UNIVERSITÉ DU QUÉBEC À MONTRÉAL

THE FREQUENCY AND COLLOCATION OF MODAL VERBS IN ENGLISH AS A SECOND LANGUAGE TEXTBOOKS AS COMPARED TO STANDARD ENGLISH CORPORA.

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN DIDACTIQUE DES LANGUES

PAR

MAXIMILIANO EDUARDO ORLANDO

NOVEMBRE 2009

UNIVERSITY OF QUÉBEC IN MONTRÉAL

THE FREQUENCY AND COLLOCATION OF MODAL VERBS IN ENGLISH AS A SECOND LANGUAGE TEXTBOOKS AS COMPARED TO STANDARD ENGLISH CORPORA.

MASTER'S THESIS
PRESENTED AS A PARTIAL REQUIREMENT
FOR THE MASTER'S DEGREE IN LANGUAGE TEACHING

BY

MAXIMILIANO EDUARDO ORLANDO

NOVEMBER 2009

ACKNOWLEDGEMENTS

I am really indebted to the following people. They have all played a vital role in the making of this thesis. First, I would like to thank Professor France Boutin, who first was my Second Language Education teacher at UQAM and then reposed trust in me and in this project as Thesis Director. I am also most grateful to Professor Tom Cobb for his comments to improve this thesis and for all the hard work and time he has put in the construction of his site, *www.lextutor.ca*. This site has given me access to all the tools without which this work would not have been possible. I must really thank Professor Lori Morris for her guidance at the beginning of this project and Professor Andréanne Gagné for agreeing to read this thesis, for her ideas and suggestions. I have no words to express my gratitude to the assistant to the Department of Language Teaching at UQAM, Miss Diane Pellerin, for her academic counselling while I was a student.

There have also been people outside academia who have certainly been an important source of motivation in my decision to study for a Master's degree in Language Teaching. My parents, whose examples of hard work, bravery and determination in my childhood days gave me the patience and strength I needed to finish this thesis. Juan Ascárate, whose encouragement also helped me to see this research through. Patrick Dudgeon, whose priceless advice when I decided to go into ESL teaching when I was very young has had a bearing on my decision to learn more about second language education. Finally, my students, whose effort and determination to become better language users have been really inspiring.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

adv	Adverb
AmE	American English
BNC	British National Corpus
BNCW	British National Corpus Written
BNCS	British National Corpus Spoken
BrE	British English
EFL	English as a Foreign Language
ELT	English Language Teaching
ESL	English as a Second Language
inf	Infinitive
L1	First Language
L2	Second Language
MELSQ	Ministère de l'Éducation, du Loisir et du Sport du Québec
MVb	Modal Verb
n.d.	Not Dated
pers.pron	Personal Pronoun
Vb	Verb in Active Voice
past pple	Past Participle

LIST OF SYMBOLS

$\bar{x}$	Mean
$\sum$	Sum
n	Number
t	*t* score
r	Correlation
p	Significance

RÉSUMÉ

Le but principal de cette recherche est de savoir si les manuels scolaires d'anglais langue seconde du deuxième cycle de l'école secondaire du Québec approuvés par le Ministère de l'Éducation, du Loisir et du Sport du Québec en 2007 préparent les étudiants à utiliser les verbes modaux ainsi que les structures des verbes modaux par le biais de fréquences semblables à celles de l'anglais écrit ou de l'anglais oral. Dans ce but, les fréquences des verbes modaux et de leurs structures dans un corpus contenant les manuels scolaires déjà mentionnés ont été comparées avec celles dans le corpus d'anglais écrit et dans le corpus d'anglais oral du Corpus National Britannique en utilisant des outils statistiques.

Par rapport aux fréquences des verbes modaux, la seule différence significative entre les trois corpus (le corpus de manuels scolaires, le corpus d'anglais oral et celui d'anglais écrit) correspond aux contractions des verbes modaux. Tandis qu'elles sont très nombreuses en anglais oral, leurs fréquences dans le corpus de manuels scolaires sont très proches de celles du corpus d'anglais écrit. En effet, les manuels en question préparent rarement l'étudiant à utiliser les contractions *'ll* et *'d*. L' étude montre aussi que la préférence pour les verbes *can* et *will* parmi les locuteurs d'anglais langue seconde (Montero, Watts et Garcia-Carbonell (2007) et Debbie (2009)) coïncide avec les verbes modaux les plus fréquents dans le corpus de manuels scolaires.

