

The use of *if*-conditionals in English: a comparative study of a Finnish upper secondary school textbook series, *the British National Corpus* and *the International Corpus of Learner English*

University of Tampere
School of Modern Languages
and Translation Studies
English Philology
Pro Gradu Thesis
May 2007
Virpi Turunen

Tampereen yliopisto

Kieli- ja käännöstieteiden laitos

Englantilainen filologia

Turunen, Virpi: The use of *if*-conditionals in English: a comparative study of a Finnish upper secondary school textbook series, *the British National Corpus* and *the International Corpus of Learner English*

Pro Gradu-tutkielma, 59 sivua + 3 liitettä

Toukokuu 2007

Tiivistelmä

Tämä pro gradu-tutkielma tarkastelee *if*-konditionaalien käyttöä englannin kielessä. Tutkimus toteutettiin vertaamalla *if*-konditionaalien esiintymistä suomalaisessa lukion *Culture Café*-oppikirjasarjassa sekä kahdessa korpuksessa, jotka olivat *the British National Corpus* ja *the International Corpus of Learner English*. Tarkoituksena oli selvittää, missä määrin oppikirjojen sisältö vastaa korpustutkimuksen tuloksia keskittyen *if*-konditionaalien esiintymistiheyteen ja luokitteluun ja siten päätellä, kuinka autenttisenä kyseisen konditionaalirakenteen käyttöä voidaan pitää näissä oppikirjoissa.

Tutkielman tavoitteena oli tarkastella *if*-konditionaalien esiintymistä puhutussa ja kirjoitetussa englannin kielessä (*the British National Corpus*) sekä suomalaisten englannin kielen opiskelijoiden käyttämässä kirjoitetussa kielessä (*the International Corpus of Learner English*). Kirjasarja, johon korpuksia verrattiin, koostuu kahdeksasta eri kirjasta. *BNC*-korpus on kokoelma sähköisessä muodossa olevia tekstejä, jotka edustavat erilaisia tekstityyppejä puhutusta ja kirjoitetusta englannin kielestä. Sekä puhutusta että kirjoitetusta aineistosta valittiin satunnaisotannalla 250 *if*-konditionaalilauseetta tutkimusta varten. *ICLE*-korpus sisältää 19:ää eri kansalaisuutta edustavien englannin kielen opiskelijoiden kirjoittamia kirjoitelmia. Tästä korpuksesta tutkimuksen kohteeksi valittiin suomalainen alakorpus, josta tutkittiin 197:ää *if*-konditionaalia.

If-konditionaaleja analysoitiin sen perusteella, minkälaisia verbiyhdistelmiä ne sisälsivät pää- ja sivulauseissa. Konditionaalilauseita tutkittiin sekä formaalisten (muodollisten,kieliopillisten) että semanttisten (merkitysopillisten) ominaisuuksien kannalta.

Tutkimuksen tulokset osoittavat, että korpuksissa esiintyvä autenttinen materiaali sisältää huomattavasti suuremman määrän erilaisia *if*-konditionaalirakenteita kuin tutkitut oppikirjat. Lisäksi oppikirjoissa ilmenneiden konditionaalirakenteiden suhteellinen esiintymistiheys ei vastaa todellisessa kielenkäytössä käytettyjen rakenteiden esiintymistiheyttä, vaan oppikirjoissa tunnutaan painottavan tiettyjä yksittäisiä rakenteita. Semanttiselta kannalta tarkasteltuna oppikirjojen ja korpusten tulokset eivät eroa kovinkaan paljon toisistaan. Tuloksista käy siis ilmi, että *if*-konditionaalirakenteiden kannalta *Culture Café*-kirjasarjaa voidaan pitää semanttisesti, mutta ei formaalisesti autenttisenä.

Asiasanat: autenttisuus, *if*-konditionaali, korpus, oppikirja

Table of Contents

1. Introduction	1
1.1. Objectives and scope of the study	2
1.2. Structure of the study	3
1.3. Textbook authenticity	4
1.4. Previous studies	6
2. Formal aspects of <i>if</i> -conditional constructions	14
3. Semantic aspects of <i>if</i> -conditional constructions	17
4. Materials	20
4.1. <i>Culture Café</i> - textbook series	21
4.2. <i>The British National Corpus</i>	22
4.3. <i>The International Corpus of Learner English</i>	24
5. Methods	28
5.1. Corpus based vs. corpus-driven approaches	28
5.2. The quantitative paradigm	30
5.3. Collection of materials	32
5.4. Methods of analysis	33
6. Findings	35
6.1. Analysis of formal aspects	35
6.2. Analysis of semantic aspects	49
6.3. Summary of findings	52
7. Conclusion	54
Bibliography	57
Appendices	60

1. Introduction

Conditionals form a part of the English language that seems relatively difficult for many learners of English as a second or foreign language. Since conditionals appear in various forms and may convey numerous different meanings, they constitute a complicated area of the English grammar system that may seem utterly troublesome not only for students but also for some teachers of English. Although many researchers have studied conditionals and created different kinds of methods for teaching them, this semantically and contextually complex linguistic construction may still remain a puzzle for teachers in practical classroom situations.

The present tendency in English language teaching is towards using authentic materials instead of English that is considered standard or governed by normative rules. The task of bringing authentic language into foreign language classrooms is to a large extent dependent on the authors and producers of different textbook series. Bridging the gap between authentic language and textbook language is a challenge that requires a careful analysis of the various features of language represented in the textbooks. For comparison, large amounts of authentic language data need to be examined to gain a comprehensive insight into the nature of the linguistic construction under investigation.

The concept of authenticity has become a subject of discussion in the pedagogic circles and therefore present and future language teachers are inevitably expected to become aware of the nature of authenticity, to apply new approaches in their work and to introduce authenticity to their students, particularly in the form of textbooks. As a future language teacher this topic has relevance to my work as well and for this reason I decided to study textbook authenticity in particular and to focus on the use of *if*-conditionals.

1.1. Objectives and scope of the study

The purpose of the present study is to consider the most essential properties of *if*-conditional constructions in English and to examine these constructions as they appear in a selected series of English textbooks and additionally, to compare this occurrence to their actual realization in real language use by carrying out two corpus studies.

The research question of this comparative study is whether there is a correlation between the occurrence (i.e the frequencies and classifications) of conditionals in the textbooks and in the corpora. Thus, the study analyses whether the language used in the textbooks can be considered authentic when it comes to this specific linguistic feature.

Conditionals serve as a good example of a grammatical structure that may prove to be relatively difficult for Finnish learners of English on many levels. Firstly, whereas many different verb-forms can occur in conditional sentences in English, the equivalent structure in the Finnish language is always the same (*-isi*). Secondly, the semantics of the conditional can vary substantially in the English language, as can be noted in chapter 3 where the semantic aspects of conditionals are considered in more detail. The Finnish conditional structure does not denote this kind of variation of meaning.

The reason why this study is only interested in *if*-conditionals is based mainly on practical issues. There are, naturally, many more types of conditional constructions that could be studied (such as *in case*, *on condition that*, *unless* and conditionals with inversion) but it would require a far more comprehensible and extensive study to analyse all the potential structures. In addition, such a variety of constructions would be very complicated, if not impossible, to be processed with the corpus tools available. For this reason, this small-scale study is limited to the examination of *if*-conditional constructions.

Since this study is concerned with *if*-conditionals in particular, other conditional forms will not be included in the study at all. The main focus is on the verb-form combinations that constitute the conditional sentences. Basically, the *if*-clause and the main clause may both contain almost any type of verb form: auxiliaries, simple present, simple past, present perfect, past perfect, present continuous, imperative etc. The diversity of forms is particularly clear in spoken language where even many non-standard verb-form combinations are frequently used.

Considering the wide variety of forms that may constitute a conditional sentence and an *if*-conditional in particular, it can be concluded that although there are certain forms that are considered more standard than others, there are still many forms that are frequently used outside the scope of school English in everyday language. The different verb-form combinations that are presented in the materials are listed in their entirety in chapter 6 where the findings of the present study are examined more closely.

1.2. Structure of the study

Section 1.3 elaborates on textbook authenticity, a concept which is at the centre of textbook analysis. It is followed by a description of the preceding studies relevant to this area of research. Chapter 2 introduces the formal aspects and chapter 3 the semantic aspects of *if*-conditional constructions. Chapter 4 focuses on presenting the materials under investigation: an upper secondary school English textbook series called *Culture Café, the British National Corpus* and *the International Corpus of Learner English*. The aim is to introduce the most essential properties of the materials and to justify the selection of this particular material. Chapter 5 focuses on the methodology applied in this study and explains the research process

starting from the collection of materials and proceeding to the introduction of the methods of analysis. In addition, this chapter introduces the concepts of corpus based and corpus-driven approaches and presents the most important characteristics of the quantitative paradigm. The findings of the study will be discussed in chapter 6. The conclusion summarises the main points of the study and gives some suggestions for future research.

1.3. Textbook authenticity

One of the underlying intentions behind this study is to consider the authenticity of the selected textbook series with special attention drawn to the appearance of *if*-conditional constructions. Therefore, it is necessary to introduce the concept of textbook authenticity as an essential motivation behind this study. The ultimate aim of the study is, after all, to gain more information on how *if*-conditionals are presented in the textbooks and how well this presentation corresponds to the language used in real-life situations.

Mauranen (2004, 91) defines the concept of authenticity either in a purely factual sense as meaning 'genuine', or as something so similar to the real thing that it is almost like the thing itself. Still, the word 'authentic' is a relative concept. Mauranen is of the opinion that any record of language that is used outside the classroom that is brought to learners inside the classroom is necessarily severed from its original context and is inauthentic in the sense of not being 'genuine'. According to Widdowson (2000, 7), on the other hand, in a pedagogic context authenticity is not a characteristic of texts as such but one which users bring to texts through the process of recontextualisation. Therefore, the meaning of the concept is largely dependent on the person's view of the language and the process of language learning as well as on the classroom environment.

According to Römer (2005, 277), “The issue of whether to use artificial, invented or natural, authentic language examples in coursebooks and reference works provides plenty of food for discussion among linguists” . Still, the matter of authenticity is not only of interest to linguistic scholars, but most closely to teachers and other people who work in the pedagogic sphere on a daily basis and are in close collaboration with the actual target audience of the teaching materials, the students.

Some think that using authentic material does not put enough emphasis on the subject or grammatical construction being taught and that text material invented for this specific purpose is much more convenient since it can be elaborated almost infinitely to serve the purposes of the teacher. In addition, using invented language also allows the students to work with flawless and functional language without having to encounter mistakes or discrepancies. Others, on the other hand, find that if students are not equipped with real-life language examples they will not be able to use language competently in authentic situations. Römer (2005, 280) points out that “Pupils and students have to learn how to deal with “messiness” and “untidiness” in language, because in real-life communicative contexts they are unlikely to encounter the kind of well-formed sentences that their EFL coursebooks present them with”.

Using authentic texts in coursebooks is a reliable way to make sure that students are taught how the language actually works and how it is used by real people in real situations. By familiarising themselves with genuine language, students are given the opportunity to discover the underlying conventions behind the target language in more natural contexts. Still, it should be taken into consideration that the starting point in any learning situation should be the students, not the texts themselves. This is what the authors and publishers of textbooks should also note when considering the students’ interests in the process of manufacturing teaching materials.

In this study, the term ‘authentic’ is used to refer to the materials collected from the BNC corpus. ‘Authentic’ is, therefore, restricted to the materials provided by native speakers but not to the materials written by Finnish learners of English. Some might argue that also the language produced by non-native speakers of English could be considered authentic since it is produced by real people in real-life situations. Still, this study aims to emphasize the differences and similarities between the language used by native speakers of English, the language presented in the Finnish textbooks of English and the language produced by learners of English. Therefore, the study takes as a starting point the authenticity of the language used by native speakers of English and compares the language represented in the textbooks and in the learner corpus in relation to this authentic native speaker language.

1.4. Previous studies

Perhaps the best way to start the discussion on the use of conditionals in textbooks and their relation to actual language use is to look at the earlier studies conducted in similar areas of research. Since the interest of the present study is not entirely limited to textbook analysis or, on the other hand, entirely to corpus study, the sources used in this study include some that are directly concerned with textbook analysis and grammar specifically, and others that have to do with corpus study.

