0
making	10.585	2.428	1.407	7.669	make	17.489	2.110	1.815	1.949	paying	5.282	5.703	0	0
playing	11.990	4.538	3.070	5.091	try	18.419	3.682	1.016	1.828	giving	4.224	3.676	0	0
taking	10.087	2.526	0	0	need	17.149	2.902	0.872	1.382	seeing	3.662	2.868	0	0
asking	10.361	3.845	0	0	leave	16.206	3.195	1.019	1.840	using	4.143	3.382	0	0
walking	10.518	4.978	2.409	5.932	know	14.638	1.945	0	0	putting	4.031	4.326	0	0
running	10.297	4.156	2.800	6.644	ask	14.629	3.066	0	0	sitting	4.033	5.523	0	0
happening	10.044	4.562	0	0	look	12.512	2.010	1.185	1.663	asking	3.536	4.183	0	0
watching	10.078	5.048	2.772	5.656	let	16.755	4.280	0	0	telling	3.335	3.738	0	0
living	9.319	4.216	1.589	3.594	keep	13.445	3.069	1.026	1.866	thinking	3.068	2.746	0	0

Table 4.24 Association Measures Of the Most Frequent Verb Tenses in Both Corpora

Future Perfect					Present Perfect					Present Perfect Continuous				
	BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC	
	T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score
seen	5.006	5.783	0	0	got	169.996	7.639	9.941	5.993	doing	14.988	4.494	0.959	4.623
done	4.170	3.885	0	0	had	46.834	2.065	1.126	0.599	going	11.947	5.014	0	0
had	2.846	2.062	0	0	done	48.871	3.073	1.778	1.039	talking	11.375	7.341	0	0
noticed	3.305	8.154	0	0	gone	40.605	3.631	-0.026	-0.021	working	9.085	4.986	3.864	8.737
gone	3.141	4.238	0	0	seen	37.307	3.682	1.477	1.935	trying	8.987	6.208	1.993	8.122
told	2.575	5.229	0	0	come	29.378	3.401	2.114	2.869	saying	8.091	3.575	0	0
come	2.299	4.023	0.955	4.477	put	28.259	4.259	0.628	1.426	looking	7.492	4.618	0.984	5.980
finished	2.633	7.685	1.727	8.570	made	24.378	2.567	2.838	3.284	getting	5.806	3.263	0	0
changed	2.382	5.185	0	0	said	23.851	2.127	0.822	1.256	having	1.914	0.528	0	0
read	2.204	6.138	0	0	heard	26.645	4.986	1.884	4.114	waiting	6.097	6.508	0.993	7.173
heard	2.413	6.084	0	0	taken	20.170	2.381	2.241	2.708	taking	5.845	3.521	0	0
got	2.228	3.468	0	0	given	20.411	2.756	1.051	1.962	using	5.481	3.532	1.724	7.758
eaten	1.982	6.826	0	0	happened	21.654	4.570	1.274	3.330	sitting	5.645	5.852	0	0
gathered	1.994	8.301	0	0	lost	20.423	4.628	1.301	3.640	coming	5.131	3.426	0	0
achieved	1.707	6.138	0	0	finished	20.514	5.578	1.649	4.377	happening	5.360	5.545	0	0
worked	1.828	3.538	0	0	changed	17.938	3.188	3.073	5.146	telling	4.906	4.165	0	0
received	1.696	5.590	0	0	written	16.092	3.539	1.224	2.892	running	5.019	4.873	0.989	6.521
paid	1.670	4.814	0	0	found	14.839	2.357	0	0	playing	5.115	4.907	0.963	4.771
run	1.666	4.715	0.990	6.644	worked	12.508	1.760	1.522	2.065	thinking	4.620	3.173	0	0
given	1.550	3.253	0	0	told	14.875	2.639	1.083	2.094	living	4.752	5.058	1.716	6.758

Table 4.24 Association Measures Of the Most Frequent Verb Tenses in Both Corpora

Past Perfect				
	BNC-Spoken		TEFL CC	
	T-Score	MI Score	T-Score	MI Score
got	35.487	6.295	1.503	2.917
gone	20.281	4.656	1.296	1.990
done	17.014	3.160	-0.448	-0.535
had	13.064	1.758	0.692	0.736
seen	12.469	3.675	2.120	4.268
said	10.749	2.710	0	0
come	10.656	3.592	1.270	3.295
left	9.970	4.472	1.679	5.035
made	9.257	2.832	1.585	3.558
taken	9.208	3.034	1.579	3.496
finished	10.097	6.655	1.982	6.803
happened	9.446	5.282	0	0
forgotten	9.011	6.516	0	0
lost	8.423	5.188	0.960	4.651
given	7.427	2.935	0.873	2.973
heard	8.410	4.824	0	0
bought	7.353	4.859	0.904	3.376
died	7.051	4.988	0	0
worked	5.845	2.343	0	0
thought	7.021	2.636	0.779	2.176

4.6.3. Simple Past Tense

In BNC-Spoken Simple Past Tense forms collocation strength ranking is almost similar to their frequency list. Therefore, we can say that relative frequencies and association measures are in compliance in BNC-Spoken in terms of Past tense items. However, there are some verbs which have high and low collocation strengths. The verbs *'ask, look, feel and happen'* are more significant as collocations in BNC-Spoken compared to their normalized scores. On the other hand, 2 verbs – *know and tell* –are weak collocates in Past Tense sentences. The remaining verbs are more or less on the same level as collocates.

However, the same similarity cannot be observed in terms of TEFL CC. According to table, there are 5 verbs observable in TEFL CC which has similar collocation strengths as BNC-Spoken; and therefore, it can be deduced that these verbs are represented uniformly in TEFL CC. Nevertheless, other remaining 15 verbs differ considerably. 7 verbs including '*come, want and take'* have low collocation strengths while '*start, ask, tell and see'* have high scores compared to their BNC-Spoken counterparts. To sum up, it can be assumed that Simple Past verb items have mostly different co-occurrence values.

4.6.4. Past Continuous Tense

Like Simple Past tense items, verbs used in Past Continuous form is almost exactly the same as their normalized frequency ranks in previous chapter. Only one verb – *walk*- can be considered significant. It has more collocation strength than the frequency analysis. The other verbs are on the same rank on the list as their frequencies show.

On the other hand, TEFL CC collocation scores indicate a different ranking on the list. First of all, 6 items are not represented at all in course books. Therefore, it is not possible to do a collocation analysis for these verbs. 5 verbs – *look, work, get, talk and come*- have low co-occurrence values while '*play, run, watch, walk, sit and live'* have more strengths as collocations in course books. These verbs, as a result, seem not to be representing authentic language in Past Continuous verb usages. The remaining 3 items

comply with their BNC-Spoken counterparts. In general, we can say that TEFL CC has weak collocations scores of 20 most frequent verb forms.

4.6.5. Future Tense

Future Tense collocation scores have also similarities with normalized scores presented in previous section. Collocations scores of the most verbs are on the same rank with frequencies of those items. There is only one verb – *let* – which has more strength as collocation in BNC-Spoken.

The same inconsistency as in previous tenses in TEFL CC can be observed in Future forms, too. 3 verbs – *let, know and ask*- have no occurrence in course books and there are no scores for them on the table. An interesting observation is that 3 other verbs – *need, say and do* – cannot be considered as collocates in Future forms since T-scores and MI values are considerably low (0.872, -0.268 and 0.026). Although they have high normalized frequencies, their strengths as collocations are relatively low. As can be seen on the table, there are only a few verbs which have similar ranks in terms of collocation strengths. These are '*go, get, give, take, tell and try'*. The remainings are used more than their BNC-Spoken counterparts as collocations. The same level of inconsistency within two corpora is present in Future Tense, too.

4.6.6. Future Continuous Tense

Frequency listing and collocation strengths rankings regarding Future Continuous verbs are quite similar with each other in BNC-Spoken. Only the verbs – *pay and sit* – are strong collocations while '*have'* is the weak one in Future continuous forms. The remaining are consistent with collocation scores listings.

However, as discussed in the previous chapter, there is quite little occurrence of the most frequent verbs in TEFL CC and these are '*take and work*'. As a result, as can be seen on the table, the other items have no representations in course books, and collocation analysis cannot be done. It can be assumed that Future continuous forms are not represented in TEFL CC.

4.6.7. Future Perfect Tense

In BNC-Spoken, there is little data on the use of Future Perfect verb forms. As can be observed in previous chapter, the normalized scores of the verbs were very small in number and a comparison couldn't be done. In addition, TEFL CC corpus doesn't provide us with enough samples on Future perfect verbs. The comparison can be seen on Table 4.24. There are only 3 verb forms out of 20 represented in BNC-Spoken, with very small T-scores and they are not the collocates in Future Perfect sentences. As a result, it can be stated that Future Perfect Tense forms are not represented in TEFL CC and they are weak collocates compared to BNC-Spoken.

4.6.8. Present Perfect Tense

Unlike Future Perfect and Future Continuous, Present Perfect has enough data in both corpora. The frequency listing of the Present Perfect forms can be examined in previous section. The co-occurrence listing of the most frequent verbs with their association measures are shown on Table 4.24.

Frequency and collocation ranking of the most frequent verbs are quite similar to each other. Almost all items are in the same rank on both columns. There are only 2 items – finish and take- which have disparity in terms of collocation strength. While the verb finish is a strong candidate as a collocation with 20,154 T-score compared to frequency listing, the verb *take* is a weak collocate with 20,170 T-score.

As for TEFL CC, the similarity within frequency listing and collocation measures is not present for the verbs analyzed. The TEFL CC has 6 items which are similar to BNC-Spoken as collocation candidates. However, the other 16 items are quite different. For instance, 4 items – *say, put, go and find* – have small T-scores close to 0; and thus, seem to be underused as a collocate in TEFL CC. Among the most frequent verbs, *do, see and have* are weak collocates compared to BNC-Spoken. Interestingly, the Present Perfect Tense has more strong collocates than other course book tenses. 8 items are used as strong collocates while they have low ranking in BNC-Spoken. As for weak collocates –*do, see and have* - in TEFL CC, they are on the lower level of the list compared to BNC-Spoken.

4.6.9. Past Perfect Tense

The frequency listing of the most frequent verb forms used in BNC-Spoken and their co-occurrence listing is almost the same with each other. Only 2 verbs – *happen and finish* – have higher ranks in the list compared to frequency listing. The remaining verbs are quite similar ranking.

On the other hand, as explained in the previous section, Past Perfect verb forms are not many in number in both corpora; and it has a few items from most frequent verbs in TEFL CC. For instance, a comparison of the verbs '*say, happen, forget, hear, die and work*' is not possible since they are not represented in TEFL CC. In addition, 6 other verbs have very small T-scores close to 0.

As can be observed in Table 4.24, there is not even one item similar to BNC-Spoken regarding the use of Past Perfect items. The remaining 8 verbs are either strong or weak collocates in Past Perfect forms. It can be deduced from the scores that TEFL CC does not present enough data on the use of Past Perfect forms and they do not form strong collocates in the sentences.

4.7. Modals

Previous section shows similarities and differences between BNC-Spoken and TEFL CC regarding their frequencies, log-likelihood comparison, most frequent verbs used in Tensed sentences, and their collocation strengths. In this section, a similar methodology were used on the use and comparison of Modals. To extract Modal forms from BNC-Spoken and TEFL CC, some special formula were devised and used. See Appendix 2 that indicates the Sketch Engine modal structures extraction formula.

4.7.1. Modals in Positive Form

By using the formula in Appendix 2, frequencies of each sentence containing modals were extracted from both corpora. Later they were normalized to per million words since it will be impossible to make comparison with the TEFL CC only with over all frequencies of each structure. Normalized scores are computed by dividing the tense frequencies to the total number of words in the corpus and then multiplied by 1 million.

Table 4.25 shows the normalized scores and log-likelihood values of modals taken from both corpora.

Table 4.25 Modals in Positive Form Comparison

Modals Positive Comparison	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
can	3.052	3.892	33,35	+
may	376	708	36,56	+
might	667	213	66,57	-
will	4.300	3.534	22,91	-
could	1.354	990	17,04	-
would	3.374	1.874	125,68	-
should	881	1.260	22,58	+
must	487	884	40,67	+
shall	229	219	0,07	-
ought to	102	50	5,07	-
Need	6	0	-	-
Dare	8	0	-	-

A clear picture of how modals are treated in authentic language, that is BNC-Spoken, can be seen on Table 4.25. The distribution of modals in real language data shows us how authentic language makes use of them in daily life. Thus, the current picture of modal structures presents us with some useful data of authentic language.

On the table, it would come as no surprise that a clear overuse of modal structures *can and will* in positive forms with 4.300 pmw and 3.052 pmw occurrences is observable at the first sight. They have the largest shares in all modal sentences in BNC-Spoken. However, it is interesting to note that *will* is used much more often than *can*, a thing clearly overlooked by many people. Many textbooks and language teachers start teaching can before they teach *will*.