Par contre, les fréquences des structures des verbes modaux dans ce dernier sont proches de celles de l'anglais oral. Effectivement, l' étude indique que les manuels scolaires préparent l'étudiant à utiliser des verbes modaux dans des fréquences semblables à celles de l'anglais écrit mais dans des structures plus simples.

ABSTRACT

The main objective of this study is to know whether the English as a second language (ESL) textbooks approved by the Ministère de l'Éducation, du Loisir et du Sport of Québec in 2007 and meant to be used by secondary school cycle two students in Québec as from 2008 prime the learner to use both modal verbs and modal verb patterns by means of frequencies similar to those in spoken English or in written English. To this end a corpus containing the above-mentioned textbooks was built, and the frequencies of modal verbs and of modal verb patterns in this corpus, in the written corpus and in the spoken corpus of the British National Corpus were compared by carrying out statistical analyses.

As regards modal verb frequencies, the only significant difference between the three corpora (the textbook corpus, the corpus of spoken English and the corpus of written English) is the frequency of contracted modal verb forms. While they are numerous in spoken English, their frequencies in the textbook corpus and in the written English corpus are very close. Indeed, the targeted ESL textbooks rarely prime the learner to use the contracted forms *'ll* and *'d*. The study also shows that *can* and *will* are the two most frequent modal verbs in the textbook corpus, which coincides with the finding that ESL speakers overuse these two modal verbs (Montero, Watts and Garcia-Carbonell (2007) and Debbie (2009)).

In contrast, the frequencies of modal verb patterns in the textbook corpus are similar to those in spoken English. The ESL textbooks selected for analysis tend to prime the learner to use modal verbs by means of frequencies similar to those in written English but in simpler structures.

INTRODUCTION

English as a second language (ESL) textbooks play a very important role in second language education since they are priming agents (Hoey, 2005). The theory of priming is based on the assumption that the frequent contact a speaker has with a word and its collocates prepares him/her to incorporate ready-made phrases to which he/she resorts to communicate in writing and in speaking (Hoey, 2005; Sinclair, 2007). Ideally, among other sources, ESL textbooks should prime the learner to use authentic English as native speakers do (Ministère de l'Éducation, du Loisir et du Sport du Québec (MELSQ), 2008). However, this is not always the case. Several studies have shown that instead of teaching authentic language, i.e. instead of priming the ESL learner to use the language as it is actually used, ESL textbooks follow a prescriptive approach, i.e. they show the learner how the language should be used (Tognini-Bonelli, 2001; Lyung 1990; Willis, 1990). What is more, it is the written language that ESL textbooks usually focus on to the extent that they prime the learner to speak written language (Willis, 1990; Lewis, 2000a; Lewis, 2000b). A case in point is dialogues: "Scripted dialogues usually have more in common with written language than with spoken language" (Willis, 1990, p. 125). Bearing this problem in mind, i.e. that ESL textbooks tend to prime the learner to use written English even for oral communication, it was thought that comparing the language of ESL textbooks with the language of written and spoken English corpora would indicate to what extent ESL textbooks prime second language (L2) learners to use written English, spoken English or both. To this end, a set of words whose frequencies and collocations could be analysed, a set of ESL textbooks and corpora of spoken and written English had to be selected.

The main objective of this study is to find out whether the ESL textbooks approved by the MELSQ in 2007 and meant to be used by secondary school cycle two students as from 2008 prime the learner to use both modal verbs and modal verb patterns by means of frequencies similar to those in spoken English or in written English. This will be done by carrying out two tasks. The first task will be to compare the frequencies of modal verbs in a corpus containing the above-mentioned textbooks with their frequencies in a corpus of spoken English, the British National Corpus of Spoken English (BCNS), and in a corpus of written English, the British National Corpus of Written English (BCNW). The second task will consist of comparing the frequencies of modal verb patterns in the same textbook corpus with their frequencies in the BNCS and in the BNCW. This will be done by using descriptive statistics and by calculating correlations and *t* scores.