One of the most important sources used in the present study is Ute Römer’s (2005) study *Progressives, Patterns, Pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics*. The book offers interesting guidelines as to how a corpus-driven study of a grammatical feature can be carried out and therefore, it plays an important role in the general theoretical framework and the underlying research setting in the

study of the corpora utilized in this study. In her study, Römer concentrates on the examination of progressive forms by the means of a comparative analysis based on a corpora of spoken British English and a small corpus of German EFL textbooks. Although Römer's study is far more comprehensive than this study and includes additionally a number of pedagogical implications which are not included in the present study, Römer's study served as a model and as an important source of information throughout the research process.

In her book, Römer (2005, 1) argues that most studies conducted to analyse the use of progressive forms lack a broad empirical basis. Instead, she says that theoretical conclusions are drawn from samples of "rather intuitive examples which have been invented for the purpose of illustrating a particular point, or which have been taken from earlier linguistic studies and grammar books" (Römer 2005, 1). Therefore, she concludes, it is necessary to conduct a study based on systematic empirical analysis including large amounts of quantitative language data.

Römer (2005, 1-2) defines the following as the basic study questions of her research: How are progressives actually used in spoken English? How are their different forms distributed? In which lexical or syntactical contexts do they usually occur? What do different progressive forms typically express? and Is it possible to identify a "generally valid" behaviour of progressives, or do different verbs show largely dissimilar context and function patterns?

To answer the study questions, Römer analyses 10,171 progressive verb forms in context presented in the spoken component of *the British National Corpus* and the spoken British subsection of *the Bank of English*. In addition, she includes in her study an analysis of two most widely used and most recently published textbook series in Germany in order to find out whether any significant differences exist between the use of the progressive constructions in real English and so-called "school" English.

In her summary of findings, Römer (2005, 268) states that many recent corpus linguistics researchers have conducted studies related to some central lexical-grammatical features in natural language corpora and compared it with language teaching materials. It seems that all of these studies have come to similar conclusions: there are significant differences between real native-speaker English presented in the corpora and “school” English that appears in EFL and ESL coursebooks and grammars. Equivalent results were also found by Mauranen (2004, 90) who concludes that “what is taught as functional language use is not necessarily in agreement with what’s frequent in the language, or even appears at all”. The results of Römer’s comparative analysis on the progressives in the two spoken corpora and the textbook corpus conform to this general conclusion on the mismatches between “real “ English and “school” English. Römer (2005, 271) states that by comparing the findings in the data drawn from the corpora “we are able to show that progressive forms show largely different average and verb-specific distributions of a number of function and context features in “school” English and real spoken English”. Therefore, progressive patterns are presented differently in both types of English and are indeed used inconsistently inside and outside the EFL classroom. Some examples of the most significant pattern deviations in the textbooks include:

- general over-representation of progressives with present time orientation
- under-representation of progressives with past time orientation
- overuse of future time reference progressives
- under-representation of negated progressives
- overuse of a couple of verb-preposition collocations
- missing of some typical verb-object patterns, for example *asking+you*, *bothering+me*, *liking+it*, *listening+to+the+noun (group)*, *telling+me*

These results lead to the conclusion that EFL German coursebooks seem rather unrealistic when it comes to a lexical-grammatical feature like the progressive. Römer (2000, 277) states that “real-life language with its typical usage patterns has not quite entered the classroom” and that “pedagogical descriptions are rather inadequate and the classroom does not meet reality” (Römer, 2005, 273). For this reason, Römer additionally proposes some further pedagogical implications as to how these inconsistencies should or could be taken into consideration in the classroom. Nevertheless, these pedagogical suggestions will not be discussed further in the present study, since the main interest is on the differences in the presentation of a certain grammatical feature in the selected corpora and textbook series. I will return to Römer’s study later on in chapter 6 where the results of the present study are analysed in the light of Römer’s conclusions.

As Römer’s research serves as an example of a comparative study related to linguistic corpora and textbook analysis, it is also necessary, considering the interests of the present study, to discuss some additional studies concerned in particular with the conditional form.

An interesting article by David Maule (1998), called “Sorry, but if he comes, I go: Teaching conditionals”, introduces a pedagogical point of view related to the presentation of conditional sentences traditionally displayed to students in the EFL classroom. The study conducted by Maule points out the contradiction between the variety of conditional forms that exist and the classroom procedure as to how few of the different conditional forms are actually taught and dealt with in practice.

The background assumption in Maule’s research was that current textbooks and pedagogic grammars more or less ignore the wide variety of conditional constructions and concentrate rather on the type *if + present simple* → *will + verb* pattern. For this reason, he

claims, students fail to recognize different conditional constructions and are thus incapable of producing distinct patterns in their own language use. Maule (1998, 118) concludes that if students are only familiar with the structure *if + present Simple* → *will + verb*, they will not be able to function appropriately in situations where this structure is not acceptable. He also adds that "the idea that one structure can be used for every situation...seems to be based more on teacher convenience than on student need". Maule suggests that course books and grammars should include the full range of possible conditional constructions regardless of their assumed frequency of use rather than simplifying the language and "storing up trouble for the future" (Maule 1998, 118), as he puts it. Fulcher (1991, 168) states that "The traditional conditionals are both common and difficult enough to be taught as presented in current course books, while other forms may be overlooked through such an approach". She agrees with Maule's conclusion, suggesting that students should certainly be faced with less frequent forms, preferably in natural contexts, to make them become aware of the range of possibilities in their use.

To find out how the conditional forms are realized in actual language use, Maule collected a sample of 100 conditionals from television. According to his hypothesis, the pattern *if + present simple* → *present simple* is particularly problematic to EFL students, presumably because it is continuously ignored in the classroom. The results of his small-scale study indicated that this conditional construction in fact outnumbered the occurrences of the type *if + present simple* → *will + verb* although it is regarded less central by the authors of coursebooks and grammars. Maule is not alone with his judgement; Kennedy (1998, 166) also concludes that "Regrettably for students of English in many parts of the world, pedagogy sometimes tends to focus on the less frequently used structures and functions. Indeed, in many grammars and course books the most common type of conditional (simple present + simple present) is ignored entirely".

Maule (1998, 118) argues that if students are only exposed to certain conditional types and are taught to believe that these are the only correct forms to express the conditional, they might come to believe that other constructions are somehow colloquial or cases of sub-standard usage. Still, in the sample Maule collected for his study, some 90 per cent of all real non-past conditionals represented other than the type *if + present simple → will + verb*. If this sample can be considered as an illustrative example showing the common qualities of conditional constructions in general, there is certainly a contradiction between what is taught in the classroom and what is realized in actual language use.

The results of Maule's small scale study indicate that only 7 of the total 100 conditional sentences conformed to the type *if + present simple → will + verb*. Of the total 100 sentences, 61 instances represented real non-past conditionals, 19 represented the unreal non-past form *if + past simple → would + verb* and 6 represented the unreal past form *if + past perfect → would have + verb*. Real past and real mixed time (non-past + past) constructions constituted 9 instances of all conditional constructions.

The outcome of Maule's research is that although the conditional type indicating real non-past conditionality makes up the majority of all conditional instances, a crucial proportion of the total number of the conditional sentences also contains other constructions that should not be neglected when dealing with the conditional in the classroom. The relevance of Maule's study will be referred to in more detail later when discussing the results of the present study.

Kennedy (1998, 165) introduces a study conducted by Wang (1991) in which the distribution of the verb-form combinations is analysed in a sample of 2,017 *if*-conditions in the Brown Corpus and 2,307 *if*-conditions in the LOB Corpus. Wang discovered 76 different combinations in Brown and 103 in LOB. The 24 most frequent verb-form combinations account for about 75 per cent of all the *if*-conditionals included in the sample. In each corpus,

approximately 75 per cent of the *if*-conditionals represent open or real conditions and the remaining 25 per cent include hypothetical or unreal conditions.

The results of Wang's study (1991, my source here Kennedy 1998, 165) are introduced in the table below. It presents the verb-form combinations in conditional sentences in the Brown and LOB corpora. These 10 categories representing different verb-form combinations are formed of the 24 most frequent verb-form combinations, and they are listed from 1 to 10 according to their frequency in Wang's study, starting from the most frequent verb-form combination.

Table 1: Verb-form combinations in conditional sentences in the *Brown* and *LOB* corpora (adapted by Kennedy from Wang 1991,58)

% of *if*-tokens

Verb form in <i>if</i>-clause	Verb form in main clause	Brown	LOB
Simple present	Simple present	22.0	22.0
Simple present	<i>will/shall/be going to</i> + stem	13.2	12.5
Simple past	<i>would/could/might</i> + stem	11.3	11.1
Simple past	Simple past	6.7	6.8
Simple present	<i>should/must/can/may/ought to</i> + stem	10.0	6.4
Past perfect	<i>would/could/might have</i> + past participle	3.9	4.1
<i>were (to)</i>	<i>would/could/might</i> + stem	4.0	4.0
<i>can</i> + stem	Simple present	1.1	3.2
Simple present	<i>would/could/might</i> + stem	1.9	2.4
Simple present	Imperative	1.7	2.0
Total		75.8	74.5

According to table 1 there is not very much difference between American (Brown) and British English (LOB) when it comes to the use of the conditional. The construction *simple present* + *simple present* is the most frequently used *if*-conditional type in each corpus. Kennedy (1998, 165) also introduces an older study carried out by Hill (1960) in which it was noted that at least 324 different combinations of finite verb forms could occur in the main and subordinate clauses of conditional sentences where the subordinate clause was introduced by *if*. Wang discovered that in his own study, only 76 of the 324 possible combinations introduced by Hill

occurred in the Brown corpus and 103 in the LOB corpus (Kennedy 1998, 165). Kennedy does not present the distribution of all the verb-form combinations of Wang's study in his book but it becomes clear that about 25 per cent of the materials represented in both corpora are further divided into 66 different types in the case of Brown and into 93 different types in LOB. The distribution of the verb-form combinations is, therefore, very uneven. This leads to the conclusion that there is a wide range of variation in the use of the *if*-conditional in authentic language use.

2. Formal aspects of *if*-conditional constructions

In order to be able to identify and classify the conditional constructions that appear in the materials, let us begin by considering the formal aspects of conditionals. In the following I will introduce the most common ways of classifying conditional constructions and also give some insights into a more detailed approach as to how conditionals can be considered from a linguistic point of view.

According to Palmer (1988, 149), a conditional sentence is basically made up of two parts: the *if*-clause or protasis and the main clause or apodosis. The function of the conditional sentence is to indicate the dependency of one proposition upon another. Palmer (1988, 149) also points out that every conditional sentence without ellipsis is of the semantic form 'If *p*, then *q*'. Conditional sentences come in many forms and the difficulty that lies behind the complexity of using the conditional results precisely from this variety of forms that the language user needs to choose from. There are three main types of conditional sentences that are usually distinguished when discussing the basic structure of conditionals.

The following three types of *if*-conditionals are traditionally described in EFL textbooks and grammars and are, therefore, mentioned in this context. The examples of these three types are taken from Palmer (1988, 149).

Type 1: *if* + simple present (*will*-future)

[1] If John comes, Mary will leave.

In this type of conditional sentence, it is possible and also very likely that the condition presented in the protasis will be fulfilled. In this case, Mary's leaving is dependent upon John's coming.

Type 2: *if* + simple past (conditional I)

[2] If John came, Mary would leave.

The conditional can be either real (for example, "If John comes, Mary will leave") or unreal, as exemplified in the above sentence. The use of the past tense refers to the unreality of the proposition. As Palmer puts it "In real conditionals the speaker merely presents the propositions that are linked, without indicating any views about them. In unreal conditionals he indicates that he has doubts about the propositions, or, in particular about the proposition expressed in the protasis" (Palmer 1988, 150). In this case, it is possible but very unlikely that the condition will be fulfilled.

Type 3: *if* + past perfect (conditional II)

[3] If John had come, Mary would have left.

This type of conditional relates to the past and therefore it is impossible that the condition will be fulfilled. Palmer points out that these conditionals are often counterfactual since they relate two propositions that are contrary to the factual situation.