Would comes in the third place with 3.374 pmw occurrences in BNC-Spoken followed by could with 1.354 pmw and should 881 pmw occurrences in order. The remaining modals are *might, must, may, shall and ought to* in order. The modals *need and dare* have very little occurrences to take into consideration.

A short picture of modals in BNC-Spoken is presented above and their frequency ranking can be seen on the table. As for the TEFL CC representations of modal structures, the formula in Table 4.27 is used to extract them from Sketch Engine Corpus Query System. Table 4.25 also shows the normalized frequencies of modals represented in TEFL CC.

The figures on the table clearly illustrate the extent to which the frequencies affirmative modal verb forms in TEFL CC differ completely than in BNC-Spoken. First observation is that underuse of *will*, which has 3.534 pmw occurrences against 4.300 pmw in BNC-Spoken. This is clearly an underrepresentation of the item. Next, while *can* is on the second place in occurrence in authentic English, it has the top frequency in course book corpus with 3.892 pmw occurrence. Just like in natural language, *would* is on the third place in course book corpus too. However, its frequency is almost half of its BNC-Spoken counterpart with 1.874 pmw against 3.374 pmw occurrences.

As for other modals, the modal item *should* has 1.260 pmw occurrences in TEFL CC against 881 pmw. In this case, it is used more frequently than its BNC-Spoken counterpart. It is followed by *could* with 990 pmw and must with 884 pmw occurrences. The former is underused compared to BNC-Spoken samples. On the other hand, *must* is a relatively overused item with 884 pmw against 487 pmw occurrences.

A similar observation can be made on frequencies of *may and might*, which are on the last ranks of the BNC-Spoken modals. In TEFL CC orpus, *might* has 213 pmw occurrences; a number quite low compared to *might* with 667 pmw in natural language. Yet, *may* is excessively used in TEFL CC with 708 pmw.

As the least used items, *shall and ought to* do not differ much from BNC-Spoken. Actually, they can be considered as the only overlapping items in both corpora. Shall has 219 pmw in TEFL CC against 229 pmw; *ought to* has 50 pmw against 102 pmw occurrences. The items *need and dare* do not have any occurrences in TEFL CC at all to make comparison with BNC-Spoken. After all, they are not used enough to be considered significant in natural language either. Thus, they cannot be compared.

The last two columns in Table 4.25 show the Log-likelihood figures after comparing frequencies from both corpora. As stated earlier, Log-likelihood value is significance test of the comparison. Normalized frequencies and Log-likelihood values match up in showing the modal usages in both corpora. There are four modal items – *can, may should and must* – which are overused in TEFL CC. On the other hand, *might, will, could and would* are the underused items in TEFL CC. *Shall and ought to* have very little Log-likelihood values to make comparison. Thus, *shall and ought to* representation in course books seem to reflect real life usage. Lastly, *need and dare* have too few occurrences in BNC-Spoken and no occurrences in TEFL CC; therefore, they are not significant to compare.

4.7.2. Modals in Negative Form

Modals in negative sentences do not resemble much to their affirmative counterparts. There are more over used items and full and contracted negatives differ completely from negatives in total. Table 4.26 shows the distribution of all negative modals in two corpora.

Table 4.26 Modals in Negative Comparison

Modals Negative Comparison	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
can	1.162	1.391	6,71	+
may	41	38	0,05	-
might	39	13	3,95	-
will	527	439	2,45	-
could	349	539	13,91	+
would	535	219	37,92	-
should	141	407	51,75	+
must	31	201	63,56	+
Shall	14	0	-	-
ought to	2	19	8,39	+
Need	12	63	16,44	+
Dare	8	0	-	-

In affirmative modals, *will* is the most used item followed by *can* in BNC-Spoken. In addition, there is not too much gap on usage between them. However, in

negative sentences, situation is the opposite. *Can* in negative form is the most used modal verb with 1.162 pmw occurrences while *will* in negative form has only 527 pmw occurrences in BNC-Spoken. In fact, *would* is used more than *will* with 535 pmw occurrences in real life language.

On the other hand, *could* has 349 pmw and *should* has 141 pmw occurrences, which are quite low compared to affirmative forms. Other negative modals – *may, might, must, shall* etc – have very few occurrences as can be observed on the Table 4.26. An interesting observation is the frequency of *must* in negative forms. It is generally used in situations when somebody or something is prohibited from something. However, in natural English, it has a quite low pmw value like 31 in the corpus.

In TEFL CC, the frequency ranking of negative modals is similar to positive ones. *Can* in negative form is again the most used modal in TEFL CC with 1.391 pmw occurrences. *Will*, on the other hand, is not the second most used negative modal with 439 pmw. It is *could* with 539 pmw occurrences, which are more than *will*. *Will* is followed by *should, would and must* in negative forms with occurrences more than 200 pmw. An interesting item to consider in TEFL CC is need. It is used more than *may and might* in negative sentences.

Comparing the significance of those frequencies between two corpora, their Log-likelihood values are taken into consideration. Among the modals in negative forms, *could need, should and must* are the most significant ones in that they are relatively overused compared to BNC-Spoken. On the other hand, would can be considered as the only underused item in TEFL CC with 37,92 Log-likelihood value. There a few more underused items on the table, but with low Log-likelihood values. So, the difference is not considered as significant. *Can* in negative form is also overused with a low value of 6,71. Lastly, since *shall and dare* in negative forms have no occurrences in TEFL CC, they cannot be compared to BNC-Spoken and they do not have Log-likelihood values.

Above is the comparison of total frequencies and Log-likelihood values of modals in negative form. In addition to this, it would be wise to distinguish and

compare both full and contracted forms of negative modals. Table 4.27 demonstrates the distribution of negative modals between full and contracted forms in both corpora.

Table 4.27 Contracted and Full Negatives Comparison in Both Corpora

Negative comparison	TEFL CC Contracted Negative N-pmw	BNC-Spoken Contracted Negative N-pmw	TEFL CC Full Negative N-pmw	BNC-Spoken Full Negative N-pmw
can	1.310	1.090	81	72
may	0	0	38	41
might	0	5	13	34
will	420	472	19	55
could	514	334	25	15
would	219	497	0	38
should	395	122	13	19
must	194	4	6	27
Shall	0	13	0	2
ought to	0	0	19	2
Need	63	9	0	3
Dare	0	7	0	1

Contracted forms are one of the main features of authentic language. As can be observed on the table, contracted forms in most of the items are used more than full forms. This fact is true for both corpora. For instance, in the use of *can* in negative forms, the distribution is roughly same with 1.310 pmw full forms and 81 pmw contracted forms in TEFL CC; and 1.090 pmw full forms and 72 pmw contracted forms in BNC-Spoken. The figures of items *may, might and will* are close to each other in both corpora.

However, the table presents us with some interesting facts after comparison. First, modal item *could* in full and contracted forms is overused compared to BNC-Spoken. However, *would* is underused in full forms and there is no occurrence of *would* in contracted forms in TEFL CC. In addition, while the item *should* is overused in full forms, it is underused in contracted forms but with little difference.

The most interesting observation is, perhaps, the use of *must*. Both in affirmative and in negative modals, *must* is the overused item in TEFL CC. The difference is huge in contracted negatives with 194 pmw against 4 pmw in BNC-Spoken. However, in full negatives must is the underused item with 6 pmw against 27 pmw.

The affirmative item *need* has no occurrence at all in TEFL CC. Yet, *need* in negative form is quite common and has more occurrences (63 pmw against 12 pmw) than its BNC-Spoken counterpart. Although it has 9 pmw occurrences in contracted forms in BNC-Spoken, it has 63 pmw occurrences in TEFL CC. In fact, it has no occurrence in full forms against 3 pmw in BNC-Spoken. The remaining items – *shall, need and dare* – are not covered in course books at all both in contracted and full forms.

4.7.3. Modals Total

In previous sections, affirmative, full and contracted negative frequencies and their significance values were discussed. In this section, modal forms in total in both corpora will be covered. Table 4.28 indicates the total distribution and their Log-likelihood values of modals in both corpora. Since affirmative forms occupy almost %90 of all forms, negative frequencies added does not make much difference on the table.

Table 4.28 Modals Total Comparison

Word Modals Total	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
can	4.214	5.283	39,42	+
may	417	746	32,90	+
might	706	226	70,53	-
will	4.827	3.973	25,35	-
could	1.703	1.529	2,91	-
would	3.909	2.093	160,50	-
should	1.022	1.667	53,60	+
must	518	1.084	73,48	+
shall	244	219	0,40	-
ought to	103	69	2,05	-
need	18	63	10,65	+
dare	15	0	-	-

Log-likelihood distribution shows which items are overused and underused. Overused modals, just like in positive forms, include *can, may, should, must and dare*. *Dare* in positive form is not among the over used items. However, when negatives and positives are added, it becomes an overused item. The modal *must* has the highest Log-likelihood score with 73,48. It is almost twice more used than its BNC-Spoken counterpart. *Should* is the other overused item with 53,60 Log-likelihood value. *Can and may* have similar values around 30.

There are more underused items observed than overused ones in TEFL CC. *Would* is almost twice more used in BNC-Spoken than in TEFL CC with 160,50 log-likelihood value. Another underused item with 70,53 Log-likelihood value is *might*. It has a 226 pmw value in TEFL CC against 706 pmw in BNC-Spoken. The last significant underused item is *will* with 25,35 Log-likelihood value. Other remaining three modals – *shall, ought to and could*- are also underused, but with low Log-likelihood scores. Thus, they are not considered as significant.

4.8. Verb Usage in Modals

In the previous section, the frequency of modals in positive and negative form with their N-pmw values were explained in detail. Same procedure will be used to explore the immediate right collocates of each modal forms with their N-pmw values comparison, Log-likelihood data and their collocation strengths. This way, how a certain modal behaves in context can be explored.

The modal *can*, being the most overused item, is the main characteristic feature of course book language with 5.283 pmw. Therefore, it can be inferred that it is one of the key elements to make comparison since the Log-likelihood value is too much between two corpora. Table 4.29 shows the top 20 verb distribution of the modal *can* with their pmw and Log-likelihood values.

Table 4.29 Can Verb Distribution Comparison

Top 20 Verbs Used with Can	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
do	232,21	187,99	1,42	-
get	205,33	81,46	15,34	-
see	181,49	187,99	0,04	+
remember	102,62	18,79	16,42	-
go	109,13	131,59	0,68	+
be	217,42	375,99	14,87	+
tell	47,59	12,53	5,80	-
say	57,57	43,86	0,56	-
take	48,18	43,86	0,06	-
put	42,27	6,266	7,61	-
afford	38,21	50,13	0,53	+
make	43,87	81,46	4,02	+
find	37,79	81,46	5,93	+
hear	34,74	31,33	0,05	-
use	35,34	43,86	0,30	+
think	32,38	12,53	2,51	-
understand	26,54	43,86	1,48	+
give	31,02	18,79	0,89	-
come	30,52	18,79	0,82	-
imagine	21,13	31,33	0,67	+

Though the difference of normalized frequencies of *can* in both corpora is high, the verb distribution does not show too much divergence in TEFL CC compared to BNC-Spoken. That is, most of the top 20 verbs used with the modal *can* has, more or less, the same distribution in both corpora. However, some individual verbs need more attention. The verbs *make, find and be* are the most overused items in TEFL CC compared to BNC-Spoken. The verb *be* with 14,87, *make* with 4,02 and *find* with 5,93 positive Log-likelihood values are distinct features in TEFL CC. On the other hand, 4 other verbs – *get, remember, tell and put* – have negative Log-likelihood values. They are less frequent than their BNC counterparts. The remaining most frequent verbs used with the modal *can* have too small Log-likelihood values so they cannot be compared. As a result, the remaining verbs in course book corpora can be considered as authentic and reflect real life language in terms of frequency.

Table 4.30 Could Verb Distribution Comparison

Top 20 Verbs Used with Could	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
be	194,09	162,93	0,83	-
do	98,06	56,39	3,30	-
get	76,42	31,33	5,42	-
go	43,37	56,39	0,56	+
see	41,76	31,33	0,45	-
put	20,63	0	-	-
say	23,25	0	-	-
take	20,37	12,53	0,55	-
find	17,92	43,86	4,17	+
tell	17,08	0	-	-
make	18,09	25,06	0,38	+
afford	12,76	6,26	0,64	-
come	14,29	25,06	1,04	+
believe	11,75	12,53	0,01	+
use	13,78	6,26	0,81	-
hear	10,06	6,26	0,26	-
give	11,50	6,26	0,45	-
remember	6,42	6,26	0,00	-
buy	6,42	0	-	-
help	6,85	31,33	7,21	+

Table 4.30 shows the top 20 verb distribution of the modal *could* with their pmw and Log-likelihood values. *Could* is one of the least significant under used items in over all modal comparison, as shown in previous sections. A similar observation can be made in distribution of top 20 verbs which are used with *could*. Almost all of the verbs are underused in TEFL CC except a few such as help and find with 7,21 and 4,17 pmw values. Actually, these two are the most significant values on the whole table. The remaining verbs have either too low significance value or no occurrence at all. Only the verbs *get and do* can be observed as the only significant underused items with 5,42 and 3,30 pmw values. Other underused items are not considered significant and, therefore, the TEFL CC seems to represent authentic language use. In addition, there are four items which are not present in TEFL CC at all. These are the verbs *buy, tell, say and put*. Since they do not occur in TEFL CC, they cannot be compared to their BNC-Spoken counterparts As a result, it can be assumed that verb representations of the

modal *could* is similar to authentic language except a few items, as shown in Table 4.30.