The following is the outline of the dissertation. Chapter 1 will deal with the research problem. In it, the research questions and the objectives of this study will be discussed. Reference to its significance, limitations and assumptions will also be made. The theoretical background will be dealt with in Chapter 2. The methodology used will be introduced in Chapter 3. Findings will be presented in Chapter 4 and discussions and conclusions will follow in Chapter 5. The appendixes and references will be presented at the end.

CHAPTER I

THE RESEARCH PROBLEM

Before the research problem of the present study is discussed, a few terms will be defined, mainly those of *collocate*, *collocation*, *word form*, *lemma* and *token*. After discussing the research problem, the objectives and the research questions will be presented. Its significance will then be established and several assumptions will be made before presenting the limitations of its scope.

1.1 Definition of Terms

According to Sinclair, a "[...] word which occurs in close proximity to a word under investigation is called a collocate of it" (Sinclair, 1991, p. 170). The word under investigation Sinclair speaks of is called a *node*. Stubbs provides a more specific definition of *collocate* when he says that it is "[...] a word form or lemma which co-occurs with a node in a corpus" (Stubbs, 2001, p. 29).

A lemma is the concept traditionally referred to by the term *word*. Stubbs (2001) gives the example of the lemma *want*. A learner who knows the word *want* and who can distinguish between *want*, *wants*, *wanted* and *wanting* knows one lemma, *want*, and four word forms. According to Stubbs (2001), lemmas are not directly observable; they are abstract classes of word forms. Indeed, only word forms can be counted in a corpus: "Each word form which occurs in a text is a word-token" (Stubbs, 2001, p. 133).

Since the present study deals with the frequencies of modal verbs and of modal verb patterns in three different corpora (a textbook corpus, the BNCS and the BNCW) each modal verb form will be counted separately. For example, the lemma *will* has the following word forms: *will*, *'ll*, *will not* and *won't*. In the corpus of ESL textbooks selected for the present study, there are 4,236 tokens of *will*, 74 tokens of *'ll*, 157 tokens of *won't* and 157 tokens of *will not* per million words and therefore, by totalling these tokens, 4,624 tokens of the lemma *will* are obtained. This approach will help to present a detailed analysis of the frequencies of modal verbs and of modal verb patterns and to avoid generalisations that might hide individual differences between different word forms.

Interest in the collocates of a node (in the context of this research, in the collocates of each modal verb form, as discussed above) has led to the study of the smallest units of language such as morphemes, of lexical phrases and also of words in context, i.e. of collocations (Tognini-Bonelli, 2001). A *collocation* is the accompaniment of two or more words (Stubbs, 2001). Stubbs explains that not all collocations are of interest to corpus linguistics. He argues that only frequent events are worth studying and provides a statistical definition by saying that collocation is frequent co-occurrence (Stubbs, 2001, p. 29). Co-occurrence is the number of times a node is accompanied by one or more collocates. The frequent co-occurrence of modal verbs with their collocates in ESL textbooks enables us to find out modal verb patterns.

1.2 Statement of the Problem

Hoey (n.d.) explains that frequent encounters with co-occurrences of words or collocations in speaking and in writing result in mental associations of the node and its collocates. So much so that the knowledge a speaker has of a node is not that of the node in isolation, but of the node together with the words it comes along with. To support this theory, the author introduces two aspects of collocations: their pervasiveness and their subversiveness.

Hoey (2005) argues that collocations are pervasive for two reasons: firstly because it is probable that collocations or frequent co-occurrence of words are features of all lexical items, and secondly because clauses "[...] are made up of interlocking collocations such that they could be said to reproduce, albeit with important variations, stretches of earlier sentences [...]" (Hoey, 2005, p. 5). Therefore, sentences would not be abstract entities that contain collocations. On the contrary, sentences are built using collocations. However, the author says that while the pervasiveness of collocations has been given a lot of attention, little has been said to explain why language users reproduce them (Hoey, 2005; Hoey n.d.), i.e. why they prove to be subversive. Hoey proposes the theory of priming, which explains that this frequent co-occurrence of words becomes part of a language user's knowledge of a node:

> The ubiquity of collocation challenges current theories of language because it demands explanation, and the only explanation that seems to account for the existence of collocation is that each lexical item is primed for collocational use. By primed, I mean that as the word is learnt through encounters with it in speech and writing, it is loaded with the cumulative effects of those encounters such that it is part of our knowledge of the word that it co-occurs with other words. Hoey (n.d.).