The categorization of conditional constructions may differ somewhat from one source to another. For example, some authors (such as Ferguson, 2001) prefer referring to the *will*-future as the first conditional, to *if* + simple past as the second conditional and as *if* + past perfect as the third conditional, respectively.

Palmer (1998,149) introduced the three forms of *if*-conditionals that are traditionally included in the EFL textbooks. As can be concluded from table 1, these three types of the

conditional do not, in fact, constitute the three most frequent if-conditional constructions in real-life language usage. Nevertheless, these results seem to be consistent with Maule's results. Maule (1998, 122) states that the traditional division of the three most common types of conditionals places too much emphasis on the idea of standard structures and "places unnecessary restraints on the student's natural curiosity and communicative ability". He continues by claiming that presenting conditionals in this form may even lead to rejecting perfectly valid alternative forms.

As stated by Ferguson (2001, 64), the traditional paradigm of the three most common types of conditionals is not sufficient enough to cover many, perhaps most, conditionals. This is because most conditional constructions do not conform to these three classifications as they include many combinations of different tenses between protasis and apodosis, as for example *If he has sent you an invitation, I won't go*. For this reason, Ferguson (2001, 64) concludes, "the traditional EFL account is not only not explanatory-being a list of main patterns- but does not accurately characterise the *range* of forms observed in naturally occurring conditional sentences". He introduces an alternative characterisation of conditionals, which uses semantic terminology to differentiate between open and hypothetical conditionals or real and unreal conditions.

3. Semantic aspects of *if*-conditional constructions

As stated above, conditional sentences can be described not only in terms of formal aspects but also referring to their semantic features. This type of analysis draws on the hypotheticality of the conditional, that is to say, considers whether the conditional has a high or low probability of realisation. Therefore, conditionals can be divided into open and hypothetical conditions. Leech and Svartvik (1994, 107) describe conditional clauses as “related to reason clauses, but they discuss the consequence of something which may or not be a real event”. To illustrate this point further, Leech and Svartvik give the following examples:

- [4] I'll lend Peter the money, *because he needs it*.
 [5] I'll lend Peter the money *if he needs it*.

In the first example the speaker of the sentence knows for a fact that Peter needs the money whereas in the second example the speaker does not know whether he does or not. The second sentence can also be referred to as an open condition. According to Leech and Svartvik (1994, 107) this term is used because “the truth or falsehood of what the sentence describes is ‘open’, i.e. unknown”. In other words, “Open conditions are neutral: they leave unresolved the question of the fulfilment or nonfulfilment of the condition” (Greenbaum and Quirk 1990, 317). Ferguson (2001, 64) adds that this type of conditional differs from a conditional representing a higher degree of hypotheticality in that it is expressed using the indicative without tense backshifting.

The other type of conditional denotes unreal or hypothetical condition, for example

- [6] I would lend Peter the money *if he needed it*.

In this case, the assumption behind the speaker's statement is that Peter does not need the money and therefore the hypothetical past tense is used. This type of conditional can also be referred to as counterfactual, as the realisation of the condition is very unlikely or even impossible. Greenbaum and Quirk (1990, 317) explain that "A hypothetical condition...conveys the speaker's belief that the condition will not be fulfilled (for future conditions), is not fulfilled (for present conditions), or was not fulfilled (for past conditions) and hence the probable or certain falsity of the proposition expressed by the matrix clause". In this case, 'matrix clause' refers to protasis, i.e. the *if*- clause. According to Ferguson (2001, 64), unlike open conditionals, hypothetical conditions do include tense backshifting. This can also be noted from the example above, which has future time reference with past tense and conditional in the protasis.

According to Kennedy (1998, 165), open conditions can be divided further into the categories of factual and predictive conditions. Factual conditions denote real facts, as in

[7] I get a headache if I eat too many nuts.

Predictive conditions, on the other hand, are related to the future which is indicated by the simple future form:

[8] If you're not careful, you'll fall down.

Hypothetical conditions can also be divided into subcategories called improbable and counterfactual conditions. As the name suggests, improbable conditions are related to sentences in which the condition is not likely to happen or be true, for example

[9] I would take a coat if I thought it would rain.

Counterfactual conditions indicate that the statement expressed in the protasis is not true, for example

[10] If I were you, I'd stay here.

Wang studied the distribution of *if*-conditional sentences across these semantic categories and in Kennedy's book (1998, 166) it is stated that open conditions are approximately three times as frequent as hypothetical conditions in the samples taken from Brown and LOB. The results are very similar in both corpora, although Brown describes American English and LOB British English. In both corpora, factual conditions account for almost 50 per cent of all the conditional sentences and predicative conditions for over 25 per cent. Improbable conditions form 14 per cent of the instances and the least frequently used counterfactual condition slightly over 10 per cent.

The materials included in this study are also examined from this semantic point of view to observe the comparative analysis from a different point of view and to gain an alternative way of approaching the data. The analysis of the semantic features of the *if*-constructions represented in the materials is introduced in more detail in section 6.2.

4. Materials

The materials chosen for this study include a series of textbooks and two corpora representing the language use of native and non-native speakers of English. The reason for selecting these three very different types of samples of language is that textbooks can be considered a norm according to which the language is taught in the classroom and then again the corpora give us an insight into how the language actually works in real life. The study approaches the *if*-conditional from three different perspectives: the textbooks, the BNC corpus and the ICLE corpus. Applying several different sources of data to study the same phenomenon can also be referred to as data triangulation (<http://www.tele.sunyit.edu/traingulation.htm>). This type of approach combines different sources of data to give more credibility to the quantitative analysis and to gain different viewpoints to the same subject of study.

Corpus analysis constitutes a useful tool in examining detailed information about the linguistic structure of sentences in large amounts of data (Oostdijk 1991, 57). It offers a means for testing hypotheses based on formal grammars and therefore serves as a testing ground for the analysis of linguistic structures, their frequency and distribution. A corpus provides a tool that serves particularly well for observing repeated sequences and patterning in language, without depending on conventional linguistic categories (Mauranen 2004, 102).

As corpora became more common from the 1960s onwards, teachers and other people in the pedagogical sphere started to apply them in different forms in their work. According to Sinclair (2004, 2), teaching lexical and phraseological structures required a higher priority and therefore corpus work became an issue of interest for language teachers. Nowadays corpora are regarded by many teachers as practical instruments that can be used more and more on a daily basis in the classroom. Easy and effortless access has become possible since standard small computers and user-friendly softwares are available for most

normal tasks, websites are accumulating fast, and corpora have become almost a part of the pedagogical landscape (Sinclair 2004, 2).

The decision to use both written and spoken parts of the BNC corpus instead of concentrating only on one aspect of language production has to do with the question of authenticity. While the textbooks and the written part of the BNC offer a means for examining written language only, corpus study on the spoken part makes it possible to conduct the analysis of an even more spontaneous spoken material. As concluded by Mindt (1986, 129), most foreign language courses are designed to acquire the ability to communicate in everyday-life situations. This aim emphasises the role of spoken language in teaching and therefore corpora of the spoken language with particular regard to every-day conversations are useful for pedagogical purposes (Mindt 1986, 129). Little by little, spoken language has gained ground in language teaching, and nowadays the ability to express oneself orally is probably considered by most even more essential than written skills. Therefore, including the analysis of both written and spoken parts of the BNC corpus can be considered justified. Examining the material produced by the specific audience that the textbook series is aimed at provides practical evidence on the way the students themselves use the *if*-conditional. *The International Corpus of Learner English* was selected for this purpose.

4.1. *Culture Café*- textbook series

The textbook series selected for the study is *Culture Café*, a series of course books used widely in Finnish upper secondary schools. The books were published between the years 2002 and 2005, and therefore represent a relatively recent part of the variety of textbooks used in Finnish upper secondary schools in general. The entire *Culture Café*-series includes eight

textbooks, which are all included in the analysis. The first six books represent the compulsory courses and the last two optional courses. For mainly practical reasons, the textbook analysis concentrates only on the actual text chapters without taking into consideration any exercises, extracts, songs, essay titles, instructions or grammar sections.

The reason for ending up analysing upper secondary school material in particular was that it can be assumed to include relatively challenging constructions and a variety of grammatical forms that offer more data than materials aimed at lower levels. In addition, what makes upper secondary school materials an interesting object of study is that the assumed target audience represents a relatively wide scale of language users when it comes to the level of their knowledge of English. There are surely many students who acquire a very fluent level of English during upper secondary school and, on the other hand, there are some whose knowledge does not correspond to the minimum requirements set for the matriculation examination. Also the learning objectives and needs vary from one learner to the next. This inevitably sets more challenges for the producers of the materials.

Selecting *Culture Café* as the subject of the textbook analysis rather than some other textbook series in the market resulted partly from the fact that it was readily available in the public library. In addition, the series in question has been marketed as authentic and the texts have been claimed to be thought-provoking as well as versatile. Therefore, it was assumed that this series serves well for the purposes of the present study.

4.2. *The British National Corpus*

The British National Corpus (BNC) is a collection of samples of written and spoken language from a wide variety of sources. It contains 100 million words that represent British English

from the latter part of the 20th century. The corpus was compiled between the years 1991 and 1994. The contents of the corpus is described on the official web pages of *the British National Corpus* as follows:

The **written part** of the BNC (90%) includes, for example, extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays, among many other kinds of text. The **spoken part** (10%) includes a large amount of unscripted informal conversation, recorded by volunteers selected from different age, region and social classes in a demographically balanced way, together with spoken language collected in all kinds of different contexts, ranging from formal business or government meetings to radio shows and phone-ins.
[\[http://www.natcorp.ox.ac.uk/corpus/index.xml\]](http://www.natcorp.ox.ac.uk/corpus/index.xml) Accessed on 14 April 2007

Language corpora provide language workers with evidence of how language is really used, evidence that can then be used to inform and substantiate individual theories about what words might or should mean. Traditional grammars and dictionaries tell us what a word *ought to mean*, but only experience can tell us what a word *is used to mean*

[\[http://www.natcorp.ox.ac.uk/using/index.xml\]](http://www.natcorp.ox.ac.uk/using/index.xml). With the help of a corpus it is possible to work on substantially larger amounts of data than by manual work. More and more applications become available for the reach of researchers, teachers, linguists and scholars as the variety and scope of different kinds of corpora are developed. Nowadays these electronic collections of texts have established their place in the very heart of linguistic research as the starting point for many language-related studies.

Choosing the BNC corpus does not imply that it is assumed that British English is somehow more frequent in teaching materials or more present in the EFL classroom than American English, for instance. Nor is it argued that this should be the case. Utilising a British native speaker English corpus for reference was a practical choice since it is available on-line. Furthermore, as stated above, it enables us to compare written and spoken data and is therefore appropriate for this study.

4.3. *The International Corpus of Learner English*

The International Corpus of Learner English (ICLE) was created by the Louvain Centre for English Corpus Linguistics. This computerised databank contains over 2 million words of essay writing by advanced learners of English as a foreign language from 11 different mothertongue backgrounds. The essays are written by 3rd and 4th year university students of English (Granger et al. 2002, 27).

According to Granger et al. (2002, 27), the ICLE comprises 3,640 essays and a total number of 2,500,353 words distributed in 11 national subcorpora. The Finnish subcorpus contains 390 essays and 278,103 words. It is further divided into three parts according to the mothertongue(s) of the writer: Finnish, Swedish from Finland and bilingual Finnish-Swedish essays. The material collected for the Finnish subcorpus was collected at the Universities of Helsinki, Joensuu, Jyväskylä, Oulu and Turku.

According to Granger et al. (2002, 1), the purpose of the ICLE is to allow the researchers to compile their own tailor-made corpora on the basis of their specific criteria. The aim is that the corpus can “contribute to giving theories of second language acquisition a more solid empirical foundation and lead to the production of more learner-aware pedagogic material” (Granger et al. 2002, 1). Thus, the ICLE was compiled to provide researchers with a database of authentic material to support their studies and, ultimately, to improve the level of EFL material introduced to students. This underlying intention is mentioned frequently in the ICLE handbook where Granger et al. (2002, 44) states that “the corpus should help improve the ‘learner fit’ and hence efficiency of ELT pedagogical materials aimed at higher or intermediate or advanced learners of English”.