Table 4.31 May Verb Distribution Comparison

Top 20 Verbs Used with May	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
be	126,96	137,86	0,14	+
want	5,83	0	-	-
say	6,67	0	-	-
find	5,49	6,26	0,02	+
get	4,48	18,79	3,94	+
need	3,80	18,79	4,68	+
think	3,97	12,53	1,82	+
feel	2,78	6,26	0,50	+
come	3,04	6,26	0,41	+
wish	2,28	0	-	-
go	3,04	12,53	2,57	+
know	2,70	0	-	-
remember	1,94	0	-	-
take	2,62	6,26	0,57	+
seem	1,60	0	-	-
affect	1,09	0	-	-
help	1,26	0	-	-
make	1,94	0	-	-
change	1,26	12,53	5,36	+
sound	1,01	0	-	-

As for the modal *may*, Table 4.31 shows its top 20 verbs distribution with their pmw and Log-likelihood values. The modal *may* is one of the most overused item in total in TEFL CC with 32,90 positive Log-likelihood value. However, the top 20 verbs distribution in TEFL CC is not in compliance with its overall frequency as can be observed on the table. Almost all occurrences of *may* is with the verb '*be*' which has 137,8 pmw positive value against 126,9 pmw in BNC-Spoken. Its representation in TEFL CC is similar to BNC-Spoken with 0,14 Log-likelihood value. As for the remaining verbs, only 3 of them – *change, need and get* – have high significance values. 10 verbs out of 20 most frequent verbs do not have any occurrence in TEFL CC at all.

The last 6 verbs such as *take, come and feel* have roughly the same significance values; thus, they can be considered authentic.

The modal *must* is the most significant positive modal item with 73,48 Log-likelihood value as shown in the previous section. It is almost twice more frequent in course book than in BNC-Spoken. However, its verb distribution in TEFL CC is not as diverse as the modals presented earlier. Just like the modal *may*, 10 verbs are not even represented in the whole TEFL CC such as *get*, *come* and *put*. Table 4.32 summarizes the top 20 verb distribution comparison values of must in both corpora.

Table 4.32 Must Verb Distribution Comparison

Top 20 Verbs Used with Must	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
be	154,44	269,46	10,98	+
admit	19,86	0	-	-
say	16,39	12,53	0,16	-
get	9,89	0	-	-
do	10,90	12,53	0,47	
go	7,354	50,13	25,04	+
remember	3,88	6,26	0,02	+
make	5,41	18,79	4,58	+
take	3,88	31,33	14,23	+
put	2,87	0	-	+
come	3,29	0	-	-
confess	2,02	0	-	-
know	3,21	6,26	0,69	-
tell	2,36	0	-	+
give	2,62	18,79	9,46	-
forget	1,43	0	-	+
ensure	1,43	0	-	-
try	1,77	0	-	-
look	2,28	6,26	1,07	-
ask	1,77	0	-	+

The verb '*go'* on Table 4.32 has the highest significance with 25,04 positive Log-likelihood value. It basically means that although it is used few times in BNC-Spoken, it is the second most frequent verb used with the modal *must* in TEFL CC with 50 pmw value. Four other verbs -*take, be, give and make*- are also used more than their

BNC-Spoken counterparts with high significance values as can be observed on Table 4.32. The remaining verbs are either not even present in TEFL CC and thus cannot be compared; or have too low significance values.

The modal *should* is also among the most significant items with 1.667 pmw in TEFL CC against 1.022 pmw in BNC-Spoken. However, just like the modal *must* the verbs used with *should* in TEFL CC is not diverse and significant in numbers. Table 4.33 shows the top 20 verb distribution comparison values of should in both corpora.

Table 4.33 Should Verb Distribution Comparison

Top 20 Verbs Used with Should	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
be	307,87	238,12	2,70	-
think	45,56	0	-	-
do	28,57	12,53	1,80	-
go	19,44	43,86	3,53	+
say	20,28	0	-	-
get	15,63	0	-	-
take	10,313	50,13	12,27	+
imagine	7,26	0	-	-
know	9,46	6,266	0,19	-
put	6,42	25,06	4,83	+
come	6,76	0	-	-
make	7,43	6,26	0,03	-
happen	3,88	0	-	-
give	4,81	18,79	3,62	+
try	3,71	12,53	2,00	+
pay	3,29	6,26	0,33	+
let	2,78	6,26	0,50	+
ask	3,12	6,26	0,38	+
look	3,88	0	-	-
keep	2,87	31,33	14,23	+

7 verbs – *think, say, get, imagine, come, happen and look* - some of which are among the most frequent verbs in English language, do not have any occurrence in TEFL CC at all, so a comparison cannot be made. On the other hand, the verbs – *take, go, put, give and keep*- have high positive significance values, which means that they are overused in TEFL CC compared to their authentic counterparts. For instance, the verb *keep* has 31 pmw value in TEFL CC against 2 pmw; thus, it has 14,23 positive

Log-likelihood value. 4 underused verbs can be observed on Table 4.33. However, their Log-likelihood values are not high, so they are not considered as significant. The verb *be* is the most common item in both corpora with 307 pmw in BNC-Spoken against 238 pmw. Yet, its Log-likelihood value is negative 2.70. That means the value is not high enough to be considered significant and the verb *be* can be regarded as representing the real life modal usage. The remaining verbs, just like verb *be*, have too low significance value and it can be concluded that TEFL CC seems to be using them as in BNC-Spoken.

The modal *would* is the third most frequent modal in BNC-Spoken as it is also in TEFL CC. In addition, it has 160,50 negative Log-likelihood value, which means the TEFL CC use it less often than real life samples as shown in previous section. Top 20 verb distribution of the modal *would*, as shown on Table 4.34, also proves this fact. 9 verbs do not have even occurrences in TEFL CC and a comparison cannot be made although they are quite frequent in BNC-Spoken.

Table 4.34 Would Verb Distribution Comparison

Top 20 Verbs Used with Would	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
be	695,96	288,26	48,52	-
say	115,55	18,79	19,81	-
go	73,20	37,59	3,33	-
get	61,62	6,26	12,99	-
do	66,10	25,06	5,29	-
think	46,40	0	-	-
come	44,88	6,26	8,32	-
want	37,19	0	-	-
take	30,85	37,59	0,22	+
put	23,92	0	-	-
make	30,09	12,53	2,08	-
expect	20,37	0	-	-
give	24,09	12,53	1,06	-
know	25,78	0	-	-
suggest	16,315	0	-	-
need	19,02	0	-	-
love	15,89	0	-	-
mind	15,38	12,53	0,09	-
agree	12,00	0	-	-
let	12,172	0	-	-

We can observe mostly underused items in TEFL CC and almost all of them have high significance values. For example, the verbs – *be and say* – have 48,52 and 19,81 negative Log-likelihood values respectively, which means they are relatively underused in TEFL CC compared to the BNC-Spoken. The remaining verbs, except for *take* which is the only overused item with low significance value, are all underused in TEFL CC. Only two of them seem to have low significance value. As a result, only 3 verbs – *give, mind and take* – are similar to real life language use.

Table 4.35 shows the top 20 verb distribution comparison values of the modal *will* in both corpora. However, the use of *will* was explained in detail in section 4.5.5 where the verb usages of tenses are explained in detail. Since the Future Tense *will* and modal *will* cannot be explored separately in Sketch Engine, their verb distribution table is same with just one exception. In tense verb distributions the verb *be* was excluded from the list since it is the most common item in almost all tenses by default. So, modal *will* verb distribution is the same as Future Tense verb distribution except for the verb *be* and the same analysis can be made. The verb *be* has 21,83 positive Log-likelihood value, which means it has more frequency in TEFL CC with 1290 pmw against 912 pmw in BNC-Spoken. The verb *do* is the most underused item with 12,09 significance value. 8 verbs seem to represent real life language since they are close to 0; but, the others are underused in TEFL CC. See Section 4.5.5 for a detailed analysis.

Table 4.35 Will Verb Distribution Comparison

Top 20 Verbs Used with Will	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
be	912,45	1290,91	21,83	+
get	184,53	81,46	11,51	-
go	162,13	144,13	0,33	-
do	169,06	68,93	12,09	-
see	107,18	43,86	7,62	-
give	91,88	31,33	8,48	-
come	90,36	75,19	0,43	-
take	87,49	43,86	4,21	-
tell	68,97	37,59	2,70	-
put	53,84	12,53	**7,29**	-
find	49,95	37,59	0,53	-
say	57,56	6,26	11,84	-
make	43,53	43,86	0,00	+

try	33,72	12,53	2,78	-
need	33,13	18,79	1,16	-
let	26,29	6,266	3,49	-
leave	27,89	12,53	1,68	-
know	33,05	0	-	-
ask	23,33	0	-	-
pay	19,78	12,53	0,48	-

As for the remaining items, 4 modals – *might, need, ought to and shall* –have a few occurrences in all top 20 most frequent verbs. So, only a few comparisons can be made in 4 of them. For the modal *might*, it has only the verb be as the only and most significant item with 28,46 negative significance value. The other verbs – *find, take and help* – which are the only verbs occurring in TEFL CC, have too low significance value; so, they are similar to their BNC-Spoken counterparts. The remaining 16 verbs don't have any occurrences at all in TEFL CC at all although they all have occurrences in BNC-Spoken as can be observed on Table 4.36.

Table 4.36 Might Verb Distribution Comparison

Top 20 Verbs Used with Might	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
be	200,51	43,86	28,46	-
get	24,51	0	-	-
say	15,89	0	-	-
do	16,14	0	-	-
want	10,82	0	-	-
go	11,24	0	-	-
think	10,14	0	-	-
like	9,29	0	-	-
find	8,70	6,26	0,12	-
come	8,53	0	-	-
need	6,50	0	-	-
take	5,07	6,26	0,04	+
help	3,97	6,26	0,18	+
call	4,05	0	-	-
make	5,32	0	-	-
see	4,73	0	-	-
happen	3,38	0	-	-
know	3,46	0	-	-
put	2,53	0	-	-
ask	2,28	12,53	3,45	+

The modal need just like *might* has only two occurrences of top 20 most frequent verbs which are *be* and *go*. The other verbs do not have any occurrences at all in the whole TEFL CC. In addition, these two verbs are overused compared to BNC-Spoken with 2,49 pmw and 2,38 pmw significance values respectively. Table 4.37 shows the verb distribution values of the modal *need*.

Table 4.37 Need Verb Distribution Comparison

Top 20 Verbs Used with Need	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
be	3,12	12,53	2,49	+
do	1,09	0	-	-
bother	0,67	0	-	-
go	0,76	6,26	2,38	+
worry	0,50	0	-	-
like	0,33	0	-	-
take	0,25	0	-	-
argue	0,16	0	-	-
concern	0,16	0	-	-
buy	0,16	0	-	-
get	0,25	0	-	-
meet	0,16	0	-	-
cost	0,16	0	-	-
put	0,16	0	-	-
know	0,16	0	-	-
elaborate	0,084	0	-	-
rush	0,084	0	-	-
lock	0,084	0	-	-
sing	0,0845	0	-	-
contribute	0,0845	0	-	-

Table 4.38 shows the top 20 verb distribution comparison values of *ought to* in both corpora. Though quite few, BNC-Spoken corpus contains occurrences of most frequent verb forms. With 33 pmw, the verb *be* is the most frequent item used with the modal *ought to*, which is followed by *go* with 5,9 pmw and *do* with 4,6 pmw respectively. On the other hand, the TEFL CC contains only 4 instances of all top 20 verbs with equal pmw distribution of 6,2 pmw for each verb. Their comparison within two corpora, however, doesn't present significant results except the verb *ask*. Although all other 3 verbs are used evenly in both corpora, the verb *ask* is used more in TEFL CC

with 2,38 positive Log-likelihood value. As a result, the *ought to* representation in course book corpus is quite few and doesn't reflect authentic language use as can be observed on Table 4.38.