Hoey (n.d.) goes on to argue that the way speakers are primed to use a word is not permanent. New encounters of a node with the same collocates will reinforce this priming of this node; while new encounters with other collocates will loosen it. Since the way a word is primed will vary according to the domain in which it is used, it can be assumed that the subversive nature of collocations stems from language users' sensitivity to frequency effects and to the range of contexts in which frequency effects can be found. Then, it could be said that there are as many collocational primings as there are individuals. However, first language (L1) speakers share a substantial body of collocational knowledge. It is the second language speaker's ability to gain collocational knowledge that determines, in part, language learning success. Having said that, Hoey's theory of priming inevitably leads to the following question: what are the implications of the assumption of the subversiveness of collocations in L1 and in L2 learning?

According to the author, native learners are primed differently from non-native ones. Firstly, L1 learners interact with and are primed by a larger number of other L1 speakers than L2 learners are. Secondly, L1 learners usually use the target language in a wider variety of

contexts and communicative situations, which can be recreated in the L2 classroom context and therefore will not be spontaneous. Thirdly, the range of lexical items, words and clauses to which the L1 learner is exposed is broader than in L2 learning. Consequently, L1 and L2 learners necessarily differ in their collocational knowledge.

Nonetheless, in recent years, e-communication has allowed for more substantial L2 contact with written and spoken examples of the target language. Even though this phenomenon has attenuated very heavy L2 learner reliance on traditional sources of priming such as textbooks, particularly in the case of more advanced learners able to engage in exchanges, language teaching materials still provide shortcuts to what are perceived to be reliable primings and, in some cases, are the only sources for them (Hoey, 2005).

The use of textbooks in ESL teaching may result from the fact that people generally accept lexical and grammar content of ESL textbooks as being what is needed for L2 learners. However, the perceived reliability of textbook primings has been recently challenged by various corpus linguists. Tognini-Bonelli (2001) cites a study by Lyung (1990), who analysed the vocabulary used in fifty well-known ESL textbooks and compared it with its frequency in a corpus compiled for the Cobuild project. The latter has been used for many purposes, including the production of the Collins dictionary, of grammars and of textbooks (see next paragraph). Lyung came to the conclusion that the fifty ESL textbooks under consideration contained a higher percentage of simple concrete words and a smaller proportion of abstract words than the Cobuild corpus, which leads Tognini-Bonelli to conclude that "[...] the language of ELT textbooks falls very short indeed in terms of representing natural usage" (Tognini-Bonelli, 2001, p. 40).

Willis (1990) arrives at a similar conclusion, supporting it with an example of a textbook-corpus mismatch. He explains that many pedagogic grammars explain to ESL learners that *some* is used in affirmative sentences, while *any* is used in negative and in interrogative ones. However, the author presents the concordances of the word *any* in the *Collins COBUILD English Course, Level 2*, where it is used in affirmative sentences much more frequently than in interrogative sentences (Willis, 1990, p. 49). Tognini-Bonelli (2001) deals with the same

words, but first goes through the presentation of *some* and *any* in a pedagogic grammar (Thomson and Martinet's *A Practical English Grammar*), and then goes into their use in general corpora to arrive at the following conclusion: "The evidence from the corpus, while indeed supporting some of the evidence from the pedagogic grammar, shows a far wider degree of variation with respect to the prescribed structures" (Tognini-Bonelli, 2001, p. 15).

In response to the problems evoked above, several writers propose the use of corpus-based syllabi and materials in ESL teaching: "The content and the sequencing of the curriculum and the weight given to different items in classroom activities can benefit from drawing their insights from corpus evidence" (Tognini-Bonelli, 2001, p. 40). Hoey (2005) proposes the use of corpus-based monolingual dictionaries and corpus-based descriptive grammars, and criticizes those grammars that present "[...] fabricated illustrations of grammatical points" (Hoey, 2005, p. 186). One of the benefits of using corpus-based materials is discussed by Hunston and Francis (2000): "If the learner wishes to sound 'natural', 'idiomatic', or 'native-like', it is argued, he or she needs to use the collocations, the phraseologies and the patterns of English that native speakers automatically choose" (Hunston and Francis, 2000, p. 268). Willis even uses authentic and spontaneously produced texts (by *spontaneously produced* texts the author means unscripted and unrehearsed conversations among native speakers which have been recorded in a studio) in his *Collins Cobuild English Course* (Willis, 1990).