The ICLE is usually used to carry out descriptive studies to provide more accurate information on learner lexis, (lexico-)grammar and discourse features by referring to a limited

data sample (Granger et al. 2002, 43). Most studies including the ICLE are corpus-based and the corpus is used to provide material to exemplify a pre-existing theory (Granger et al. 2002,4). Still, in this study, the materials are not interpreted on the basis of any theory but, instead, I will take a corpus-driven method to approach the data. This approach will be discussed in more detail in chapter 5.

According to Nadja Nesselhauf (2004, 125), native speaker corpora are convenient for collecting data on how native speakers usually use language. Still, as Nesselhauf points out, when considering language teaching it is more essential to find out the typical difficulties that the learners of a certain language face when studying the target language and for this particular purpose it is more beneficial to use the data offered by learner corpora alongside with the native speaker corpora. This way it is possible to identify and analyse mistakes, correct forms and over- and underuse of certain linguistic features, as opposed to native language use. Nesselhauf also concludes that in addition to being able to distinguish the difficulties that EFL students have when studying the target language, learner corpora can also be used to improve pedagogic material by making it correspond more closely to typical native speaker language use.

The view of utilising learner corpora merely as representations of learner errors is challenged by Melinda Tan (2005) who wrote an article on learner corpora and language authenticity. In the article she argues that learner corpora investigations rarely consider the link between language and culture and the growing role of English as lingua franca. This leads to the common practice that `peculiar` patterns of use in learner language are classified as errors or inauthentic language use when they could, in fact, be regarded as “reflections of their users’ social and cultural identities” (Tan 2005, 127). Tan claims that the conclusions commonly driven from learner corpora data imply that learner language is erroneous because it contains unnatural and inauthentic usages when compared to native speakers’ language use.

She goes even further by concluding that “if we were to examine the ‘authenticity’ criteria by which researchers base their claims, we would find that it is very much based on imperialistic assumptions about the ownership of English, rather than the present role of English as lingua franca” (Tan 2005, 128). She argues that the impression of native speaker language as the only representative of ‘real’ or ‘authentic’ English is conveyed mainly by publishers of EFL coursebooks.

One could argue that both of these opinions on the role of learner corpora can be justified. The line between authentic and inauthentic material is not clear-cut and therefore generalizations cannot easily be avoided when considering the level of authenticity of any kind of corpus data. Labelling different varieties of English as either authentic or inauthentic is also problematic since both native and non-native varieties represent the language as it is used by a certain group of people and therefore it can be regarded as authentic to that particular group of language users. There are by far more non-native speakers of English than there are native speakers and thus it could be argued that the normative rules of the language should be based on the principles governing the variety used by the majority of English speakers.

Nevertheless, the present study will take a ‘purist’ view (as Tan calls it) on the matter of correctness of standard English and the incorrectness of any other varieties of English and considers the language use of native speakers of English as an example of authentic language. I chose to approach the use of *if*-conditionals from this point of view because the textbooks, the BNC and the ICLE all represent a different perspective on the same subject. Studying the ICLE sub-corpus that includes texts written by Finnish learners of English enables us to examine the way language forms are realised in practice after studying a foreign language for a considerable period of time. Access to the BNC and the ICLE simultaneously

provides a comprehensive amount of material for a comparative study between the language use of native and non-native speakers of English.

5. Methods

As the purpose of the present study is to analyse the use of conditionals through a comparative research referring to a series of textbooks and two corpora, the next logical step after defining the characteristics of this grammatical phenomenon is to examine the materials by focusing on the conditional constructions that appear in them.

The actual research process starts by identifying the different conditional structures that are included in the textbooks. Then a quantitative analysis is conducted by dividing the conditionals into groups according to the type they represent. A similar method is used to analyse the randomly selected section of the BNC and the ICLE. For the BNC, a limited portion of the material is randomly selected for reference by using the thinning option provided by the program itself. For the ICLE, the sample is reduced by restricting the search to the essays written only by monolingual Finnish speakers and further by selecting only the first *if*-clause from each text. The occurrence of the conditional forms in each source is then compared to each other according to the frequency and classification of the forms.

5.1. Corpus based vs. corpus-driven approaches

In her book *Progressives, Patterns, Pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics*, Römer (2005, 8) introduces the concepts of corpus based and corpus-driven approaches in the field of linguistic study. These two disciplines differ in terms of applying a distinct starting point for the analysis of the data discovered through corpus analysis. Corpus based approach is used to test pre-formulated ideas and hypotheses as well as to gain illustrative examples to support an existing theory. For

corpus-driven linguists, on the other hand, it is important “not to let other researchers’ s ideas and existing theories guide or influence them too much in approaching the data” (Römer 2005, 10). Where corpus based linguistics uses pre-tagged texts and annotated corpora, corpus-driven linguistics relies on direct processing of the text and findings derived from the data without being influenced by the interference of any outside factors, such as other researchers’ views of the language. Römer herself seems to favour corpus-driven linguistics over corpus based linguistics and also lists some advantages attained through corpus-driven linguistics. One of the most significant assets provided by corpus-driven analysis is the ability to lead to new insights into language. It also enables us to see how language really works, how it is used in communicative situations, and, consequently, find out how it ought to be taught (Römer 2005, 11).

Since Römer is of the opinion that in order to apply the corpus-driven approach the researcher needs to examine the data before forming any specific theoretical models, in this context one needs to consider the concept of ‘theory’. One could argue that no data can be approached without some pre-existing theories. ‘Theory’ in this sense is a complex concept which derives its meaning from the researcher’s own experience. Römer seems to argue that, at least in her study, a theory is considered from a relatively strict perspective meaning regularities and relations between certain characteristics whereas it could also be regarded as a pre-existing conception or opinion of the subject under analysis.

In this study, the underlying assumption is that all the conclusions will be based on the analysis of the data itself and no hypotheses will be presented in advance as such. Additionally, the data collected will be analysed without using any annotation to justify any certain theoretical background presuppositions. Therefore, considering Römer’s argument, the present study can be said to represent the corpus-driven approach. The results will be analysed in the light of the previous studies related to this particular subject-matter but any

straightforward influence of these studies on the process of collecting and examining the data will be avoided.

5.2. The quantitative paradigm

In applied linguistics, as in many other fields of study, a distinction is commonly made between quantitative and qualitative methodologies when considering the ways in which the data is collected and analysed. A useful distinction could be drawn between the concepts of 'approach' and 'method'. According to Bell (1981, 75), an approach can be defined as "an orientation to the problem of language learning which derives from an amalgam of linguistic and psychological insights into the nature of language and the nature of the learning process". A method, on the other hand, can be considered as "the application of the insights which constitute the approach to the problem of language learning". Thus, methods could be regarded as the real-life practices that have their origins in the underlying theoretical approaches.

In order to justify the selection of the methodology used in this study, let us begin by mentioning some of the most essential differences between the quantitative and qualitative paradigms. Quantitative studies usually present argumentation based on statistical methods. According to Denzin and Lincoln (1994, 4), qualitative researchers "stress the socially constructed nature of reality, the intimate relationship between the researcher and what is studied" and "seek answers to questions that stress how social experience is created and given meaning". Contrastingly, quantitative researchers emphasize the measurement and analysis of causal relationships between variables, not processes (Denzin and Lincoln 1994, 4). Therefore, quantitative research is mainly interested in the outcome of the study and not so

much in the process through which the results are achieved. In addition, a quantitative study seeks for generalisable results whereas a qualitative research mostly concentrates on ungeneralisable single case studies. Quantitative studies usually rely on large collections of material, representative samples and statistical methods whereas qualitative studies analyse smaller samples of data by examining the phenomenon closely to observe different possibilities to understand and interpret the materials (Viestintätieteellinen tutkimus, 2007). According to Brannen (1992, 4), the most important difference between these two paradigms is the way each tradition treats data. She states that “the quantitative researcher looks through a narrow lens at a specified set of variables”. Contrastingly, the qualitative researcher can be said to look through a wide lens in search for patterns of inter-relationships between a previously unspecified set of concepts (Brannen 1992, 4).

One additional difference between the quantitative and qualitative paradigms is related to data collection. In the quantitative tradition, the research instrument is “a pre-determined and finely-tuned technological tool” which is less concerned with flexibility, imaginative input and reflexivity (Brannen, 1992, 5). An example of this kind of research tool is a questionnaire which requires the respondents to choose unambiguous answers to certain questions. Qualitative research, on the other hand, relies on less clear-cut techniques such as in-depth interviews to achieve complex and discursive replies (Brannen 1992, 5). To summarize, these two different kinds of paradigms represent distinct ways of approaching the data, data collection and the analysis process. They observe and capture different realities or different aspects of reality and are, therefore, usually not regarded as substitutable for one another since they are associated with very different theoretical perspectives and different conceptualizations of research problems (Brannen, 1992, 5).

The quantitative methodology was selected to carry out the present study because the underlying aim was precisely to obtain generalisable results from relatively large samples of

data representing different sources. Therefore, individual case studies could not have been carried out to gain comprehensive demonstration of the representation of *if*-conditionals in textbooks and authentic language. Larsen-Freeman and Long (1991, 44-45) state that both qualitative and quantitative methods improve our understanding of SLA. Therefore, they conclude that we should not choose between these paradigms or between extant methodologies but rather design a research methodology which possesses the best characteristics of both approaches to examine the research question under analysis. Even if some researchers do choose to combine some of the characteristics from both types of methodologies to approach their data, this study concentrates only on the quantitative paradigm because it serves the purposes of the study most appropriately.

5.3. Collection of materials

The collection of material started from going through the textbooks from cover to cover. I analysed all the text chapters and listed all the *if*-clauses presented in the texts. To gather data from the BNC, a simple standard query was conducted by using “if” as a search word. The query was first restricted to the written subsection and then separately to the spoken subsection of the BNC. This query returned about 200 000 matches in the written part and about 48 000 matches in the spoken part. To limit the amount of data to a manageable proportion the number of hits was reduced by using the thinning option. This resulted in a random section of the sample to be displayed in an alphabetical order according to the name of each text file. I decided that 250 hits were suitable for each corpus because, on the one hand, the number needed to be small enough for practical reasons and, on the other hand, the samples needed to be representative enough to be able to make generalisable conclusions.

According to Granger et al. (2002, 47), the ICLE allows the user to create a corpus of learner essays based on a selection of 'learner variables' (such as age, gender, mother tongue etc.) and 'task variables' (such as topic, genre, length etc.). Still, the program itself does not allow any random queries on specific search words and therefore I had to select the preferred sample manually. To limit the amount of data I decided to take into account only the texts that were written by monolingual Finnish speakers because presumably the majority of the students that the textbooks are aimed at represent that group of language users. I browsed the files and, to be able to work on the texts, I copied them to Word. At this point the number of texts was 229 and the number of words 173,217. These texts included a total of 676 instances of *if* which was far more than I was planning to work on. Therefore, I decided to take into account only the first *if*-construction from each text. As not all the texts included *if*-constructions, the final number of hits was reduced to 197. This number seemed adequate and since it represents a wide variety of different language users I decided it was enough for the purposes of this study.

5.4. Methods of analysis

The analysis of the data collected from the corpora and the textbooks was relatively straightforward. The analysis of the data began by dividing each *if*-conditional sentence into a category according to the verb-form combination that comprises the sentence. The results are presented in table 2 in a descending order with the most frequent verb-form combinations in the beginning of the table. As for the material collected from the BNC, a further analysis was conducted to find out the distribution of the different textual genres in order to examine whether the text samples exemplify the variety of all textual categories available in the BNC

in a representative way. Then a semantic analysis was conducted by identifying the semantic category represented by each *if*-sentence.

6. Findings

The analyses of the most important findings of this study are presented separately for the formal and semantic aspects of *if*-conditional constructions. This division is made in order to make the analysis more explicit, since these two aspects of the conditional represent relatively different ways of approaching the data, the first relating strictly to the grammatical features of the conditional and the second dealing with the meaning behind the formal constructions.