Table 4.38 Ought To Verb Distribution Comparison

Top 20 Verbs Used with Ought to	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
be	33,98	0	-	-
go	5,91	6,26	0,00	+
do	4,64	0	-	-
get	2,62	0	-	-
take	2,11	6,26	0,83	+
see	2,11	6,26	0,83	+
say	2,02	0	-	-
look	1,52	0	-	-
put	1,35	0	-	-
know	1,52	0	-	-
start	1,01	0	-	-
try	1,01	0	-	-
write	0,92	0	-	-
make	1,18	0	-	-
ask	0,76	6,26	2,38	+
give	0,84	0	-	-
tell	0,598	0	-	-
learn	0,42	0	-	-
buy	0,42	0	-	-
let	0,42	0	-	-

Table 4.39 shows the top 20 verb distribution comparison values of shall in both corpora. Like *ought to*, *need and might*, the modal *shall* has only one occurrence, the verb *be*, with 6,2 pmw in TEFL CC against 31,8 pmw in BNC-Spoken. That means it is underused with 4,87 significance value. A s a result, the modal *shall* verb distribution in BNC-Spoken is not represented in TEFL CC since there are no more verbs as collocates of the modal *shall*.

Table 4.39 Shall Verb Distribution Comparison

Top 20 Verbs Used with Shall	BNC-Spoken N-pmw	TEFL CC N-pmw	Log-likelihood	Overuse/ Underuse
be	31,87	6,26	4,87	-
have	15,98	0	-	-
sell	4,90	0	-	-
go	4,31	0	-	-
see	3,30	0	-	-
give	2,87	0	-	-
get	2,79	0	-	-
say	3,04	0	-	-
tell	1,44	0	-	-
put	1,27	0	-	-
do	2,87	0	-	-
take	1,52	0	-	-
need	1,10	0	-	-
want	1,01	0	-	-
try	0,76	0	-	-
ask	0,76	0	-	-
move	0,68	0	-	-
buy	0,59	0	-	-
come	1,01	0	-	-
leave	0,68	0	-	-

4.9. Modals Collocation Strengths

In the previous section, 20 most frequent R1 collocates of Modal forms were investigated and their relative frequency and log-likelihood values after comparison were showed in detail. As stated earlier in previous sections, it is not always trustworthy to compare just the raw frequencies and their relative values. It has to be made sure that these collocates do not happen to be together just by chance. For that reason, T-score and MI values of the verbs which are used with the modal structures needed to be employed to determine if they really collocate or not.

In this section, the verbs, which are used with modals, were extracted from both corpora and were analyzed to compare with each other. This way, how close the course books resemble to authentic language in terms of modal usage and the order of the verbs

to be presented in course books will be determined. Table 4.40 shows the association scores of 20 most frequent verbs in all modals.

4.9.1. can

All the verbs presented in previous section were ordered according to their frequency in BNC-Spoken and their comparison was done on frequency level. The verbs are ordered in frequency order again in Table 4.40. As can be observed on the table, their T-scores are also roughly compatible with their frequency orders. However, when the verbs are ranked according to their MI scores, there can be a few discrepancies. For instance, although the verb '*be'* is one of the most used item in modal can, its MI value carries it to the last place as a collocate in BNC-Spoken. In addition, although the verbs – *afford, imagine and understand* – have high MI values, their significance levels carry them to the last places in terms of their T-scores.

In course book side, the TEFL CC appears to be using the verb '*see'* on the first place as the strongest collocate of the modal can. Yet, it is not considered significant when compared to its BNC-Spoken counterpart, and it seems to represent real life usage. The other verbs that roughly represent BNC-Spoken strong verb collocates are *do, go, be, hear, use, give and come*. That means 8 verbs in TEFL CC are used approximately with similar collocation strengths. On the other hand, 5 verbs – *find, make, afford, understand and imagine* – have high collocation strength values compared to the BNC-Spoken. In addition, 7 verbs –*get, say, take, remember, tell, put and think* – have low T-score and MI-score values in course book corpus. That means, only % 40 of all most frequent verbs are represented proportionally in TEFL CC.

Table 4.40 Association Measures of Most Frequent Verbs in Modals in Both Corpora

Can					could					dare				
	BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC	
	T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score
do	49.516	4.178	3.755	1.669	be	37.154	2.155	1.927	0.685	say	6.284	5.753	0	0
get	48.050	5.319	2.888	2.330	do	32.108	4.126	1.719	1.227	go	2.818	4.042	0	0
see	45.192	5.341	5.169	4.150	get	29.181	5.085	1.765	2.246	tell	2.158	4.837	0	0
remember	34.632	7.372	1.573	3.449	go	21.385	4.163	2.332	2.167	leave	1.924	4.721	0	0
go	34.110	4.303	3.509	2.094	see	21.183	4.413	1.928	2.860	put	1.919	4.634	0	0
be	27.488	1.127	2.621	0.596	put	15.080	4.853	0	0	touch	1.715	6.684	0	0
tell	22.845	4.749	0.879	1.402	say	14.180	2.787	0	0	do	1.669	1.979	0	0
say	22.608	2.904	2.350	2.564	take	14.124	3.471	0.915	1.504	move	1.669	4.774	0	0
take	21.795	3.521	1.992	2.016	find	13.731	4.135	2.511	4.294	ask	1.652	4.439	0	0
put	21.499	4.697	0.399	0.734	tell	13.568	4.462	0		come	1.532	3.117	0	0
afford	21.190	8.242	2.806	7.007	make	12.825	3.020	1.662	2.565	mention	1.393	6.084	0	0
make	20.137	3.107	3.146	2.971	afford	12.235	7.852	0.975	5.302	speak	1.372	5.074	0	0
find	19.839	4.019	3.363	3.892	come	11.606	3.221	1.730	2.887	carry	1.363	4.775	0	0
hear	19.784	5.372	2.069	3.744	believe	11.432	5.043	1.347	4.395	open	1.339	4.240	0	0
use	18.328	3.272	2.017	2.073	use	11.283	3.105	0.322	0.561	look	1.210	2.793	0	0
think	17.443	3.202	0.395	0.473	hear	10.511	4.776	0.848	2.717	give	1.206	2.762	0	0
understand	17.350	5.581	2.589	5.542	give	10.337	3.138	0.662	1.565	take	1.121	2.269	0	0
give	17.315	3.378	1.253	1.855	remember	8.350	4.566	0.888	3.159	fancy	0.993	7.215	0	0
come	16.821	3.125	0.966	1.177	buy	8.347	4.554	0	0	dare	0.992	6.982	0	0
imagine	15.667	6.776	2.201	6.007	help	8.189	3.658	2.090	3.933	whisper	0.991	6.843	0	0

Table 4.40 Association Measures of Most Frequent Verbs in Modals in Both Corpora

may					might					must				
	BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC	
	T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score
be	35.388	3.525	3.544	2.033	be	43.760	3.300	2.011	2.059	be	37.237	2.956	5.298	2.381
want	7.970	4.625	0	0	get	16.299	4.544	0	0	admit	15.289	8.562	0	0
say	7.754	2.970	0	0	say	12.355	3.337	0	0	say	12.623	3.415	1.215	2.830
find	7.683	4.412	0.881	3.076	do	11.575	2.622	0	0	get	9.693	3.267	0	0
get	6.354	2.975	1.530	3.099	want	10.858	4.633	0	0	of	9.499	1.018	0	0
need	6.273	3.946	1.668	4.758	go	10.373	3.315	0	0	do	8.687	2.088	0.027	0.028
think	6.183	3.349	1.276	3.358	think	10.178	3.818	0	0	go	7.926	2.735	2.467	2.967
feel	5.415	4.122	0.912	3.499	like	9.622	3.598	0	0	remember	6.566	4.973	0.943	4.130
come	5.236	2.973	0.820	2.477	find	9.594	4.192	0.963	4.754	make	6.495	2.410	1.533	3.121
wish	5.078	5.463	0	0	come	9.208	3.577	0	0	take	5.319	2.213	2.075	3.796
go	4.792	2.313	0.943	1.586	need	8.161	3.838	0	0	put	5.170	3.142	0	0
know	4.658	2.502	0	0	take	6.436	2.564	0.927	3.771	come	4.921	2.237	0	0
remember	4.626	4.824	0	0	help	6.419	3.972	0.966	4.878	confess	4.881	8.067	0	0
take	4.580	2.494	0.766	2.093	call	6.415	3.755	0	0	know	4.511	1.899	0.791	2.258
seem	3.924	3.327	0	0	make	6.386	2.355	0	0	tell	4.501	2.743	0	0
affect	3.488	4.933	0	0	see	6.036	2.371	0	0	give	4.301	2.136	1.632	4.121
help	3.454	3.208	0	0	happen	6.036	4.453	0	0	forget	3.961	4.665	0	0
make	3.404	1.785	0	0	know	4.775	1.976	0	0	ensure	3.928	4.404	0	0
change	3.344	2.871	1.371	5.017	put	4.758	2.928	0	0	try	3.902	2.751	0	0
sound	3.255	4.050	0	0	ask	4.545	2.996	1.386	5.663	look	3.868	1.968	0.790	2.250

Table 4.40 Association Measures of Most Frequent Verbs in Modals in Both Corpora

need					ought					shall				
	BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC	
	T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score
be	3.958	1.517	1.075	2.059	be	17.664	3.071	0	0	be	13.633	1.747	-0.680	-0.749
do	3.144	2.964	0	0	go	8.049	4.720	0.935	3.934	have	11.140	2.398	0	0
bother	2.824	9.253	0	0	do	6.585	3.157	0	0	sell	7.540	6.658	0	0
go	2.761	3.649	0.941	4.072	get	5.124	3.649	0	0	go	6.268	3.032	0	0
worry	2.440	7.956	0	0	take	4.597	3.632	0.977	5.441	see	5.436	2.949	0	0
like	1.756	3.036	0	0	see	4.570	3.538	0.978	5.476	give	5.254	3.337	0	0
take	1.417	2.461	0	0	say	4.144	2.699	0	0	get	4.735	2.508	0	0
argue	1.382	5.449	0	0	look	3.912	3.682	0	0	say	4.554	2.053	0	0
concern	1.381	5.419	0	0	put	3.804	4.353	0	0	tell	3.639	3.091	0	0
buy	1.357	4.624	0	0	know	3.754	3.119	0	0	put	3.398	3.029	0	0
get	1.347	2.168	0	0	start	3.301	4.407	0	0	do	3.349	1.232	0	0
meet	1.341	4.272	0	0	try	3.281	4.242	0	0	take	3.127	1.927	0	0
cost	1.333	4.118	0	0	write	3.181	4.612	0	0	need	2.909	2.372	0	0
put	1.265	3.241	0	0	make	3.088	2.516	0	0	want	2.770	2.318	0	0
know	1.018	1.837	0	0	ask	2.776	3.742	0.988	6.333	try	2.504	2.597	0	0
elaborate	0.994	7.350	0	0	give	2.709	2.802	0	0	ask	2.474	2.511	0	0
rush	0.986	6.164	0	0	tell	2.324	3.041	0	0	move	2.386	2.677	0	0
lock	0.997	8.494	0	0	learn	2.138	4.512	0	0	buy	2.379	3.313	0	0
sing	0.994	7.350	0	0	buy	2.102	4.058	0	0	come	2.325	1.605	0	0
contribute	0.980	5.661	0	0	let	2.090	3.932	0	0	leave	2.216	2.208	0	0

Table 4.40 Association Measures of Most Frequent Verbs in Modals in Both Corpora

should					will					would				
	BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC			BNC-Spoken		TEFL CC	
	T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score		T-Score	MI Score	T-Score	MI Score
be	54.454	3.356	4.116	1.590	be	88.132	2.721	11.808	2.496	be	77.241	2.749	4.178	1.381
think	22.675	5.422	0	0	get	44.912	4.690	2.946	2.450	say	34.414	3.853	1.385	2.317
do	15.891	2.882	-0.707	-0.585	go	41.719	4.399	3.852	2.345	go	27.118	3.671	1.556	1.455
go	13.863	3.542	2.055	2.162	do	40.002	3.244	0.700	0.342	get	24.658	3.527	-0.151	-0.203
say	13.718	3.126	0	0	see	33.541	4.106	2.058	2.171	do	22.321	2.309	-0.099	-0.069
get	12.251	3.332	0	0	give	31.481	4.469	1.895	2.712	think	21.584	3.665	0	0
take	9.688	3.024	2.634	3.861	come	30.936	4.215	3.112	3.297	come	21.176	3.625	0.410	0.760
imagine	9.199	6.964	0	0	take	30.026	3.906	2.044	2.136	want	19.726	4.068	0	0
know	9.128	2.862	0.680	1.645	tell	27.547	4.809	2.165	3.107	take	16.403	2.822	2.135	2.962
put	8.050	3.706	1.904	4.387	put	24.177	4.571	1.023	1.854	put	15.631	3.820	0	0
come	7.547	2.678	0	0	find	22.734	3.947	2.120	2.896	make	15.548	2.507	0.892	1.438
make	7.441	2.274	0.473	0.923	say	21.248	2.428	-0.245	-0.316	expect	14.987	4.852	0	0
happen	6.384	4.091	0	0	make	19.002	2.620	2.069	2.198	give	14.709	2.958	1.153	2.438
give	6.138	2.419	1.580	3.508	try	18.889	4.201	1.023	1.854	know	14.428	2.524	0	0
try	5.923	3.222	1.279	3.387	need	17.949	3.420	1.297	1.993	suggest	13.288	4.523	0	0
pay	5.667	3.433	0.924	3.721	let	16.999	4.794	0.518	1.054	need	13.174	3.038	0	0
let	5.321	3.760	0.834	2.587	leave	16.776	3.708	1.026	1.867	love	13.095	4.475	0	0
ask	5.261	2.887	0.703	1.753	know	16.188	2.463	0	0	mind	12.818	4.326	1.329	4.060
look	5.244	2.141	0	0	ask	15.228	3.584	0	0	agree	11.345	4.383	0	0
keep	5.109	3.013	2.153	4.747	pay	14.210	3.815	1.259	3.189	let	11.301	4.103	0	0

4.9.2. could

The frequency rank and collocation ranks are not so much different as verb tenses as can be observed on the Table 4.40. In the modal could, the frequency ranking and T-score rankings of the verbs are also the same just like the modal *can*. As stated earlier, the verbs which have high values of both T-score and MI values are considered as strong collocates. Thus, all the verbs on the list can be considered significant in BNC-Spoken.