The assumptions that L1 and L2 learners differ in their collocational knowledge and that ESL textbooks prime learners to become sensitive to collocations and to their frequencies in ways that may not be necessarily authentic, or as Hunston and Francis (2000) put it, natural, idiomatic, or native-like, to the extent that they may prepare them to speak written English, have raised our interest in knowing whether the ESL textbooks produced in Québec and meant to be used by secondary school cycle two students as from 2008 prime L2 learners to use written English or spoken English. Indeed, the issue that ESL textbooks teach English by illustrating how it should be used instead of how L1 speakers actually use it raised the problem that, in the end, the input that the ESL learner receives from textbooks consists mainly of examples of "correct" written English. As Willis (1990) explains, this emphasis on

correct written language leads the ESL learner to prime it and to employ it in oral environments (see *Introduction*). Is it true then, as Willis (1990) puts it, that there is an attempt to teach ESL learners to speak written language? As explained in the *Introduction*, to find this out a set of words had to be selected.

Modal verbs were targeted for several reasons. Firstly, modal verbs express meaning: obligation, permission, ability, need, etc, and what makes them different from other verbs is the fact that they are not inflected and that they are a closed class (Stubbs, 2001, p. 40). The choice of a compact and well-formed group of verbs limits the subjectivity involved in word choice. Secondly, most modal verbs occur frequently: *will*, *would*, *can*, *could*, *may*, *should*, *must* and *might* figure amongst the most frequent English verbs (Leech, Rayson and Wilson, 2001). Finally, modal verbs are relatively impervious to type or topic of text effects. ESL textbooks often differ in the choice of topics they deal with and include texts of diverse genres. This makes the study of the frequency of more lexically charged items almost impossible because it is difficult to guarantee that they will occur in different ESL textbooks. Ubiquitous and more lexically neutral, modal auxiliaries are ideally suited to cross-corpora comparisons.

1.3 Objectives

The two main objectives of this study are:

1. To know whether the ESL textbooks approved by the MELSQ in 2007 and meant to be used by secondary school cycle two students as from 2008 prime the learner to use modal verbs by means of frequencies similar to the ones in spoken English or in written English.

2. To know whether the ESL textbooks approved by the MELSQ in 2007 and meant to be used by secondary school cycle two students as from 2008 prime the learner to use modal verb patterns by means of frequencies similar to the ones in spoken English or in written English.

These objectives will be attained by carrying out two tasks:

1. By comparing the frequencies of modal verbs in a corpus made up of the ESL secondary school cycle two textbooks produced in Québec in 2007 with their frequencies in a corpus of written English (BNCW) and in a corpus of spoken English (BNCS).

2. By comparing the frequencies of modal verb patterns in a corpus made up of the ESL secondary school cycle two textbooks produced in Québec in 2007 with their frequencies in a corpus of written English (BNCW) and in a corpus of spoken English (BNCS).

Therefore, in the course of the analyses, two quantitative variables will be considered: the *frequency of modal verbs* and the *frequency of modal verb patterns*, as well as a qualitative one: the *mode of communication,* spoken or written.

The *frequency of modal verbs* corresponds to the number of occurrences per million words of the full and contracted forms of the affirmative and negative forms of the targeted modal verbs (see *Modal Verbs* in Chapter 2) – *can*, *could*, *may*, *must*, *might*, *will*, *would*, *shall*, *should*, *need*, and *dare* – in the textbook corpus, in the BNCW and in the BNCS. The *frequency of modal verb patterns* refers to the number of occurrences per million words of the patterns that each modal verb form and its collocates give rise to in the textbook corpus, in the BNCW and in the BNCS. A number of questions that will make these two objectives more specific need to be addressed.

1.4 Research Questions

The first set of questions refers to the frequencies of modal verbs. Since the textbook corpus is a collection of written texts, several questions arise. Do the ESL textbooks under scrutiny prime the ESL learner to use all the modal verbs selected for this study? If so, what frequencies do they have? Is there any modal verb form or lemma that these ESL textbooks tend to prioritise as far as frequencies are concerned? How do modal verb frequencies in the

textbook corpus compare with modal verb frequencies in written English and in spoken English? Are the frequencies of past modal verb forms in the textbook corpus different from their frequencies in spoken English and in written English? How about present modal verb forms? What are the learning implications of the answers to these questions?