6.1. Analysis of formal aspects

Since not all the *if*-clauses necessarily contain conditional, the first step in the analysis was to separate the clauses that do not imply conditionality. The textbooks represent a total of 72 *if*-clauses, 68 of which are conditionals.

In the case of the BNC corpus analysis, it had to be taken into consideration that there are also some sentences that could not be included in the study because they are unclear. These sentences are mostly partial with some parts completely missing or so obscure that they cannot be included in the analysis. Unclear in this context means that a sentence could not be classified because it was incomplete or completely unintelligible. The corpus data also included some sentences that could not be classified because some parts of the sentences had remained unclear when the corpus itself was compiled. The following examples of unclear sentences are taken from the spoken corpus:

[11] I'll never talk to you again if you <unclear dur=2> <unclear dur=2> (KC9 6865)

[12] So sixteen to the three quarters is find the fourth root and cube it or if you like cube sixteen and then find the fourth root. (K6J 459)

[13] <laugh>, I'm very <unclear> I like, if they haven't got a scratching board, scratch what's in the <unclear> (KB7 13279)

The written section includes 250 *if*-sentences, 13 of which are classified as unclear and 58 of which are not conditional. Thus, the number of hits in the written part is reduced to 179. The spoken part of the random query also contains 250 *if*-sentences. 42 of these sentences were unclear and 41 *if*-sentences were not conditional. As might be expected, the reason why there are so many sentences that remain unclear in the case of the spoken texts is that naturally occurring speech takes place in authentic situations where hesitation, interruptions, overlaps and repetition are common and thus reduce the number of full, comprehensible sentences. The number of *if*-conditional sentences in the spoken part of the data is therefore 167.

The sample taken from the ICLE is comprised of 194 *if*-sentences, 173 of which denote conditional. There were 35 texts that did not include any *if*-constructions.

The following table presents the number of *if*-sentences in each source. The results are divided into different groups according to the verb-form combinations they represent in the *if*-clause and in the main clause.

Table 2: Verb-form combinations in *if*-sentences in the *Culture Café*-textbook series, the written and spoken subsections of the BNC corpus and the ICLE corpus

Verb form in <i>if</i> -clause	Verb form in main clause	Textbooks	Written	Spoken	ICLE
Simple present	Simple present	10	38	43	55
Simple present	<i>will/shall/be going to</i> + stem	20	26	31	19
Simple past	<i>would/could/might</i> + stem	9	30	17	26
Simple present	<i>should/must/can/may/ought to</i> + stem	4	19	7	24
Past perfect	<i>would/could/might have</i> + past participle	5	13	2	3
Simple present	Imperative	0	9	9	0
Simple past	Simple past	5	6	4	4

Simple present	<i>would/could/might</i> +stem	3	4	6	7
Simple present	-	0	0	8	0
Simple past	Simple present	0	0	7	1
Present perfect	Simple present	1	2	2	3
<i>can</i>	<i>will</i>	2	2	0	0
<i>is (to) + stem</i>	Simple present	0	4	0	0
Present continuous	Simple present	1	1	2	1
<i>can + stem</i>	Simple present	1	1	1	3
Present perfect	<i>will</i>	0	3	0	0
<i>will/be going to</i>	Simple present	0	2	1	1
Simple past	-	0	0	3	0
Present perfect	<i>can</i>	1	1	0	0
<i>would</i>	Simple present	0	0	2	1
Simple past	<i>would have + past</i> participle	1	0	1	0
Simple present	Present perfect	1	0	1	0
Simple present	<i>have to</i>	0	2	0	0
Past continuous	<i>would</i>	0	0	2	0
<i>could</i>	<i>would</i>	0	1	1	1
Present perfect	Present perfect	0	1	1	0
Simple present	Present continuous	0	0	2	0
<i>were (to)</i>	<i>would/could/might +</i> stem	1	0	0	1
Present continuous	<i>will</i>	1	0	0	0
Simple present	<i>is likely to + Present</i> continuous	1	0	0	0
<i>is to + stem</i>	<i>has to + stem</i>	1	0	0	0
<i>must</i>	Imperative	0	1	0	0
<i>could</i>	Imperative	0	1	0	0
<i>would</i>	<i>would</i>	0	1	0	3
Simple present	Simple past	0	1	0	1
Past participle	Simple past	0	1	0	0
Past participle	<i>would</i>	0	1	0	0
<i>is to</i>	<i>will</i>	0	1	0	0
Past participle	<i>should</i>	0	1	0	0
<i>will/be going to</i>	<i>might</i>	0	1	0	0
<i>was to</i>	<i>might</i>	0	1	0	0
Past perfect	Imperative	0	1	0	0
Simple present	<i>is likely to be</i>	0	1	0	0
Present perfect	<i>is likely to be</i>	0	1	0	0
<i>would</i>	Simple past	0	1	0	0
<i>could</i>	Simple past	0	0	1	0
Simple past	Past continuous	0	0	1	0

Simple past	Imperative	0	0	1	0
<i>may</i>	Simple present	0	0	1	0
Present perfect	<i>could</i>	0	0	1	0
<i>was to say</i>	Simple present	0	0	1	0
Past perfect	<i>would</i>	0	0	1	1
Simple past	<i>will</i>	0	0	1	0
Simple past	<i>should</i>	0	0	1	0
<i>could</i>	-	0	0	1	1
Present continuous	-	0	0	1	0
Simple past	<i>were likely to have been</i>	0	0	1	0
Present perfect	-	0	0	1	0
Past perfect	Simple present	0	0	1	0
<i>would</i>	<i>could</i>	0	0	0	1
Past participle	Present perfect	0	0	0	1
<i>would have</i> +Past participle	<i>would have</i> + Past participle	0	0	0	2
<i>should</i>	Simple past	0	0	0	1
Present continuous	Present continuous	0	0	0	1
<i>can</i> + stem	<i>can</i> + stem	0	0	0	1
Present perfect	<i>should</i>	0	0	0	2
<i>be going to</i>	<i>will</i>	0	0	0	1
<i>can</i> + stem	<i>should</i> + stem	0	0	0	1
<i>would have</i> + Past participle	<i>would</i> + stem	0	0	0	1
<i>would have</i> + Past participle	simple present	0	0	0	1
<i>will</i>	<i>would</i>	0	0	0	1
<i>were (to)</i>	Simple present	0	0	0	1
Past participle	Simple present	0	0	0	1
<i>should</i>	Simple present	0	0	0	1
Not conditional		4	58	41	21
Unclear		0	13	42	0
Total		72	250	250	194

The results of this small-scale study on the *if*-conditionals in these specific materials seem to show that in real-life language this construction is, in fact, far more complex than the teaching materials imply. In the textbooks, *if*-conditionals are presented in 18 different verb-form combinations, whereas in the written part of the BNC the corresponding number is 33 and in the spoken part it is 36. In the sample taken from the ICLE, 34 different verb-form

combinations are represented. Even though the number of occurrence of each verb-form combination in this sample may not seem very significant, in many cases being only one, the mere variety of potential verb-form combinations that exist in actual language use is enough to demonstrate that the conditional constructions presented in the teaching materials do not form an accurate picture of the complex nature of the conditional.

When analysing the results presented in table two it should be noted that even if the total number of verb-form combinations seems relatively high, not all the combinations represented in the materials can be considered grammatically correct or well-formed. The verb-form combinations included in the analysis simply represent all the instances where *if*-clauses occurred in the materials under investigation. Therefore, such verb-form combinations as *would have + past participle + would + stem* that would normally be considered grammatically incorrect are taken into consideration since they appear in the materials. It would have required much more time and resources to separate the erroneous sentences from the correct ones and as one of the goals of this study is to analyse authentic language use, it would not serve the interests of the study to manipulate the samples to achieve error-free material, quite the contrary. Since the purpose of this study is not to analyse the level of grammatical correctness of *if*-sentences but the frequencies and classifications represented in the textbooks and the two corpora, I will not make further statements concerning the correctness of the sentences.

As regards the corpus data, the BNC does not enable the same random query to be conducted more than once. For this reason, the data obtained in the query had to be saved in a file format separately from the corpus itself to be able to return to the same random sample for further reference. It is, however, possible to save the particular web page that the data is on but it is not possible to work on the data or perform an additional post-query option after saving the page that displays the actual text files. Because of this limitation I decided to save

the genre classification information separately in a text file to be able to make a more thorough analysis of the data relating to the textual categories that are represented in the samples. Therefore, Lee's genre classification was saved separately for both written and spoken material.

Lee's genre classification serves to analyse the different text types that are represented in the sample. Exploring the genre labels one can retrieve text files that exemplify certain textual categories. The classification scheme created by David Lee contains 46 different genres of written texts and 24 genres of spoken texts. The reason why this type of classification of text types is relevant to this study has to do with the representativeness of the samples. In order for the random samples to be considered representative enough they should contain text extracts from as many different text genres as possible. One practical problem concerning the random queries of the BNC corpus is that it does not allow the researcher to choose, manually or using any given program, any specific text genres to be included in the random sample. Therefore, one cannot make sure that all the existing text categories are represented in the sample.

Although it is not possible to influence the number or type of textual categories included in the random samples, I concluded that the automatic categorisation of the BNC was enough to serve the purposes of this study. Tables 3 and 4 presented in the appendices show the genre classifications of the written and spoken subsections used in the present study.

The written data includes 33 of the total 46 possible genre categories. The spoken data includes 20 of the total 24 possible genre categories. Since there is no other method of collecting the random samples without having to place restrictions on other features of the text files, such as publication date, medium of text, domain, perceived level of difficulty or the sex, age or type of the author, it was considered less relevant that the random samples only represent 72% of the written categories and 83% of the spoken categories. In the end,

most of the genre categories are included and, although the samples are not in this sense perfect, they are representative enough to exemplify the general origins of the data used in this study.

At a general level, it can be concluded that over half of the data (63,2%) from the written part of the corpus is comprised of the six most representative textual categories: fictional prose (novels and short stories), popular lore (popular magazines), miscellaneous texts, academic prose (politics, law and education), commerce and finance economics and non-academic texts (social and behavioural sciences). There are some categories that are not covered at all but as stated earlier, since it is not possible to influence the type of textual categories included in the random sample and the sample nevertheless includes most text types, the absence of some categories is not considered relevant.

As for the spoken part of the random sample, the distribution of the hits in the different genre categories is more uniform than in the written part. The most commonly represented genres include face-to-face spontaneous conversations, business or committee meetings, interviews (job interviews and other types), unclassified texts (miscellaneous spoken genres), broadcast discussions (TV or radio discussions) and non-tertiary classroom discourse. These categories account for about 75% of the total spoken data. Four spoken categories are not represented in the sample.

As stated in the introduction, this comparative study intends to examine whether there is correlation between the occurrence of conditionals in the textbooks and in the corpora. According to the data collected from the textbooks and the two corpora, there is a significant difference between the occurrence of *if*-conditionals presented to students in the teaching materials and in the natural language used in written and spoken usage. As it is not practicable to analyse the total frequency of use (number of hits per million words, for example) based on the textbook analysis, one alternative way of looking into the occurrence of *if*-conditionals

and being able to compare the types of conditionals presented in the materials is to examine the distribution of different kinds of *if*-sentences in the form of percentages. This enables us to consider the relative frequencies of each type regardless of the different number of instances in the textbooks and in the corpus.

The following table illustrates the distribution of *if*-clauses in the textbooks, in the written and spoken sections of the BNC and in the ICLE. The verb-form combinations are put in order in percentage terms according to the number of occurrences in total. Only eleven of the most common verb-form combinations are included in this analysis because they represent the majority of *if*-clauses and are, therefore, the most important constructions. The last three verb-form combinations in the table all occur equally many times in the samples and therefore each combination is included in the analysis of the most frequent verb-form combinations. The number of occurrences is considered horizontally based on table number two and the frequency of each verb-form combination is counted in relation to the total number of *if*-sentences in each source. For this reason the order of the constructions in table 5 does not seem accurate since the percentages of each source are considered together to be able to compare the results to one another.