As for TEFL CC, the table presents us with valuable data. Most of the verbs used with the modal *could* have no similarity to BNC-Spoken items. There are 7 verbs – *go, see, make, afford, hear, give and use*– which are used with roughly the same strength values. So, these verbs are considered to be represented in TEFL CC. However, 4 verbs including *be, get, do and take* have low collocation strengths in TEFL CC compared to BNC-Spoken. Besides, 5 other verbs –*find, help, come, believe and remember* – have more collocation strengths in course books. Lastly, the remaining 4 verbs – *put, buy, tell and say* – do not have any occurrence in course books at all; so, they cannot be compared and their collocation strengths cannot be assessed. To sum up, the modal *could* used in TEFL CC has just only 7 verbs which reflect authentic language use in terms of collocation strengths.

4.9.3. may

The modal *may* is one of the most overused items in total in course book corpus with 32,90 positive Log-likelihood value. However, as stated earlier, the modal *may* has only 10 verbs out of 20 most frequent verbs. In addition, their strengths as collocates are not as strong as in BNC-Spoken. T-score ranking of the verbs is just the same as the frequency ranking. So, BNC-Spoken is consistent with its collocation strengths.

However, the same similarity cannot be observed in terms of TEFL CC. According to the Table 4.40, there are only 5 verbs –*be, get, think, feel and come* - observable in TEFL CC which has similar collocation strengths as BNC-Spoken; and therefore, it can be deduced that these verbs are represented uniformly in TEFL CC. However, other 4 verbs – *need, change, go and take*- have high collocation strengths in

TEFL CC although they are not considered significant. In addition, only 1 verb *-find-* has low strength value compared to BNC-Spoken. The remaining 10 verbs do not occur in TEFL CC at all; so, they cannot be compared to the BNC-Spoken. As a result, for the modal *may*, TEFL CC does not seem to be representing the authentic collocation pairs in BNC-Spoken.

4.9.4. must

The modal *must* is the most significant positive modal item with 73,48 Log-likelihood value as shown in the previous section. It is almost twice more frequent in TEFL CC than in BNC-Spoken. The frequency ranking and verbs significance ranking are similar to each other as can be observed on Table 4.40. Just like the previous modals presented, the modal *must* also has consistency with frequency of collocates and their strength values.

On the other hand, TEFL CC collocation scores indicate a different ranking on the list. First of all, 10 items are not represented at all in course books. Therefore, it is not possible to do a collocation analysis for these verbs. Only 2 verbs – *be and remember* – have similar strength rankings compared to BNC-Spoken corpus. 6 other verbs – *go, take, give, make, know and look* – have high co-occurrence values; so, they are considered as strong collocates of the modal must in TEFL CC. On the other hand, the verbs *do and say* have low co-occurrence values and thus, different strength ranking. These 8 verbs, as a result, seem not to be representing authentic language in the modal must usages. In general, we can conclude that course books have weak collocations scores of 20 most frequent verb forms in the modal *must*.

4.9.5. should

The modal *should* is also among the most significant items with 1.667 pmw in TEFL CC against 1.022 pmw in BNC-Spoken. However, just like the modal *must* the verbs used with should in TEFL CC is not diverse and significant in numbers. The modal *should* collocation scores have also similarities with normalized scores presented in previous section. Collocations scores of the all verbs are on the same rank with

frequencies of those items. Thus, it can be concluded that BNC-Spoken frequency scores and collocation strength values comply with each other.

The same inconsistency as in previous modals in TEFL CC can be observed in the modal *should*, too. There are several verbs which are not represented at all. Since there are 7 verbs which do not occur with the modal should in TEFL CC, they cannot be compared to their BNC-Spoken counterparts. These items include the most frequent verbs in the whole corpus like *think, say, get, come* etc. In addition, 9 other verbs have high collocation strength values compared to BNC-Spoken verbs. However, 3 of them – *ask, let and pay* – have such low strength values (below 1 as can be observed on Table 4.40) that they cannot be considered as collocates in the modal should. Although they have average normalized frequencies, their strengths as collocations are relatively low. Besides, although 3 verbs –*do, make and know* – have similar ranking compared to BNC-Spoken, they cannot be considered significant as the values are too low. Only the verb *b*e is similar to its BNC-Spoken counterpart and has similar ranking with high strength value. As a result, the modal *should* does not seem to represent authentic language use.

4.9.6. would

Frequency listing and collocation strengths rankings regarding the modal *would* verbs are quite similar with each other in BNC-Spoken. Actually there is no discrepancy between two rankings just like the modals presented earlier as can be seen on Table 4.40. The modal *would* collocations also seem to comply with the normalized frequencies.

However, as discussed in the previous chapter regarding the modal *would*, although it is among the most frequent modals in both in TEFL CC and BNC-Spoken, it has only half of the most frequent verbs represented in TEFL CC. In addition, only 3 verbs – *be, go and say* – are similar to their BNC-Spoken usages in c TEFL CC. The remaining verbs have either too low or high strength values; so, they do not represent to authentic language usages. For instance, the verbs – *take, mind and give* – have high collocation strength values in TEFL CC. Besides, since 4 other verbs – *make, come, get and do* – have too low T-score values, they are not considered as collocates of the

modal *would* in TEFL CC. It can be assumed that the modal *would* collocations do not seem to represent real life language usages except 3 verbs.

4.9.7. will

The modal *will* collocation scores have also similarities with normalized scores presented in previous section. Collocation scores of the all verbs are on the same rank with frequencies of those items. Thus, it can be concluded that BNC-Spoken frequency scores and collocation strength values of the modal *will* comply with each other.

On the TEFL CC side, however, the same inconsistency as in previous modals in TEFL CC can be observed in the modal will, too. 2 verbs – *know and ask*- have no occurrence in course books and there are no scores for them on the Table 4.40. An interesting observation is that 3 other verbs – *let, say and do* – cannot be considered as collocates in the modal will since T-scores and MI values are considerably low (0.518, -0.245 and 0.700). Although they have high normalized frequencies, their strengths as collocations are relatively low. As can be seen on the table, there are only a few verbs which have similar ranks in terms of collocation strengths. These are '*be, go, get, take and try*'. However, the verbs '*see, give and put'* have low strength values in TEFL CC compared to BNC-Spoken. The remaining 7 verbs have more significance as collocations than their BNC-Spoken counterparts. The same level of inconsistency within two corpora is present in the modal *will*, too.

4.9.8. might

Frequency and collocation ranking of the most frequent verbs used with the modal *might* are all the same. All of the items are in the same rank on both columns. An interesting observation to make is that the verb '*be*' is the most frequent item used with *might* with 200 pmw in BNC-Spoken as stated in the previous section. Therefore, it can be stated that the verb '*be*' is the strongest collocate. The other items have also have enough occurrence – though few – in the corpora to assess their strengths.

As for TEFL CC, the similarity within frequency listing and collocation measures is not present for the verbs analyzed. Only 5 items occur in course book corpus out of 20 most frequent verbs used with the modal *might*. Thus, it wouldn't be

wrong to assume at first hand that the use of modal might in the TEFL CC does not resemble to BNC-Spoken in terms of collocation strengths. Although these 5 verbs have strength values as can observed on Table 4.40, they are too low to be considered as strong collocates of the modal might in TEFL CC.

4.9.9. need

The modal *need* does not have many occurrences and examples of collocations in both corpora to make assumptions compared to the previous modals. The most frequent verbs used with the modal might are *be and do* with 3,12 pmw and 1,09 pmw values, which are quite low in such a big corpus like BNC-Spoken. The per-million-word counts of remaining verbs are even lower than those. The frequency ranking and collocation strength rankings are quite similar except the ones which have too low T-score values below 1, and they are not considered as good collocates, either, as can be observed on Table 4.40.

As for the TEFL CC usage of the collocational behavior of the modal *might*, it has only two verbs occurring, which are *be* and *go*. Although they are more frequent – 12 pmw and 6 pmw - in TEFL CC than in BNC-Spoken, their collocation values are not as high as the BNC-Spoken ones. The verb '*be*' has 1,075 and the verb '*go*' has 0,941 T-score values, which do not make them strong collocates of the modal '*need*.' Since there no other verbs in TEFL CC, a further analysis cannot be done. The course cook corpus doesn't have enough occurrence of top 20 verbs; so, it does not resemble to BNC-Spoken collocations of the modal '*need*.'

4.9.10. ought to

Table 4.40 shows the top 20 verb collocation strengths comparison values of *ought to* in both corpora. Though quite few, BNC-Spoken corpus contains occurrences of most frequent verb forms, enough to make a generalizations. The frequency ranking of the modal *ought to* is the same as its collocation strength values ranking. Thus, it can be assumed that collocations are used in accordance with their frequency ranking in BNC-Spoken regarding the modal *ought to*.

On the other hand, as stated before in the previous section, the TEFL CC contains only 4 instances of all top 20 verbs with equal pmw distribution of 6,2 pmw for each verb. In addition, their T-score values are too low to be considered as collocates of the modal *ought to*, as can be observed on Table 4.40. As a result, since there are no more verbs occurring with the modal *ought to*, it can be deduced that TEFL CC does not resemble to authentic BNC-Spoken in terms of collocations of the modal *ought to*.

4.9.11. shall

The frequency ranking and collocation strength ranking of the modal *shall* is quite the same as can be observed on Table 4.40. However, on TEFL CC side, the modal *shall* has only one occurrence of the most frequent 20 verbs, which is the verb '*be*.' Yet, its T-score is below 0; therefore, it is not considered as a strong collocate of the modal shall. To sum up, the modal shall is not represented in TEFL CC in terms of its collocations.

CHAPTER 5

DISCUSSION and CONCLUSION

5.1. Discussion and Conclusion

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". The main purpose of this study was, from certain grammatical perspectives, 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.

The comprehensive literature has shown that there is a lot of controversy between researchers on the use and effectiveness of authentic materials in language classrooms. Nevertheless, their implementation in language classrooms has gained more supporters in recent years. The literature on authenticity proved that most students like dealing with authentic materials, but no empirical research was done about whether or not students' language competence improved after being exposed to authentic material. Only the advantages of using real life examples and students' feelings on authentic material were taken into consideration.

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.

Although course books are the central elements of language education, especially in Turkey, authenticity in course books does not seem to get necessary attention from researchers. There are few studies, but none of them surveyed the

linguistic content of course books. Almost all of them investigated learner attitudes, perceptions, or their face value. On the other hand, corpus, as a large database of authentic language collected and stored in computers, proved as an invaluable resource for authentic language content to be included in course books. This can be the only reason why corpus findings should be used for learning purposes, say, to determine what language features to include in course books. 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.

However, there is little effort to use this authentic corpus data in course books. Although publishers praise their books on the closeness to real life, they do not seem to be using this precious authentic corpus data. Scholars like Römer (2005), Gilmore (2004), Anping (2005), and Hyland (1994) analyzed the linguistic quality of some course books using corpus-based research analysis on a specific language point. They compared the real life language use with that used in course books, but they were not so comprehensive. Teachers, students, researchers all complain about the ineffectiveness of course books in Turkey. However, in Turkish context, though, there isn't any corpus-based authenticity analysis of any course books. As Mindt (1997) observed, "corpus-based studies of grammar can do much to bring the teaching of English into accordance with actual language use" (p.50). Therefore, this study aimed to explore the authenticity level of ELT course books used in Turkey from corpus linguistics point of view. As a result, whether Turkish language learners are exposed to real language data or not, or in which density were discovered and problematic items were shown in detail. The findings in this study showed that the language used by ELT course books used in high schools in Turkey had little similarity to real-life language features. In addition, this study showed how to use corpus data to evaluate course materials

In this respect, by using-corpus-based techniques, three corpus-informed data collection tools were administered; frequencies of items converted per million words (pmw) values, Log-likelihood values and collocations strength values (MI and T-score values extracted from Sketch Engine Corpus Query Tool).

Since it was impossible to compare all the features of the language, the commonly used grammatical points – tenses and modals – were then tried to be explored both in BNC-Spoken and TEFL CC and the data were compared afterwards. In the following pages, the research questions are discussed in line with the findings obtained from the data.