The second set of questions deals with modal verb patterns. What frequencies do modal verb patterns have in the corpus of ESL textbooks in question? Are these frequencies similar to the ones in spoken English or in written English? Do modal verb patterns in the textbook corpus tend to be simple as they are supposed to be in spoken English or are they more complex as they are expected to be in written English? (see *ESL Textbooks as Priming Agents of Written and Spoken English: Modal Verbs* in Chapter 2). What are the possible implications of these answers in ESL learning, i.e. in the patterns textbooks prime students to use with individual modal verbs?

1.5 Significance of the Study

The fact that ESL textbooks are a key component of in-school ESL learning led us to think that the language contained therein merited careful research attention. Describing the frequencies of the modal verbs and of the modal verb patterns that occur in the ESL textbooks selected for study, measuring them and comparing them with their frequencies in a corpus of spoken English and in a corpus of written English to come to a better understanding of the modal verb forms and of the modal verb patterns that L1 and L2 speakers can observe and absorb is a scientifically interesting undertaking. Indeed, this study will indicate if the above-mentioned textbooks prime the learner to use modal verbs and their patterns in frequencies similar to those in written English or in spoken English.

1.6 Assumptions of the Study

The design of the present study is based on three assumptions. The first one is the fact that the two corpora chosen, the BNCW and the BNCS, are representative of standard spoken and written English. As it will be discussed in the *Theoretical Background* section, corpora

that contain a balanced mix of sources are needed for the present study. It is the study of specific types of English or genres that requires genre-related corpora.

Secondly, the frequencies of modal verbs and of modal verb patterns are believed to be relatively similar in the four textbooks under scrutiny for two reasons. Firstly, because all ESL textbooks writers in Québec need to follow the same set of guidelines established by the MELSQ and, secondly, because the textbooks selected for study are meant to be used at the same school level. Therefore, since it is up to the school commissions or private school authorities to use one or some of the ESL textbooks set by the MELSQ, the language they contain is expected to be uniform.

Thirdly, most learners who attend secondary school cycle two have studied English for several years and are at the end of a continuum that started in elementary school. As a consequence, it is assumed that, at this level, these learners are exposed to a wide range of modal verb forms and to the most complex and varied modal verb patterns in the whole school system. Indeed, in each of the ESL textbooks included in the present study, the use of modal verbs to perform different communicative functions is amongst the main grammar goals. In effect, the four textbooks contain tables and charts at the back that illustrate and explain how modal verbs function in English.

1.7 Limitations of the Study

The present study provides data about the frequencies of modal verbs and of modal verb patterns in the ESL textbooks to be used by secondary school cycle two students in Québec as from 2008. In total, they are four textbooks and, as explained in the *Methodology* section, a corpus containing the four of them was built. This study has allowed the comparison of these frequencies in this textbook corpus with those in a corpus of written English (BNCW) and in a corpus of spoken English (BNCS). Consequently, no conclusions may be drawn about any of the four textbooks in particular. Focusing on the frequencies in any of them would have limited the scientific scope of this research. At the same time, it must be said that its findings apply to the textbooks under scrutiny only, which means generalizations of the

sort "modal verb frequencies in ESL textbooks are similar to the ones in written English" will not be adequate for the purpose of this work. As Mahlberg puts it, in a corpus approach to grammar, grammar is "[...] seen as a set of generalizations about the behavior of words in texts, and these generalizations have to be related to the texts on which they are based" (Mahlberg, 2007b, p. 193).

In addition, the use of ESL textbooks in the classroom must be addressed. It needs to be borne in mind that while some teachers may opt not to use the entire book but to select a few sections and supplement them with tailor-made activities depending on the learners' needs and wants, other teachers may cover all the units, which might lead to much attention given to reading and writing activities. This is to say that chances are that the input the ESL textbook writers intend the learner to receive regarding modal verbs and modal verb patterns will not be equal to his/her actual priming and output. Consequently, both classroom management and textbook use will definitely have a considerable bearing on the learners' priming. As a result, the frequencies of modal verbs and of modal verb patterns in ESL textbooks may not have the same priming effect on different groups of learners.