Table 5: Verb-form combinations in *if*-sentences in *Culture Café*, in the written and spoken parts of the BNC corpus and in the ICLE corpus% Distribution of *if*-sentences

Verb form in <i>if</i> -clause	Verb form in main clause	Textbooks	Written	Spoken	ICLE
Simple present	Simple present	13,9	15,2	17,2	28,3
Simple present	will/shall/be going to + stem	27,8	10,4	12,4	9,8
Simple past	would/could/might + stem	12,5	12	6,8	13,4
Simple present	should/must/can/may /ought to + stem	5,6	7,6	2,8	12,4
Past perfect + past participle	would/could/might have	6,9	5,2	0,8	1,6
Simple past	Simple past	6,9	2,4	1,6	2,1
Simple present	would/could/might +stem	4,2	1,6	2,4	3,6
Simple present	Imperative	0	3,6	3,6	0
Present perfect	Simple present	1,4	0,8	0,8	1,5
Simple past	Simple present	0	0	2,8	0,5
Simple present	-	0	0	3,2	0
Total (%)		79,2	58,8	54,4	73,2

Perhaps the most striking observation that arises from the analysis of this table is that there is significant variation between the proportion of the *if*-sentences included in the eleven most frequent verb-form combinations when one considers the textbooks and the BNC corpus data. Almost 80 per cent of the *if*-constructions presented in the textbooks belong to one of these eleven groups. The results represented in the ICLE suggest a similar outcome with over 70 per cent of the verb-form combinations belonging to the most common constructions. For comparison, only about half of the authentic data represented in the written and spoken parts of the BNC is included in the most common constructions. This, again, serves to illustrate the wide variety of different constructions used in natural language as opposed to the restricted number of constructions represented in the textbooks.

As one could expect, the most commonly used *if*-conditional construction in the textbooks is *simple present + will/shall/be going to + stem* since it forms 27,8% of all the *if*-

conditionals in the books. What is more surprising, though, is the fact that in this specific textbook series the second most common verb-form combination is *simple present + simple present*. According to Maule (1988, 117), this construction is typically omitted from teaching materials and considered less central by the authors of textbooks and is therefore problematic to students. It seems that at least in this case, his theory and the data presented in this study do not correspond to one another. Of course, this conclusion might be due to the small amount of textbook data used in this study and therefore could be studied in more detail to achieve more generally applicable results if more material was included in the analysis. What seems even more unexpected is that the construction *simple present + simple present* is actually the most frequently used *if*-construction in the ICLE. It seems that, contrary to the common assumption, the students are in fact capable of using this particular construction, even if it is presumably not represented frequently enough in the teaching materials. *Simple present + simple present* is the most frequent verb-form combination also in both parts of the BNC.

The construction *simple present + will/shall/be going to + stem* is the second most frequent verb-form combination in both types of BNC corpus data although it is somewhat more common in the spoken part.

There seems to be some variation in the frequencies of other verb-form combinations regarding the written and spoken parts of the BNC. The construction *simple past + would/could/might + stem* is almost twice as common in written data as in the spoken data and this is also the case with *simple present + should/must/can/may/ought to + stem*. *Simple present + simple present* is by far the most recurrently used *if*-conditional type in the spoken sample of the corpus.

The last type included in table 5 represents the structure *simple present + -* which means that the main clause or apodosis has been left out entirely and the only clause forming the conditional construction is the protasis, or *if*-clause. When examining table 2 one can

notice that this type of construction only appears in spoken language in which sentences are typically disconnected in many cases. Some of these elliptical sentences are nevertheless included in the analysis because they include a complete *if*-clause and therefore can be considered to typify the general nature and distribution of the verb-forms used in conditionals in spoken language.

The fact that the vast majority of the *if*-constructions represented in the ICLE belong to the most commonly used constructions seems to indicate that the use of the *if*-conditional is similar to that represented in the textbooks. Still, as mentioned earlier, the total number of different verb-form combinations in *if*-constructions is 18 in the textbooks and 34 in the ICLE. This suggests that the variety of different *if*-constructions is, in fact, very close to the variety represented in the BNC. The results concerning *the International Corpus of Learner English* serve as an interesting example of the differences between the actual language use of the students and the language presented in the textbooks. When comparing the number of verb-form combinations presented in table 5 with the total number of verb-form combinations presented earlier in table 2, it becomes clear that the majority (73,2%) of the *if*-conditionals used by students belong to the most frequently used classes and the remaining instances are divided into as many as 23 different classes. This could indicate that the *if*-conditionals represented frequently in the textbooks are familiar to the students and are therefore used more often than other constructions. There are additionally many structures that are only used by the students and not presented in other sources at all. These verb-form combinations cannot, therefore, be considered to exemplify conventional language use but are derived from the students' own ideas about the patterns in the language. These impressions that affect the way the students form different constructions can have their origins, for example, in the first language that interferes with the acquisition of the foreign language, or in outside factors such as language adopted from alternative sources of data. Such alternative sources include, for

example, television, the internet, radio and newspapers. Thus, almost all of the *if*-conditionals used by the students represent the ones presented recurrently in the teaching materials or, alternatively, exemplify constructions that are formed by the students themselves according to the ideas they have about the rules of the English grammar.

When comparing the results of the textbook analysis and the corpus analysis more closely it is clear that many types of *if*-conditional constructions are generally over-represented in the textbooks. This is the case, for example, with *simple present + will/shall/be going to + stem*, *simple past + simple past*, *simple present + would/could/might + stem* and *can + will*. The construction *simple present + will/shall/be going to + stem* is an illustrative example of overuse with 27,8% of all the *if*-sentences representing this verb-form combination in the textbooks opposed to the just over 10% in the BNC. Only 9,8% of the *if*-sentences presented in the ICLE sample exemplify this particular construction which also serves to indicate that it is over-represented in the textbooks. The only pattern omitted from the textbooks but quite frequently included in the corpus data, is the imperative.

In her study, Römer (2005, 271) mentioned relatively many instances where the use of the progressive forms in German EFL textbooks did not correspond to the data collected from the corpus. She claimed that the representation of the progressive is inconsistent and unrealistic and that in many cases some progressive forms are overused, unused or not used frequently enough in the teaching materials. If one compares Römer's results with the results of the present study from this point of view, it could be argued that the textbook series analysed in this study seems to be more equivalent to the authentic material than in Römer's case. Still, the results show some degree of similarity in that the structures represented in the textbooks give a simplified picture of the linguistic construction under analysis.

As mentioned earlier, the findings of this study differ somewhat from those of the study conducted by Maule (1998, 117) when considering the relative frequency of the

construction *simple present + simple present*. According to Maule's small scale research, this verb-form combination is traditionally under-represented in textbooks although used very frequently in authentic language. The present study found out that the structure in question is, in fact, the second most common structure in the series *Culture Café*. In addition, as Maule also concluded, *simple present + simple present* is clearly the most common *if*-conditional construction utilised in authentic language use and can, thus, be considered to be replaced, to some degree, by *simple present + will/shall/be going to + stem* in the textbooks. This might be true also in the case of the results represented in the present study, where the number of the construction *simple present + will/shall/be going to + stem* is much more significant in the textbooks than in the corpora. In this context it should also be stated, though, that although the textbooks do not represent the full range of possible *if*-conditional constructions, the number of different types of *if*-conditionals is relatively high, at least in the light of Maule's findings.

Wang's study (1991, my source here Kennedy 1998, 165) examined the verb-form combinations in conditional sentences in the Brown and LOB corpora. It was concluded that in these two corpora, about 75 per cent of all the *if*-conditionals represent one of the ten most common structures and that the distribution of the different verb-form combinations is relatively standard when comparing the samples of both corpora. In other words, the frequency of each construction was similar in both Brown and LOB corpora. When comparing the results presented in Wang's study and the results of the present study, one should take into consideration that both Brown and LOB represent written language and should not, therefore, be compared with the results of the spoken subsection of the BNC as such. For this reason, a comparative analysis will only be done based on the findings concerning the written part of the BNC. Table 6 presents the verb-form combinations in conditional sentences in the written part of the BNC corpus. Some combinations that are included in this table are not included in table five which presents the overall view of the

distribution of the verb-form combinations in all sources according to the combinations that are most common when adding the percentages together.

Table 6: Verb-form combinations in conditional sentences in the written part of the BNC corpus

% Distribution of *if*-sentences

Verb form in <i>if</i> -clause	Verb form in main clause	Written
Simple present	Simple present	15,2
Simple past	<i>would/could/might</i> + stem	12
Simple present	<i>will/shall/be going to</i> + stem	10,4
Simple present	<i>should/must/can/may/ought to</i> + stem	7,6
Past perfect	<i>would/could/might have</i> + past participle	5,2
Simple present	Imperative	3,6
Simple past	Simple past	2,4
Simple present	<i>would/could/might</i> +stem	1,6
<i>is (to)</i> + stem	Simple present	1,6
Present perfect	<i>will</i>	1,2
Total (%)		60,8

Examining the results presented in tables 1 and 6 one can observe the differences and similarities in the distribution of the verb-form combinations in conditional sentences in this study and in Wang's study. Analysing table 6 it can be noted that, as in table 1, the verb-form combination used most frequently in *if*-conditional sentences is *simple present + simple present*. In Brown and LOB corpora, the proportion of this structure (22%) is somewhat bigger than in the BNC (15,2%).

In the random sample of the BNC, the distributions of the constructions *simple present + will/shall/be going to + stem* and *simple past + would/could/might + stem* are in the opposite order in relation to one another when compared to Brown and LOB. Still, the difference is not very significant, the proportion of *simple present + will/shall/be going to* being 10,4% in the BNC, 13,2% in Brown and 12,5% in LOB and the proportion of *simple past + would/could/might + stem* being 12% in the BNC, 11,3% in Brown and 11,1% in LOB.

Thus, it could be concluded that the results of the present study correspond to the results of Wang's study when it comes to the three most frequent verb-form combinations in conditional sentences.

For the most part the verb-form combinations represented in the studies are very similar with only some minor variation in the frequencies. The most important differences can be found in the structures *is (to) + stem + simple present*, *present perfect + will*, *were (to) + would/could/might + stem* and *can + stem + simple present*, from which the first two are not included in Wang's table of the ten most frequent *if*-structures and the last two are not among the ten most frequent structures of the sample analysed in this study. In fact, the construction *can + stem + simple present* can be found only once in the sample of the BNC and the construction *were (to) + would/could/might + stem* that comprises 4 percent of the *if*-sentences of both Brown and LOB, does not appear in the sample of the BNC at all.

6.2. Analysis of semantic aspects

The analysis of the semantic features of the *if*-conditionals in the textbooks, in the written and spoken parts of the BNC and in the ICLE reveal that the majority of the constructions represent open conditions, with the factual sense being the most common type of condition. In each source, the predictive sense is the second most common semantic category. As shown in table 7, the results of the semantic analysis of the textbook series resemble to a large extent the results of the corpus study.

Table 7: Semantic categories of *if*-conditional constructions in *Culture Café*, in the written and spoken parts of the BNC corpus and in the ICLE corpus

% of *if*-conditional sentences

Semantic category	Textbooks	Written	Spoken	ICLE
<u>Open conditions</u>				
factual	48,5	57	56,9	68,8
predictive	36,8	20,1	21	13,9
<u>Hypothetical conditions</u>				
improbable	4,4	15,6	15,6	10,4
counterfactual	10,3	7,3	6,6	6,9
Total (%)	100	100	100	100

The distribution of the semantic categories is almost identical in the written and spoken parts of the BNC corpus. Although the results of the textbook analysis indicate that from a semantic point of view the materials seem rather authentic there are some differences as well. The major difference between the semantic categories represented in the textbooks and in the corpora is that the proportion of the improbable conditions is much smaller in the textbooks while the counterfactual and predictive conditions outnumber the proportions represented in the BNC and ICLE corpora.

In the ICLE, the differences in the distribution of the semantic categories is more significant when compared to the textbooks than when compared to the results in the BNC. Particularly the proportion of the predictive conditions is much smaller in the learner corpus than in the textbooks. Almost 69 per cent of all the *if*-conditionals represented in the ICLE are factual conditions, whereas only just over 48 per cent of the conditionals in the textbooks represent this type.