Research Question 1: What is the degree of authenticity of language course books used in Turkey compared to spoken part of BNC?

The grammatical items –Tenses and Modals - , their normalized frequencies and immediate right verb association values in BNC-spoken and their comparison with TEFL CC indicated that on the whole, the language course books used in high schools in Turkey do not reflect authentic language use in the target language. As shown in previous chapter, although there was a % 27 similarity between two corpora in terms of normalized frequencies, this percentage reduced even more to % 15 when the verb choice and their association values was taken into consideration in Tenses. With modals, the same amount of similarity (%27 similarity with normalized frequencies; %15with verb association values) was observed. It can be assumed that language learners in high schools in Turkey seem to be faced with too little real life examples.

Implications: These 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 a more authentic course materials.

Research Question 2: How much do specific grammar points (tenses and modals) in course books resemble to the grammar of BNC-Spoken corpus?

The overall results of the comparison of the data extracted from both corpora indicate that in general the TEFL CC items which were selected to make comparisons do not reflect the real life usages. As the tables in Chapter 4 indicate, majority of the items which were analyzed has little or no representation in the whole TEFL CC. The first step in data analysis was to compare pmw (per million words) frequency values of Tenses and Modals in both positive and negative sentences. Almost all Tenses in course books are not in compliance with authentic language examples in terms of their pmw values. Only Past Continuous and Past Perfect Continuous Tense have similar Log-likelihood scores, but they are not considered significant as there are not enough data both in BNC-Spoken and TEFL CC.

As for Modals comparison of pmw values, the similarity is no less different than in Tenses. Almost all modal forms are either overused or underused in TEFL CC. In addition, their order of importance (in terms of frequency) is completely different in both corpora. For instance, although the modal must is one of the most used items in TEFL CC, its rank is not that significant in BNC-Spoken. Other modals are presented in Chapter 4.

Implications: These discrepancies between authentic corpus and TEFL CC 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 prioritized in classroom settings in Turkey, some changes and improvements regarding Tenses and Modals 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, as can be observed in Chapter 4, 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 amount 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 opportunities and possibility to be prepared for real life communication situations.

A second implication to make is on the order of use of Modals in TEFL CC. A similar observation as in Tenses was made in course books regarding the presentations of modals and their frequency ranking. Although they were presented in a pattern, their frequencies were quite different. We can conclude that the modal representations in course books are even different from their frequency ranking. In terms of frequency, in course books modals were introduced in a pattern like:

can, will, would, should, could, must, may, might, shall, ought to and need

However, the BNC-Spoken corpus suggests a quite different pattern like:

Will, can, would, could, should, might, must, may, shall, ought to and need

As can be observed, the ranking in authentic language do not seem a lot different than course book ranking of modals. Although the rank has some similarity on the whole, individual frequencies and Log-likelihood values are quite different. Therefore, we can conclude that TEFL CC represent a roughly-similar ranking, but with very different frequencies. 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.

Research Question 3: How much do vocabulary choice in course books resemble to the vocabulary of BNC-Spoken corpus?

The second step in analysis was to compare the most frequent verbs used with each grammar point (Tenses and Modals). For this, only top 20 verbs used for each tense and modal item were extracted from both corpora and their pmw values and Log-likelihood scores were compared. As a result, which verbs are overused, underused or properly distributed in course book corpus were identified. As can be observed in related tables in Chapter 4, though there are some verbs both in tenses and modals which have similar scores and values, it's clearly visible that most of the items are underrepresented in TEFL CC. In addition, some of the items do not have any representation at all in the whole corpus, which makes it impossible to make comparison with BNC-Spoken.

In the third analysis step, collocation values of the top 20 verbs for each tense and modal item extracted from both corpora were computed in order to determine whether they happen to be together just by chance or they have a strong connection. Therefore, two most used calculation scores, T-score and MI-value, were computed in Sketch Engine for the verbs extracted in the second step. Table 4.24 and 4.40 show the association measures of the most frequent verbs in tense and modals in both corpora. This analysis aimed to explore the likelihood of meeting the verbs in the related grammatical item. What makes it different form second analysis is that how much stress and importance were given to each verb in BNC-Spoken and TEFL CC. In this way, by comparing the scores, the ideal order and number of the verbs to be introduced in course materials were identified.

Implications: The verb frequencies and collocation values present us with valuable data. ELT course books seemed to be using inauthentic language content which has no similarity to authentic English. To expect a one-to-one correspondence in two corpora would be unfair to course materials. However, roughly similar values would make them moderately related-to real life at least. Those uneven values and scores noticeably prove that the language features in course materials were written or prepared with no authenticity in mind. Using real life examples, photos, charts etc do not always make language learning materials authentic. They should reflect how the language is used in daily life. That is, which items and vocabulary are used most in which structures in daily life should be presented in course materials. Learners would know which feature to focus on more. The verbs used with tenses and modals in ELT course books do not resemble to authentic interactions in daily life in terms of their frequency and representation.

Research Question 4: Are the students exposed to the real life language they may need in communication situations?

Since the ultimate aim of language learning is to be able to communicate with other speakers, the language in real life and the language in language learning materials should be similar in most ways. Due to the various findings cited in the tables in Chapter 4, the course book language in Turkey is regarded as inauthentic. Frequency of tenses and modals, frequency of most-used verbs for each item, their Log-likelihood comparison, association measures of top verbs and their comparison are the proofs of inauthentic language used in course books.

Although there are some BNC-Spoken consistent items which reflect authentic language us, the course books are, on the whole, far away from real life in terms items concerned. Therefore, students who are using or have used these course books in their schools are not exposed to the real life language examples they may need in communication situations.

Implications: The findings above indicate that language learning materials used in Turkey in all high schools were prepared by the writers' intuition on how language is

used by native speakers. The data clearly shows that almost all the time language items were used either too much or too low with no reason and with no proof to support them.

5.2. Suggestions for Further Research

The following items may be researched for further studies. This study aimed to find out whether Ministry of Education mandated language course books in high schools are authentic in terms of the language items studied. Only two grammar items – tenses and modals – were picked up to study. To get a clearer picture on authentic language use in curse books, other grammar items should be investigated as well. This way, a more comprehensive generalization can be made on authenticity in language course books.

This study focused on a more grammatical perspective by dealing with tenses and modals and their verb collocations. However, vocabulary load and distribution and functional role of the language were disregarded. A more comprehensive study covering these features should be conducted. In addition, the language has more communicative function than just grammar and vocabulary. It includes a lot of discourse features employed by speakers during conversation. Thus, discourse function should also be explored.

This study concluded that students are not exposed to real life language they may use in communication situations. However, as stated earlier, authenticity is not a one-sided entity. It may have more features such as learner authenticity, task authenticity etc. Since this study focused only on text authenticity, the other authenticity types were disregarded. To determine if a language learning activity is authentic or not, a research combining all authenticity types should be conducted.

REFERENCES

Abalı, E. (2006). *Analysıs 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

Aktan, E. (2010). *Fransızca öğretiminde şarkıların özgün materyal olarak kullanılması.* Master of Arts, Gazi University, Ankara.

Al-musallam, E. I. (2009). *College Instructors' and Learners' Attitudes to Authentic EFL Reading Materials in Saudi Arabia.* Master of Arts, King Saud University.

Allen, E. D., & et al. (1988). Comprehension and Text Genre: An Analysis of Secondary School Foreign Language Readers. *Modern Language Journal, 72*(2), 163-172.

Allwright, R. (1976). Language Learning through Communication Practice. ELT Documents (76/3).

Anping, H. (2005). *Corpus-based evaluation of ELT textbooks.* Paper presented at the the joint conference of the American Association of Applied Corpus Linguistics and the international Computer Archive of Modern and Medieval English, University of Michigan.

Ansary, H., & Babaii, E. (2002). Universal characteristics of EFL/ESL textbooks: A step toward systematic textbook evaluation. *The Internet TESL Journal for Teachers of English as a Second Language, VIII*(2). Retrieved from http://iteslj.org/Articles/Ansary-Textbooks/ website:

Arnold, E. (1991). Authenticity revisited: How real is real? *English for Specific Purposes, 10*(3), 237-244.

Bachman, L. F. (1991). What Does Language Testing Have to Offer. *Tesol Quarterly, 25*(4), 671-704.

Bacon, S. M. (1989). Listening for Real in the Foreign-Language Classroom1. *Foreign Language Annals, 22*(6), 543-550. doi: 10.1111/j.1944 9720.1989.tb02781.x

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.

Baltova, I. (1994). The Impact of Video on the Comprehension Skills of Core French Students. *Canadian Modern Language Review, 50*(3), 507-531.

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*: Longman.

Berardo, S. (2006). The use of authentic materials in the teaching of reading. *The Reading Matrix, 6 (2),* 60-69.

Biber, D. (1999). *Longman grammar of spoken and written English*. Harlow, England; [New York]: Longman.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at …: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics, 25*(3), 371-405. doi: 10.1093/applin/25.3.371

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*: Cambridge University Press.

Biber, D., & Reppen, R. (2002). What Does Frequency Have To Do with Grammar Teaching? *Studies in Second Language Acquisition, 24*(2), 199-208.

Blaz, D. (2002). *Bringing the Standards for Foreign Language Learning to Life*: Eye on Education.

Block, D. (1991). Some thoughts on DIY materials design. *ELT Journal, 45*(3), 211-217. doi: 10.1093/elt/45.3.211

Boran, G. (1999). *Using television programmes in English as a source of authentic EFL video materials: A study on students and teachers attitudes* PhD, Gazi University.

Boxer, D., & Pickering, L. (1995). Problems in the presentation of speech acts in ELT materials: the case of complaints. *ELT Journal, 49*(1), 44-58. doi: 10.1093/elt/49.1.44

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

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

Brumfit, C. J. (1980). Seven Last Slogans. *Modern English Teacher, 7/1*, 30-31.

Cane, G. T. (1998). Teaching conversation skills more effectively. *The Korea TESOL Journal* (1), 31-37.

Cetinkaya, Y. B. (2005). *Turkish College Students' Willingness to Communicate in English as a Foreign Language.* Ohio State University.

Chavez, M. M. T. (1994). *Learners' Perspectives on Authenticity [microform] / Monika M. Th. Chavez*. [Washington, D.C.] :: Distributed by ERIC Clearinghouse.

Chomsky, N. (1957). *Syntactic structures.* The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, Mass: MIT Press.

Chujo, K. (2004). Measuring vocabulary levels of English textbooks and tests. Using a BNC lem-matised high frequency word list. In J. Nakamura, N. N. Inoue & T. Tabata (Eds.), *English Corpora under Japanese Eyes* (pp. 231-249). Amsterdam: Rodopi.

Clark, R. E. (1983). Reconsidering Research on Learning from Media. *Review of Educational Research, 53*(4), 445-459. doi:10.3102/00346543053004445

Crewe, J. (2011). *How far do 'global' ELT coursebooks realize key principles of Communicative Language teaching (CLT) and enable effective teaching-learning?* Master of Arts in TEFL/TESL, University of Birmingham, Birmingham.

Crossley, S., Louwerse, M., McCarth, Y., Philip, M., McNamar, A., & Danielle, S. (2007). A Linguistic Analysis of Simplified and Authentic Texts. *The Modern Language Journal, 91*(1), 15-30. doi: citeulike-article-id:1111183 doi: 10.1111/j.1540-4781.2007.00507.x

Crystal, D. (1997). *English as a global language.* Cambridge: Cambridge University Press.

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

Çakır, İ. (2007). An overall analysis of teaching compulsory foreign language at Turkish state universities. *Journal of Language and Linguistics Studies, 3*(2), 250-265.

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

Davies, A. (1984). Simple, simplified and simplification: what is authentic? In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 181-198). London: Longman.

Devitt, S. (1997). Interacting with Authentic Texts: Multilayered Processes. *The Modern Language Journal, 81*(4), 457-469. doi: 10.1111/j.1540-4781.1997.tb05512.x

Dogancay-Aktuna, S. (1998). The Spread of English in Turkey and its Current Sociolinguistic Profile. *Journal of Multilingual and Multicultural Development, 19*(1), 24-39. doi: 10.1080/01434639808666340

Dogançay-Aktuna, S. (2004). Language planning in Turkey: yesterday and today. *International Journal of the Sociology of Language, 2004*(165), 5-32. doi: 10.1515/ijsl.2004.008

Doğançay-Aktuna, S., & Kiziltepe, Z. (2005). English in Turkey. *World Englishes, 24*(2), 253-265. doi: 10.1111/j.1467-971X.2005.00408.x

Doye, P. (1991). *Authenticity in Foreign Language Testing [microform] / Peter Doye*. [Washington, D.C.]: Distributed by ERIC Clearinghouse.

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

Dunkel, P. A. (1986). Developing Listening Fluency in L2: Theoretical Principles and Pedagogical Considerations. *Modern Language Journal, 70*(2), 99-106.

Dunlop, I. A. N. (1981). READING AND COMMUNICATION: USING AUTHENTIC MATERIALS*. *World Englishes, 1*(1), 50-55. doi: 10.1111/j.1467-971X.1981.tb00450.x

Duquette, G., & et al. (1987). The Effect of Authentic Materials in Acquiring a Second Language. *Canadian Modern Language Review, 43*(3), 479-492.

Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press.

Fernandez-Toro, M., & Jones, F. R. (1996). Going solo: Learners' experiences of self-instruction and self-instruction training. *Promoting learner autonomy in university language teaching*, 185-214.

Flowerdew, J. (1993). Concordancing as a tool in course design. *System, 21*(2), 231-244. doi: 10.1016/0346-251x(93)90044-h

Flowerdew, J. (2001). Concordancing as a tool in course design. In M. Ghadessy, A. Henry & R. L. Roseberry (Eds.), *Small Corpus Studies and Elt: Theory and Practice* (pp. 71-93). Amsterdam: J. Benjamins Publishing Company.

Gabrielatos, C. (1994). *Collocations, pedagogical implications and their treatment in pedagogical materials.* Ms, Research Centre for English and Applied Linguistics, University of Cambridge.

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.

Gabrielatos, C. (2006, 1-4 July 2006). *Corpus-based evaluation of pedagogical materials: If-conditionals in ELT coursebooks and the BNC.* Paper presented at the 7th Teaching and Language Corpora Conference (TALC 2006), University Paris 7 Denis Diderot, Paris, France.

Gajic, N. (2010). *Materials Development: Corpus Based Study of Authentic Texts for Business.* Master of Arts, San Diego State University.

Garinger, D. (2001). Textbook Selection for the ESL Classroom.

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

González, O. G. (1990). *Teaching Language and Culture with Authentic Materials.* Dissertation Abstracts International. Retrieved from http://books.google.com.tr/books?id=TqsfNQAACAAJ

Graddol, D. (1997). *The future of English? A guide to forecasting the popularity of English in the 21st century?* London: British Council.

Granger, S. (1998). The computer learner corpus: a versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer* (pp. 3 - 18). London: Longman.

Granger, S. (1999). Use of tenses in advanced EFL learners: Evidence from an error-tagged computer corpus. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora- studies in honour of Stig Johannson* (pp. 191-202). Rodopi: Amsterdam & Atlanta.

Guariento, W., & Morley, J. (2001). Text and task authenticity in the EFL classroom. *ELT Journal, 55*(4), 347-353. doi: 10.1093/elt/55.4.347

Guillot, M.-N. (2002). Corpus-based work and discourse analysis in FL pedagogy: a reassessment. *System, 30*(1), 15-32. doi: 10.1016/s0346-251x(01)00054-9

Hadley, A. O. (2001). *Teaching Language in Context*: Heinle & Heinle.

Harmer, J. (2001). *The practice of English language teaching* (3rd ed.). Harlow: Longman.

Harwood, N. (2005). What do we want EAP teaching materials for? *Journal of English for Academic Purposes, 4*(2), 149-161. doi: 10.1016/j.jeap.2004.07.008

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

Hedge, T. (2000). *Teaching and learning in the language classroom*. Oxford: Oxford University Press.

Henry, S. (2009). *The Practical Study of Languages; a Guide for Teachers and Learners*: BiblioBazaar.

Herron, C. A., & Seay, I. (1991). The Effect of Authentic Oral Texts on Student Listening Comprehension in the Foreign-Language Classroom. *Foreign Language Annals, 24*(6), 487-495.

Hill, V. J. (1996). Verb-form clustering and syllabus design. *System, 24*(4), 529-536. doi: 10.1016/s0346-251x(96)00047-4

Hillyard, L., Reppen, R., & Vásquez, C. (2007). Bringing the outside world into an intensive English programme. *ELT Journal, 61*(2), 126-134. doi: 10.1093/elt/ccm005

Hindle, D. (1994). A Parser for Text Corpora. In A. Zampolli (Ed.), *Computational Approaches to the Lexicon* (pp. 103-151). New York: Oxford University Press.

Holmes, J. (1988). Doubt and Certainty in Esl Textbooks. *Applied Linguistics, 9*(1), 21-44.

House, J. (2003). English as a lingua franca: A threat to multilingualism? *Journal of Sociolinguistics, 7*(4), 556-578.

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

HUNSTON, S., & GILL, F. (1998). Verbs Observed: A Corpus-driven Pedagogic Grammar1. *Applied Linguistics, 19*(1), 45-72. doi: 10.1093/applin/19.1.45

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learning-centred approach*: Cambridge Univ Pr.

Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for Specific Purposes, 13*(3), 239-256. doi: citeulike-article-id:3620499

Johnson, K., & Morrow, K. (1981). *Communication in the Classroom: Applications and Methods for a Communicative Approach*: Longman.

Kennedy, G. D. (1998). *An introduction to corpus linguistics*: Longman.

Kienbaum, B. E., & et al. (1986). Communicative Competence in Foreign Language Learning with Authentic Materials. Final Project Report.

Kilickaya, F. (2004). Authentic materials and cultural content in EFL classrooms. *The Internet TESL Journal, X*(7). Retrieved from http://iteslj.org/Techniques/Kilickaya-AutenticMaterial.html

Kim, D.-K. (2000). A Qualitative Approach to the Authenticity in the Foreign Language Classroom: A Study of University Students Learning English in Korea. *Texas Papers in Foreign Language Education, 5*(1), 189-205.

King, C. P. (1990). A Linguistic and A Cultural Competence: Can They Live Happily Together? *Foreign Language Annals, 23*(1), 65-70. doi: 10.1111/j.1944-9720.1990.tb00341.x

Konig, G. C. (1990). The Place of English in Turkey *The Birth and Growth of a Department: Department of English Language and Literature:25th Anniversary* (pp. 157-167). Ankara: Hacettepe U.

Koprowski, M. (2005). Investigating the usefulness of lexical phrases in contemporary coursebooks. *ELT Journal, 59*(4), 322-332. doi: 10.1093/elt/cci061

Kramsch, C. (1994). Language as Discourse - Perspectives for Language Teaching - Mccarthy,M, Carter,R. *Modern Language Journal, 78*(4), 560-560.

Krashen, S. D. (1981). *Second language acquisition and second language learning*: Pergamon Press.

Krashen, S. D. (1985). *The input hypothesis : issues and implications*. London: Longman.

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

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. London: Longman.

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

Lewkowicz, J. A. (2000). Authenticity in language testing: some outstanding questions. *Language testing, 17*(1), 43-.

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

Little, D., & Singleton, D. (1991). Authentic texts, pedagogical grammar and language awareness in foreign language learning. In C. James & P. Garrett (Eds.), *Language awareness in the classroom* (pp. 123-132). London: Longman.

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

Martinez, A. (2002). Authentic materials: An overview. Retrieved from http://www3.telus.net/linguisticsissues/authenticmaterials.html website:

Masgoret, A. M., & Gardner, R. C. (2003). Attitudes, Motivation, and Second Language Learning: A Meta-Analysis of Studies Conducted by Gardner and Associates. *Language Learning, 53*(S1), 167-210. doi: 10.1111/1467-9922.00227

Maxim, I. I. I. H. (2002). A Study into the Feasibility and Effects of Reading Extended Authentic Discourse in the Beginning German Language Classroom. *The Modern Language Journal, 86*(1), 20-35. doi: 10.1111/1540-4781.00134

McCarthy, M., & Carter, R. (1994). *Language as discourse : perspectives for language teaching*. London: Longman.

McEnery, T., & Wilson, A. (1996). *Corpus linguistics*: Edinburgh University Press.

McGarry, D. (1995). *The role of authentic texts*. Dublin: Authentik Language Learning Resources.

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

McNeill, A. (1994). *What Makes Authentic Materials Different? The Case of English Language Materials for Educational Television.* Paper presented at the the Annual International Language in Education Conference, Hong Kong.

Mera Rivas, R. M. (1999). Reading in recent ELT coursebooks. *ELT Journal, 53*(1), 12-21. doi: 10.1093/elt/53.1.12

Meunier, F., Gouverneur, C., & line. (2007). The treatment of phraseology in ELT textbooks. *Language and Computers, 61*(1), 119-139.

Meyer, C. F. (2002). *English Corpus Linguistics*: Cambridge University Press.

Mihwa, Y. (1994). *The influence of authentic versus edited texts on the reading comprehension and online reading behaviors of beginning Korean and Japanese ESL students.* 55 (10), Dissertation Abstracts International.

Miller, M. (2005). *Improving aural comprehension skills in EFL, using authentic materials: an experiment with university students in Nigata, Japan.* Unpublished master's thesis, University of Surrey, Guildford, UK

Mindt, D. (1997). Corpora and the teaching of English in Germany. In A. Wichmann, S. Fligelstone, A. M. McEnery & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 40-50). London: Longman.

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

Miura, T. (1997). *An Analysis of Aural/Oral Communications A. English Textbooks in Japanese Upper Secondary School.* MA thesis, The University of Birmingham.

Morrison, B. (1989). Using news broadcasts for authentic listening comprehension. *ELT Journal, 43*(1), 14-18. doi: 10.1093/elt/43.1.14

Morrow, K. (1977). Authentic texts in ESP. *English for specific purposes. London: Modern English Publications*, 13-17.

Nitta, R., & Gardner, S. (2005). Consciousness-raising and practice in ELT coursebooks. *ELT Journal, 59*(1), 3-13. doi: 10.1093/elt/cci001

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

Nunan, D. (1991a). Communicative Tasks and the Language Curriculum. *Tesol Quarterly, 25*(2), 279-295.

Nunan, D. (1991b). *Language teaching methodology: A textbook for teachers*. New York: Prentice Hall.

Otte, J. (2006). *Real language to real people: a descriptive and exploratory case study of the outcomes of aural authentic texts on the listening comprehension of adult ESL students enrolled in an advanced ESL listening course.* Master of Arts, Loyola University, Dissertation Abstracts International.

OWEN, C. (1993). Corpus-Based Grammar and the Heineken Effect: Lexico-grammatical Description for Language Learners1. *Applied Linguistics, 14*(2), 167-187. doi: 10.1093/applin/14.2.167

Ozgen, M. (2008). *The use of authentic captionedn video as listening comprehension material in English language teaching.* Master of Arts, Selcuk University, Konya.

Peacock, M. (1997). The effect of authentic materials on the motivation of EFL learners. *ELT Journal, 51*(2), 144-156. doi: 10.1093/elt/51.2.144

Porter, D., & Roberts, J. (1981). Authentic listening activities1. *ELT Journal, 36*(1), 37-1.

Ranalli, J. (2003). *ELT coursebooks in the age of corpus linguistics: constraints and possibilities* PhD, Birmingham.

Ranalli, J. M. (2003). *The Treatment of Key Vocabulary Strategies in Current ELT Coursebooks: Repetition, Resource Use, Recording.* MA thesis, University of Birmingham.

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

Rayson, P., & Garside, R. (2000, 1-8 October 2000). *Comparing corpora using frequency profiling.* Paper presented at the the workshop on Comparing Corpora held in conjunction with the 38th annual meeting of the Association for Computational Linguistics (ACL 2000), Hong Kong.

Reda, G. (2003). English coursebooks: prototype texts and basic vocabulary norms. *ELT Journal, 57*(3), 260-268. doi: 10.1093/elt/57.3.260

Richards, J. C. (2001). *Curriculum Development in Language Teaching*: Cambridge University Press.

Rings, L. (1986). Authentic Language and Authentic Conversational Texts. *Foreign Language Annals, 19*(3), 203-208.

Rivers, W. M. (1987). *Interactive Language Teaching*: Cambridge University Press.

Rogers, C. V., & Medley, F. W. (1988). Language With A Purpose: Using Authentic Materials in the Foreign Language Classroom. *Foreign Language Annals, 21*(5), 467-478. doi: 10.1111/j.1944-9720.1988.tb01098.x

Römer, U. (2004a). Comparing real and ideal language learner input: the use of an EFL textbook corpus in corpus linguistics and language teaching. *Corpora and language learners*, 151-168.

Römer, U. (2004b). Textbooks: A corpus-driven approach to modal auxiliaries and their didactics. In J. M. H. Sinclair (Ed.), *How To Use Corpora In Language Teaching* (pp. 185-199). Amsterdam: J. Benjamins.

Römer, U. (2005). *Progressives, Patterns. Pedagogy: A Corpus-driven Approach to English Progressive Forms, Functions, Contexts, and Didactics*: J. Benjamins Publishing Company.

Samimy, K. K., & Tabuse, M. (1992). Affective Variables and a Less Commonly Taught Language: A Study in Beginning Japanese Classes*. *Language Learning, 42*(3), 377-398. doi: 10.1111/j.1467-1770.1992.tb01341.x

Sardinha, A. P. B. (1999). Beginning Portuguese corpus linguistics: exploring a corpus to teach Portuguese as a foreign language. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada, 15*, 289-299.