Although the formal structures of the sentences give some basic guidelines as to how the *if*-conditionals are divided into semantic categories, in some cases the only determinative element behind the construction is the meaning. For example, the sentences including the structure *simple present + will/shall/be going to* usually have a predictive sense and the

sentences including *simple present* + *simple present* have a factual sense. Still, especially the sentences that include a hypothetical condition can come in several different forms and in these cases the semantic category cannot be judged from the formal features of the *if*-construction.

Some examples of the *if*-conditional constructions represented in the materials include

[14] if you walk in the desert you get sand in your shoes (factual) (BNC written, CG1 354)

[15] if Bilbo asks a riddle Gollum can't answer, he will lead the Hobbit back to sunlight (predictive) (*Culture Café* book 2, page 80)

[16] Secondly, "If such a species (extraterrestrial) existed, there would be evidence of it on Earth or in the Solar System" (improbable) (BNC written, B74 1761)

[17] This was not the first time and, if he had survived, certainly would not have been the last (counterfactual) (BNC written, BMN 2400)

Wang's (1991) results indicated that factual conditions represented about 48% of the conditional sentences, predictive conditions about 27%, improbable conditions 14% and counterfactual conditions about 10%. If the results of the semantic analysis conducted in the present study are compared to the results presented in Wang's study, it can be noticed that although there is some variation in the proportions of the distribution of the different categories the underlying setting remains the same, the only significant exception being the reverse order of improbable and counterfactual conditions in the textbooks.

There seems to be a connection between the most commonly used verb-form combinations and the most commonly used semantic categories with the structure *simple present* + *simple present* usually representing factual conditions and *simple present* + *will/shall/be going* representing the predictive conditional. Even though the students probably do not consider the conditionals from a semantic point of view as they use them, there is a clear tendency towards using open conditions, perhaps because the formal structures that

denote factual and predictive conditions are somewhat simpler than the conditions denoting hypotheticality.

All in all, open conditions form approximately 80 per cent and hypothetical conditions about 20 per cent of all the *if*-conditional constructions in each source with not very much difference in the distributions of the different semantic categories. From a semantic angle, therefore, the *if*-conditionals represented in the textbooks seem relatively authentic.

6.3. Summary of findings

All in all, it can be concluded that the results of the present study seem very similar to the results of the previous studies mentioned in the theory section but there are also some differences that should be taken into consideration. There were some obvious examples of overuse of certain conditional constructions in the textbooks and then again there was a variety of constructions that were not included in the textbooks at all. Additionally, the construction *simple present + simple present* outnumbered the construction *simple present + will/shall/be going to* in the corpus materials but it was indeed the most frequently used *if*-conditional construction used in the textbooks. Still, the structure *simple present + simple present* occurred surprisingly often in the textbooks and, as a result, it can be argued that it is not a structure as neglected as sometimes claimed. As for the written part of the BNC corpus, the results seem to indicate that the random sample of the BNC shares the most essential formal characteristics with the samples taken from Brown and *LOB*, examined in earlier studies, at least regarding the verb-form combinations in conditional sentences.

When it comes to the authenticity of the language used in the textbooks concerning this specific linguistic feature, it is concluded that the textbooks do not exemplify the whole variety of possible *if*-conditionals, and some constructions are markedly more highlighted than in the corpus material. In addition, the number of *if*-conditionals presented in the eight textbooks under analysis is relatively low. Therefore, the findings of this study indicate that, from a formal point of view, the use of *if*-conditionals is not very authentic in the textbooks and does not give an exhaustive idea of different *if*-conditional patterns. The semantic analysis, on the other hand, reveals that although there are some differences in the distribution of the different semantic categories, the textbook material seems relatively similar to the corpus data. Thus, the textbooks can be considered authentic from a semantic perspective.

The analysis of the ICLE learner corpus gave us an insight into how language is used by a group of Finnish learners of English. The results show that the students use a surprisingly wide variety of *if*-conditional constructions regardless of the fact that the number of different constructions is not nearly as big in the textbook materials. This could mean that the influence of the authentic language available in alternative sources in everyday life is equally, if not more, important to EFL learners when compared to the materials provided by teachers.

7. Conclusion

The aim of this study was to examine whether there is correlation between the occurrence of *if*-conditionals in the textbooks and in the BNC and ICLE corpora and in that way conclude whether the textbook materials can be considered authentic in respect of this linguistic structure. A series of upper secondary textbooks, two random samples of the BNC corpus and a sample of the ICLE corpus were selected to analyse this research question. For the corpus analysis, both written and spoken parts of the BNC were studied separately in order to compare the use of *if*-conditionals in two types of authentic language use, written language and speech. A sample of the ICLE learner corpus was analysed to find out how *if*-conditionals are actually used by EFL learners in a written context. All the *if*-conditional constructions presented in these materials were examined quantitatively and they were classified according to the verb-form combinations they represented. In addition, a semantic analysis was conducted by identifying the different semantic categories and classifying the *if*-conditionals represented in the materials according to their semantic features. The results of all the sources were then compared to each other to find out to what extent the different *if*-conditional constructions had similar distributions in each source.

The results of this comparative study imply that although surprisingly many *if*-conditional types are represented in the textbooks, the full variety of possible constructions may still remain unfamiliar for the students, since the diversity of forms is not expressed in the books. The results of the corpus analysis including both the BNC and the ICLE corpora reveal that all in all there are almost twice as many distinct *if*-conditional constructions in use in the corpus data as represented in the textbooks. For this reason, the language presented in this textbook series cannot be considered fully authentic from the point of view of the formal features of *if*-conditional constructions. Still, the results of the semantic study indicate that the

distribution of the semantic categories is generally speaking similar in the textbooks and in the corpora, with only a few differences, and therefore, from a semantic viewpoint the textbooks can be considered authentic. It seems that even if the use of *if*-conditionals in the textbook series does not correspond to the way it is used in the BNC corpus, the data presented in the learner corpus indicates that students are, nevertheless, able to utilise a wide variety of different constructions in productive language use.

This study observed *if*-conditional constructions in one series of textbooks, two random samples of the BNC and in a sample of the ICLE. Of course, more reliable and extensive results would be achieved if more material was included in the analysis. The small number of *if*-conditionals represented in the textbooks serves only to give approximate guidelines as to what kinds of constructions can appear in teaching materials. In a more comprehensive study, the analysis should contain more textbook material, preferably with material representing different decades for a wider chronological perspective. The scope of the analysis could also be expanded almost infinitely. Further studies could include, for example, the analysis of the discourse functions of the conditional or the analysis of a wider variety of conditional constructions, not just *if*-conditionals. In addition, further research could be done on the realisation of the conditionals in natural spoken language of EFL learners or on the learning backgrounds of the students in terms of the teaching methodologies and pedagogical materials they have been exposed to in the past. Further research could also be conducted on comparing the influence of the grammatical constructions of the student's L1 on the acquired language, for example, in cases where students use verb-form combinations that are not typical of the target language. One additional topic of study could include creating more practical suggestions as to how conditionals could be taught in the classroom more effectively, yet this would probably require a qualitative approach.

English conditionals offer a wide and multi-layered field of study for anyone trying to capture the complicated nature of this seemingly straightforward structure. The study of *if*-conditionals alone has proved to be a task that could be refined much further than what has been done in this small-scale study. The ultimate aim of this kind of study has its roots in the actual classroom practice and, naturally, it would be useful if the results could also have some practical relevance in improving the teaching materials presented to students. Whereas this study has only scratched the surface of this linguistic structure, it has hopefully given some ideas as to how the analysis of *if*-conditionals can be carried out by using a textbook analysis and a corpus analysis as a starting point.

Bibliography

Primary sources:

The British National Corpus Online

[<https://bncweb.uta.fi/cgi-bin/BNCquery.pl?theQuery=search&urlTest=yes>] Accessed on 14 April 2007

The International Corpus of Learner English

Culture Café: Book 1. Ed. Raquel Benmergui, Anni Lampinen, Sanna Leinonen-Nuorgam, Jaakko Mäki, Teijo Päckilä & Riitta Silk. Helsinki: Otava, 2002.

Culture Café: Book 2. Ed. Raquel Benmergui, Anni Lampinen, Sanna Leinonen-Nuorgam, Jaakko Mäki, Teijo Päckilä & Riitta Silk. Helsinki: Otava, 2003.

Culture Café: Book 3. Ed. Raquel Benmergui, Anni Lampinen, Sanna Leinonen-Nuorgam, Jaakko Mäki, Teijo Päckilä & Riitta Silk. Helsinki: Otava, 2003.

Culture Café: Book 4. Ed. Raquel Benmergui, Anni Lampinen, Sanna Leinonen-Nuorgam, Jaakko Mäki, Teijo Päckilä & Riitta Silk. Helsinki: Otava, 2004.

Culture Café: Book 5. Ed. Raquel Benmergui, Anni Lampinen, Sanna Leinonen-Nuorgam, Jaakko Mäki, Teijo Päckilä & Riitta Silk. Helsinki: Otava, 2004.

Culture Café: Book 6. Ed. Raquel Benmergui, Anni Lampinen, Sanna Leinonen-Nuorgam, Jaakko Mäki, Teijo Päckilä & Riitta Silk. Helsinki: Otava, 2004.

Culture Café: Book 7. Ed. Raquel Benmergui, Anni Lampinen, Sanna Leinonen-Nuorgam, Jaakko Mäki, Teijo Päckilä & Riitta Silk. Helsinki: Otava, 2005.

Culture Café: Book 8. Ed. Raquel Benmergui, Anni Lampinen, Sanna Leinonen-Nuorgam, Jaakko Mäki, Teijo Päckilä & Riitta Silk. Helsinki: Otava, 2005

Secondary sources:

Bell, R. T. 1981 *An Introduction to Applied Linguistics – Approaches and Methods in Language Teaching*. Batsford Academic and Educational Ltd. London.

Brannen, J. (Ed) 1992 *Mixing Methods: qualitative and quantitative research*. Avebury Ashgate Publishing Limited. Aldershot Hants.

Denzin, N. K. and Lincoln, Y. S. (Ed) 1994 *Handbook of Qualitative Research*. Sage Publications, Inc. Thousand Oaks.

- Ferguson, G. 2001 'If you pop over there: a corpus-based study of conditionals in medical discourse'. *English for Specific Purposes* 20: 61-82.
- Fulcher, G. 1991 'Conditionals revisited'. *ELT Journal* Volume 45/2. Oxford University Press. 164-168
- Granger, S., Dagneaux, E. and Meunier, F. 2002. *International Corpus of Learner English*. Presses Universitaires De Louvain.
- Greenbaum, S. and Quirk, R. 1990. *A Student's Grammar of the English Language*. Longman Group UK Limited. Essex. 316-321
- Hill, L. A. 1960. "The Sequence of Tenses with *If* Clauses." *Language Learning*, 10/3 and 4:165-178.
- Kennedy, G. 1998. *An introduction to corpus linguistics*. Longman. London. 165-167
- Leech, G. and Svartvik, J. 1994. *A Communicative Grammar of English*. Second edition. Longman Group Limited. Harlow, Essex.
- Maule, D. 1998. 'Sorry, but if he comes, I go: Teaching conditionals'. *ELT Journal*, vol. 42/2: 117-123. Oxford University Press.
- Mauranen, A. 2004. 'Spoken corpus for an ordinary learner'. In J. McH. Sinclair (Ed.) *How to Use Corpora in Second Language Teaching*, 89-105. Amsterdam, John Benjamins.
- Mindt, D. 1986. 'Corpus, Grammar, and Teaching English as a Foreign Language'. In Leitner, G. 1986. (Ed.) *The English Reference Grammar*. Niemeyer. 125-135
- Nesselhauf, N. 2004. 'Learner Corpora and their potential for language teaching'. In J. McH. Sinclair (Ed.) *How to Use Corpora in Second Language Teaching*, 125-145. Amsterdam, John Benjamins.
- Oostdijk, N. 1991. *Corpus Linguistics and the Automatic Analysis of English*. Rodopi B.V, Amsterdam-Atlanta, GA
- Palmer, F.R. 1988. *The English Verb*. Second edition: Longman Inc. New York.
- Römer, U. 2005. *Progressives, Patterns, Pedagogy: A corpus-driven approach to English progressive forms, functions, contexts and didactics*. John Benjamins Publishing Company, Amsterdam-Philadelphia.
- Sinclair J. McH. (Ed.) 2004. *How to Use Corpora in Second Language Teaching*. Amsterdam, John Benjamins.
- Tan, M. 2005. 'Authentic language or language errors? Lessons from learner corpus'. *ELT Journal*, vol. 59/2: 126-134. Oxford University Press.
- Wang, S. 1991. 'A corpus study of English conditionals'. Unpublished MA thesis. Victoria University of Wellington.