Schiffrin, D. (1996). Interactional sociolinguistics. In S. L. McKay, Hornberger, N.H. (Ed.), *Sociolinguistics and language teaching* (pp. 307–328). Cambridge: Cambridge University Press.

Schmidt, T. Y. (1994). *Authenticity in ESL: A Study of Requests.* Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED415686

Schulz, R. A. (1991). Second Language 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

Scott, M. (2000). *Focusing on the text and its key words.* Paper presented at the Rethinking language pedagogy from a corpus perspective:the third international conference on teaching and language corpora, Frankfurt.

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

Seidlhofer, J. (2005). *ELT Journal, 59/4*(October 2005).

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

Sherman, J. (2003). *Using Authentic Video in the Language Classroom*: Cambridge University Press.

Shook, D. J. (1997). Identifying and overcoming possible mismatches in the beginning reader-literary text interaction. *Hispania*, 234-243.

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.

Swales, J. (1980). ESP: The textbook problem. *The ESP Journal, 1*(1), 11-23. doi: 10.1016/0272-2380(80)90006-2

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

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

Taylor, D. (1994). Inauthentic authenticity or authentic inauthenticity. *TESL-EJ, 1*(2), 1-11.

Terrell, T. D. (1993). Comprehensible Input for Intermediate Foreign Language Students via Video. *IALL Journal of Language Learning Technologies, 26*(2), 17-23.

Thanajaro, M. (2000). *Using authentic materials to develop listening comprehension in the English as a foreign language classroom.* Unpublished doctorial dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

Thanasoulas, D. (2005). Discourse for Teaching Purposes, from http://www.developingteachers.com/articles_tchtraining/discourse2dimitrios.htm

Tılfarlıoğlu, F. Y., & Öztürk, A. R. (2007). An Analysis of ELT Teachers' Perceptions of Some Problems Concerning the Implementation of English Language Teaching Curricula in Elementary Schools. *Journal of Language and Linguistic Studies, 3/1*(April).

Titone, R. (1968). *Teaching foreign languages : an historical sketch*. Washington: Georgetown University Press.

Torun, F. P. (2008). *Contextualized language instruction:Exploring the role of authentic animated stories on the attitudes of young EFL learners towards learning English.* Master of Arts, Cukurova University.

Ur, P. (1996). *A Course in Language Teaching: Practice of Theory*: Cambridge University Press.

Van Lier, L. (1996). *Interaction in the language curriculum : awareness, autonomy, and authenticity*. London: Longman.

Vellenga, H. (2004). Learning pragmatics from ESL & EFL textbooks: How likely? *TESL-EJ, 8*(2). Retrieved from http://tesl-ej.org/ej30/a3.html

Walz, J. (1989). Context and Contextualized Language Practice in Foreign Language Teaching. *Modern Language Journal, 73*(2), 160-168.

Wardhaugh, R. (1986). *An introduction to sociolinguistics*. Oxford: Basil Blackwell.

Wen, C.-f. (1989). Student Attitudes during a Course with and without TV in the Language Laboratory. *Journal of Educational Media and Library Sciences, 26*(3), 238-250.

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

Widdowson, H. G. (1979). *Explorations in applied linguistics*. Oxford: Oxford University Press.

Widdowson, H. G. (1983). *Learning purpose and language use*. Oxford ; New York: Oxford University Press.

Widdowson, H. G. (1996). Comment: authenticity and autonomy in ELT. *ELT Journal, 50*(1), 67-68. doi: 10.1093/elt/50.1.67

Widdowson, H. G. (1998). Context, community, and authentic language". *Tesol Quarterly, 32*(4), 705-716.

Wilkins, D. A. (1976). *Notional syllabuses : a taxonomy and its relevance to foreign language curriculum development*. London: Oxford University Press.

Wipf, J. A. (1984). Shortwave Radio and the Second Language Class. *The Modern Language Journal, 68*(1), 7-12. doi: 10.1111/j.1540-4781.1984.tb01537.x

Xu, I. H. (2004). *Investigating Criteria for Assessing ESL Textbooks*: University of Alberta.

Young, D. J. (1999). Linguistic simplification of SL reading material: Effective instructional practice? *Modern Language Journal, 83*(3), 350-366.

Zok, D. (2010). Turkey's Language Revolution and the Status of English Today.

APPENDICES

Appendix 1. Sketch Engine Data Query Formula for Tenses

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
Present Perfect Continuous Tense	([tag="VHB.*"] | [tag="VHZ.*"]) [tag="VBN.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VDG.*"] | [tag="VHG.*"])
Present Perfect Continuous Tense Negative	([tag="VHB.*"] | [tag="VHZ.*"]) [tag="XX0.*"] [tag="VBN.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VDG.*"] | [tag="VHG.*"])
Past Perfect Tense	[tag="VHD.*"] ([tag="VVN.*"] | [tag="VDN.*"] | [tag="VBN.*"] | [tag="VHN.*"])

Past Perfect Tense Negative	[tag="VHD.*"] [tag="XX0.*"]
Past Perfect Continuous Tense	[tag="VHD.*"] [tag="VBN.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VDG.*"] | [tag="VHG.*"])
Past Perfect Continuous Tense Negative	[tag="VHD.*"] [tag="XX0.*"] [tag="VBN.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VDG.*"] | [tag="VHG.*"])
For TEFL CC	
Simple Present Tense	[tag="VVP.*"] | [tag="VVZ.*"] | [tag="VHP.*"] | [tag="VHZ.*"]
Simple Present Tense Negative	("(?I)do" | "(?i)does") ("(?i)not" | "(?i)n't")
Present Continuous Tense	([tag="VBP.*"] | [tag="VBZ.*"]) ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VHG.*"])
Present Continuous Tense Negative	([tag="VBP.*"] | [tag="VBZ.*"]) ("(?i)not" | "(?i)not") ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VHG.*"])
Simple Past Tense	[tag="VVD.*"] | [tag="VHD.*"]
Simple Past Tense Negative	("(?i)did") ("(?i)not" | "(?i)n't")
Past Continuous Tense	[tag="VBD.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VHG.*"])
Past Continuous Tense Negative	[tag="VBD.*"] ("(?i)not" | "(?i)n't") ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VHG.*"])
Future Tense	[lemma="will*."]
Future Tense Negative	([lemma="will*."] | [lemma="wo"]) ("(?i)not" | "(?i)n't")
Future Continuous Tense	[lemma="will*."] [tag="VB.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VHG.*"])
Future Continuous Tense Negative	([lemma="will*."] | [lemma="wo"]) ("(?i)not" | "(?i)n't") [tag="VB.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VHG.*"])
Future Perfect Tense	[lemma="will*."] [tag="VH.*"] ([tag="VVN.*"] | [tag="VBN.*"] | [tag="VHN.*"])
Future Perfect Tense Negative	([lemma="will*."] | [lemma="wo"]) ("(?i)not" | "(?i)n't") [tag="VH.*"] ([tag="VVN.*"] | [tag="VBN.*"] | [tag="VHN.*"])
Future Perfect Continuous Tense	[lemma="will*."] [tag="VH.*"] [tag="VBN.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VHG.*"])
Future Perfect Continuous Tense Negative	[lemma="will*."] ("(?i)not" | "(?i)n't") [tag="VH.*"] [tag="VBN.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VHG.*"])
Present Perfect Tense	([tag="VHP.*"] | [tag="VHZ.*"]) ([tag="VVN.*"] | [tag="VBN.*"] | [tag="VHN.*"])
Present Perfect Tense Negative	([tag="VHP.*"] | [tag="VHZ.*"]) ("(?i)not" | "(?i)n't")
Present Perfect Continuous Tense	([tag="VHP.*"] | [tag="VHZ.*"]) [tag="VBN.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VHG.*"])
Present Perfect Continuous Tense Negative	([tag="VHP.*"] | [tag="VHZ.*"]) ("(?i)not" | "(?i)n't") [tag="VBN.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VHG.*"])
Past Perfect Tense	[tag="VHD.*"] ([tag="VVN.*"] | [tag="VBN.*"] | [tag="VHN.*"])

Past Perfect Tense Negative	[tag="VHD.*"] ("(?i)not" | "(?i)n't")
Past Perfect Continuous Tense	[tag="VHD.*"] [tag="VBN.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VHG.*"])
Past Perfect Continuous Tense Negative	[tag="VHD.*"] ("(?i)not" | "(?i)n't") [tag="VBN.*"] ([tag="VVG.*"] | [tag="VBG.*"] | [tag="VHG.*"])

Appendix 2. Sketch Engine Data Query Formula for Modals

For BNC-Spoken	
Can	[tag="VM0*." & lemma="can" & word != "ca*" & word != "Ca*"]
Cannot	[tag="VM0*." & lemma="can"] [tag="XX0*."]
May	[tag="VM0*." & lemma="may"]
May not	[tag="VM0*." & lemma="may"] [tag="XX0*."]
might	[tag="VM0*." & lemma="might"]
Might not	[tag="VM0*." & lemma="might"] [tag="XX0*."]
Will	[tag="VM0*." & lemma= "will.*" & word != "wo.*" & word != "Wo.*"]
Will not	[lemma="will*."] [tag="XX0*."]
Could	[tag="VM0*." & lemma="could"]
Could not	[tag="VM0*." & lemma="could"] [tag="XX0*."]
Would	[tag="VM0*." & lemma="would"]
Would not	[tag="VM0*." & lemma="would"] [tag="XX0*."]
Should	[tag="VM0*." & lemma="should"]
Should not	[tag="VM0*." & lemma="should"] [tag="XX0*."]
Must	[tag="VM0*." & lemma="must"]
Must not	[tag="VM0*." & lemma="must"] [tag="XX0*."]
Shall	[tag="VM0*." & lemma="shall"]
Shall not	[tag="VM0*." & lemma="shall"] [tag="XX0*."]
Ought to	[tag="VM0*." & lemma="ought"]
Ought not to	[tag="VM0*." & lemma="ought"] [tag="XX0*."]
Need	[tag="VM0*." & lemma="need"]
Need not	[tag="VM0*." & lemma="need"] [tag="XX0*."]
Dare	[tag="VM0*." & lemma="dare"]
Dare not	[tag="VM0*." & lemma="dare"] [tag="XX0*."]
For TEFL CC	
Can	[tag="MD*." & word="(?i)can.*"]
Cannot	([tag="MD" & word="(?i)ca"] [tag="RB"]) | [word="(?i)cannot"]
May	[tag="MD*." & word="(?i)may.*"]
May not	[tag="MD*." & word="(?i)may.*"] ("(?i)not" | "(?i)n't")
might	[tag="MD*." & word="(?i)might"]
Might not	[tag="MD*." & word="(?i)might"] ("(?i)not" | "(?i)n't")
Will	[lemma="(?i)will*." & tag="MD*" & word !="'d.*"]
Will not	([lemma="(?i)will*." & tag="MD*" & word !="'d.*"] | [word="wo"]) ("(?i)not" | "(?i)n't")
Could	[tag="MD" & word="(?i)could"]
Could not	[tag="MD" & word="(?i)could"] ("(?i)not" | "(?i)n't")
Would	[tag="MD" & word="(?i)would"]
Would not	[tag="MD" & word="(?i)would"] ("(?i)not" | "(?i)n't")
Should	[tag="MD" & word="(?i)should"]

Should not	[tag="MD" & word="(?i)should"] ("(?i)not" | "(?i)n't")
Must	[tag="MD" & word="(?i)must"]
Must not	[tag="MD" & word="(?i)must"] ("(?i)not" | "(?i)n't")
Shall	[tag="MD" & word="(?i)shall"]
Shall not	[tag="VM0*." & lemma="shall"] [tag="XX0*."]
Ought to	[tag="MD" & word="(?i)ought"]
Ought not to	[tag="MD" & word="(?i)ought"] ("(?i)not" | "(?i)n't")
Need	[tag="MD" & word="(?i)need"]
Need not	[tag="MD" & word="(?i)need"] ("(?i)not" | "(?i)n't")
Dare	[tag="MD" & word="(?i)dare"]
Dare not	[tag="MD" & word="(?i)dare"] ("(?i)not" | "(?i)n't")

CURRICULUM VITAE

PERSONAL DETAILS

Name **:** Emrah PEKSOY

Place and Date of Birth **:** Andırın – March 02, 1985

Marital Status **:** Married

E-mail **:** e.peksoy@gmail.com

EDUCATIONAL BACKGROUND

2010- **:** (MA) Dicle University, Institute of Social Sciences,
Foreign Language Teaching Department, Diyarbakır

2003-2007 **:** (BA) Middle East Technical University, Faculty of Education
English Language Teaching Department, Ankara

1999-2003 **:** Mimar Sinan Anadolu Öğretmen Lisesi, Pınarbaşı/Kayseri

WORK EXPERIENCE

2009 - ….. **:** Kilis 7 Aralık University (English Language Instructor)

2007-2009 **:** Özel Büyükkoyuncu Anadolu Lisesi (English Teacher), Konya