Widdowson, H. 2000. 'On the limitations of linguistics applied'. *Applied Linguistics*. 21 (1), 7-8

Online sources:

Triangulation in research

[<http://www.tele.sunyit.edu/traingulation.htm>] Accessed on 17 May 2007

University of Oxford. 'What is the BNC?'

[<http://www.natcorp.ox.ac.uk/corpus/index.xml.ID=intro>] 2005. Accessed on 20 November 2006

Viestintätieteellinen tutkimus

[<http://www.uta.fi/viesverk/viesttiet/kaytannot/valinnat/maara.html>] Accessed on 3 May 2007

Appendices

Appendix 1**Table 3: The distribution of genre categories in the written data****David Lee's Genre Classification:**

Category	No. of words	No. of hits	Frequency per million words	No. of hits (extrapolated)	Frequency per million words (extrapolated)
W_fict_drama	45,757	1	21.85	821	17944.8
W_newsp_brdsht_nat_science	65,293	1	15.32	821	12575.62
W_non_ac_medicine	498,679	6	12.03	4,927	9879.3
W_newsp_brdsht_nat_arts	351,811	2	5.68	1,642	4667.85
W_advert	558,133	3	5.38	2,463	4413.46
W_newsp_other_commerce	415,396	2	4.81	1,642	3953.34
W_instructional	436,892	2	4.58	1,642	3758.82
W_ac_nat_science	1,111,840	5	4.5	4,106	3692.53
W_newsp_other_social	1,143,024	5	4.37	4,106	3591.79
W_ac_tech_engin	686,004	3	4.37	2,463	3590.8
W_fict_prose	15,926,677	66	4.14	54,193	3402.63
W_ac_polit_law_edu	4,640,346	18	3.88	14,780	3185.06
W_commerce	3,759,366	14	3.72	11,495	3057.8
W_pop_lore	7,376,391	27	3.66	22,170	3005.49
W_non_ac_nat_science	2,508,256	8	3.19	6,569	2618.87
W_news_script	1,292,156	4	3.1	3,284	2541.8
W_newsp_tabloid	728,413	2	2.75	1,642	2254.49
W_hansard	1,156,171	3	2.59	2,463	2130.57
W_misc	9,140,957	23	2.52	18,885	2066.01
W_non_ac_tech_engin	1,209,796	3	2.48	2,463	2036.13
W_non_ac_soc_science	4,187,649	10	2.39	8,211	1960.77
W_newsp_brdsht_nat_commerce	424,895	1	2.35	821	1932.48
W_ac_humanities_arts	3,321,867	7	2.11	5,748	1730.26
W_biography	3,528,564	7	1.98	5,748	1628.91
W_newsp_other_sports	1,027,843	2	1.95	1,642	1597.71
W_non_ac_humanities_arts	3,751,865	7	1.87	5,748	1531.96

W_religion	1,121,632	2	1.78	1,642	1464.12
W_ac_soc_science	4,247,592	7	1.65	5,748	1353.17
W_newsp_brdsht_nat_report	663,355	1	1.51	821	1237.8
W_newsp_other_report	2,717,444	3	1.1	2,463	906.48
W_newsp_brdsht_nat_misc	1,032,943	1	0.97	821	794.91
W_ac_medicine	1,421,933	1	0.7	821	577.45
W_non_ac_polit_law_edu	4,477,831	3	0.67	2,463	550.11
W_email	213,045	0	0	0	0
W_essay_school	146,530	0	0	0	0
W_essay_univ	65,388	0	0	0	0
W_fict_poetry	222,451	0	0	0	0
W_institut_doc	546,261	0	0	0	0
W_letters_personal	52,480	0	0	0	0
W_letters_prof	66,031	0	0	0	0
W_newsp_brdsht_nat_editorial	101,742	0	0	0	0
W_newsp_brdsht_nat_social	81,895	0	0	0	0
W_newsp_brdsht_nat_sports	297,737	0	0	0	0
W_newsp_other_arts	239,258	0	0	0	0
W_newsp_other_science	54,829	0	0	0	0
W_admin	219,946	0	0	0	0
Total	87,284,364	250	2.86	205274	2351.8

Appendix 2**Table 4: The distribution of genre categories in the spoken data****David Lee's Genre Classification:**

Category	No. of words	No. of hits	Frequency per million words	No. of hits (extrapolated)	Frequency per million words (extrapolated)
S_demonstratn	31,772	4	125.9	774	24375.68
S_interview	123,816	6	48.46	1,162	9382.44
S_tutorial	143,199	6	41.9	1,162	8112.46
S_parliament	96,239	4	41.56	774	8047.3
S_lect_polit_law_edu	50,881	2	39.31	387	7610.54
S_lect_soc_science	159,880	6	37.53	1,162	7266.05
S_unclassified	421,554	14	33.21	2,711	6430.08
S_classroom	429,970	13	30.23	2,517	5853.92
S_pub_debate	283,507	8	28.22	1,549	5463.46
S_speech_scripted	200,234	5	24.97	968	4834.74
S_meeting	1,384,302	34	24.56	6,583	4755.42
S_speech_unscripted	464,937	11	23.66	2,130	4580.78
S_conv	4,206,058	96	22.82	18,587	4419.13
S_consult	138,011	3	21.74	581	4208.71
S_interview_oral_history	815,540	17	20.85	3,291	4035.94
S_lect_humanities_arts	50,827	1	19.67	194	3809.31
S_brdrast_discussn	757,317	14	18.49	2,711	3579.25
S_brdrast_news	261,278	4	15.31	774	2964.14
S_sermon	82,287	1	12.15	194	2352.94
S_courtroom	127,474	1	7.84	194	1518.87
S_brdrast_documentary	41,540	0	0	0	0
S_lect_commerce	15,105	0	0	0	0
S_lect_nat_science	22,681	0	0	0	0
S_sportslive	33,320	0	0	0	0
Total	10,341,729	250	24.17	48405	4680.46

Appendix 3**Textbook materials***Culture Café* book 1

page	conditional construction
19	If anything goes wrong, she's going to be right there.
27	d'Lord won't mind if I have no toofs
42	But if you wait until I get back, I'll give you two marshmallows.
54	We'll share. If we're hungry.
54	If it wasn't boyfriends, though, he didn't know what it could be, apart from something bad.
55	"Shall we watch TV?" "If you want"
61	Anna was desperately worried that if Jamie no longer had the motivation to go to school he would never catch up
61	Tell them it's not the end of the world if they don't get exam grades
61	I honestly don't know what might have happened if we hadn't acted quickly
74	if a pub closes, this can make a huge dent in the local economy

 10
Culture Café book 2

57	I'd like her if she worked in Tesco.
57	If I hadn't had the support of my family, I wouldn't be here now.
80	if Bilbo asks a riddle Gollum can't answer, he will lead the Hobbit back to sunlight.
80	If Bilbo can't answer his riddles, well - he will simply eat the poor Hobbit.

 4
Culture Café book 3

7	Imagine if you could bring one index card of notes into the exam
14	If you do it, it's slummin'it.
14	See, if I'd started takin' school seriously I would have had to become different from me mates, an' that's not allowed.
22	she will allow me to enroll at NYU if I can maintain a B average for a year
23	if I open my mouth they'll know I'm not one of them
23	If the professor talks and talks and the textbook is seven hundred pages I'll surely be lost.
52	if I walk them they will just drag me around
66	if I overslept on a Saturday morning in the summer- he would be watering the windowboxes and would put the hose through my bedroom window and squirt me as an early morning call

- 78 If you commit a serious crime and are careless enough to leave behind a hair or
a drop of blood, Andy Urquhart and his colleagues could be your nemesis,
78 If his research is successful, a sample found at the crime scene could be used to
predict the hair, eye and skin colour of the person who left it
78 if things had gone differently, he might well have ended up neutering cats

11

Culture Café book 4

- 77 I won't reveal your name to anybody if you don't want me to
77 if somebody were to be swept away, the body would surely be found
downstream
78 she hoped that she would not fall asleep. If she did, though, she was sure that the
dog would wake her up

3

Culture Café book 5

- 17 if we are doing something important, we'll turn the phone off
33 if the solution is obvious the audience will get there before you and feel cheated
41 if a character delivered his lines while standing under the picture of the moon,
he was not thinking with his head but with his heart
41 if he stood under the sun, his words and actions were motivated by reason
49 But you don't really mean to say that you couldn't love me if my name wasn't
Ernest?

5

Culture Café book 6

- 7 Modern technology might have arisen 1,000 years earlier if the Greeks had
invented calculus
23 I'll scratch your back if you scratch mine
29 If disgust really is as crucial to our survival as some scientists believe, then
we're likely to be saying yuck for a very long time to come
53 if someone called up complaining about missing posters, or caps, or beach
towels, Hack had to take the call
53 if you quit your job and your replacement wasn't as competent, they sued you
for lost profits
53 If that's not your style, well...let's forget we spoke
54 if people realize every mall in the country's got Mercurys, we'll lose all that
demand we've worked so hard to build up
54 Even if we get found out, we burn a few million on legal fees, we get fined a
few million more...bottom-line, we're still way out in front
61 If all goes well, Bonk's boyfriend will pay for her goodies

61 If you don't see it on the people, it won't be a fashion
 69 The safety belt is strapped snugly across your body, and if that fails, the air bag
 will save your life-if it doesn't decapitate you
 79 if we made a mistake, we would regret it for quite a while

12

Culture Café book 7

5 If he is thirsty, we must thirst
 5 If he needs room, then we must fly
 5 And if he hungers, we must die
 12 if the rain forest disappears, we'll have to get our air in little bottles from the
 Evian company
 13 The ravenousness of rain-forest appetites is such that the floor of the jungle is
 nearly bare. If you don't count ants.
 34 If I need to drink tap water, I'll do it because I feel it's safe.
 40 And if recycling rates are to increase it has to be made easier, with better access
 to drop off sites and the introduction of nation-wide kerbside collections.
 40 Even if people can be persuaded to recycle more waste, it's not the end of the
 problem of local councils.
 69 And if he doesn't become a millionaire, at least he'll eat well.
 77 I've got the scientific name for it somewhere in my report here, if you're
 interested
 79 the cats had a field day with these feeble geckos- you can imagine, if any of you
 have ever owned a cat, the kind of joy these animals must have experienced to
 see their nemesis, this ultraquick lizard, and it's just barely creeping across the
 floor like a bug.
 79 the cats began to die..which to my mind would have been no great loss if it
 wasn't for the rats.

12

Culture Café book 8

6 The mood is festive, even if the first news imparted to us, almost gleefully, is
 that the polling station has run out of voting papers.
 6 Who minds waiting another hour, or two, or three, if one has already waited a
 lifetime?
 26 they could ride with the door off, giving them fresh air and a chance to get out if
 the plane was hit
 34 if you work at shaping yourself into an ideal candidate, you will start getting
 good job offers two to three years into your career
 45 I expected everyone to stop me in the street and ask me if I was the brother of
 the girl whose name was in the paper
 52 If I am in a place I know, like home, or school, or the bus, or the shop, or the
 street, I have seen almost everything in it beforehand
 52 if they are in the countryside, it might be

52 I wonder if Julie has given birth yet
53 But if I am standing in a field in the countryside I notice everything
53 the north-west and south-east corners are slightly lower than they would be if
the field was a flat inclined plane
53 it is very tiring if I am in a new place
53 if someone asked me afterwards what the cows looked like, I could ask which
one
72 If they believe they are expected to exhibit authoritarian conduct, abuse is likely
74 What if they start shooting at us?
74 they are not homicidal toward the police if we don't give them a big, fat target

15

tot.72 if-sentences