A third limitation is that the frequencies of modal verbs and of modal verb patterns in written and in spoken English presented in this study are limited to British English (BrE). No information will be provided concerning any specific features of spoken or of written American English (AmE), Australian English or of any other variety of English. Considering that the textbooks selected for this study have been made in North America for students of Québec, it is presupposed that cultural differences will be reflected in the use of modal verbs in these textbooks and in BNCW and BNCS. Biber (1987) says that many Americans think BrE is more proper and formal than AmE, while many British speakers regard AmE as informal and relaxed. Precht (2003) argues that these stereotypes, that British speakers are polite and reserved whereas Americans are direct and brash, can be examined comparing stance in lexical verbs, adverbs, adjectives and nouns that belong to the stance categories of affect, evidentiality and quantifying as well as in modal verbs. In *Great vs. Lovely: Grammatical and Lexical Stance: Differences in American and British English*, the author compares the frequencies of the words that belong to these categories in a corpus of 100,000

words containing conversations at home in America and Great Britain taken from the Longman Corpus of Spoken and Written English. When dealing with modal verbs, the writer comes to the following conclusions. Americans tend to use more lexical verbs (28.67 per 1,000 words) than modal verbs (19.82 per 1,000 words) while British speakers use them equally (24.07 and 23.09 per 1,000 words respectively). The difference in the use of modal verbs between AmE and BrE is particularly shown in the case of *shall*. When defining modal verbs, for instance, Mathews (2003) says that *shall* and *shan't*, are limited to BrE and are not commonly used in AmE.

Finally, only the frequencies of modal verbs and of modal verb patterns have been identified and analysed, not their functions. Analysing the semantic preference and the semantic prosody of each occurrence of each modal verb form in the textbook corpus, in the written English corpus and in the spoken English corpus would have been a monumental enterprise. Nevertheless, *Further Research Possibilities* will be discussed at the end of this study.

CHAPTER II

THEORETICAL BACKGROUND

The present study was conceived and conducted within the framework of corpus linguistics methodology and the theory of priming to find out whether the ESL textbooks approved by the MELSQ in 2007 and meant to be used by secondary school cycle two students as from 2008 prime the learner to use modal verbs and modal verb patterns by means of frequencies similar to the ones in spoken BrE or in written BrE. As a result, it was necessary to provide some background regarding the following concepts and topics: *corpus*, *corpus linguistics methodology, modal verbs, modal verbs in written and in spoken English, pattern priming, ESL textbooks as priming agents of modal verbs in spoken and in written English*, and *the use of modal verbs amongst ESL/English as a Foreign Language(EFL) speakers*.

2.1. Why Using a Corpus?

Corpora have been defined in many different ways and from many different angles. The focus here will be on those definitions which are central to the framework of the present study. Stubbs (1996) defines a corpus as "[...] a collection of utterances and therefore a sample of actual behavior", and utterances for him are "[...] actual behavior, spoken or written" (Stubbs, 1996, p. 233). Therefore, a corpus may contain samples of authentic spoken or written language. Tognini-Bonelli (2001) explains that in all the definitions of "corpus" she has reviewed, three features are present: the authenticity of the texts any corpus contains,

the representativeness of the language of those texts, and finally, the criteria used to select those texts as samples of real language. For Tognini-Bonelli (2001) a corpus is

> [...] a computerised collection of authentic texts, amenable to automatic or semi-automatic processing or analysis. The texts are selected according to explicit criteria in order to capture the regularities of the language, a language variety or a sub-language. (Tognini-Bonelli, 2001, p.55).

Because the ESL textbooks that have been selected for analysis are meant to teach spoken and written English (see the section *ESL Textbooks as Priming Agents of Written and Spoken English: Modal Verbs* below), it was decided to compare the textbook corpus with two corpora: a corpus of written English (the British National Corpus of Written English), and a corpus of spoken English (the British National Corpus of Spoken English). The description of each corpus will follow.

The British National Corpus (BNC) contains both written text (BNCW) and sections of transcribed speech (BNCS) and has been

> [...] produced by an academic and industrial consortium consisting of Oxford University Press, Longman, Chambers Harrop, Oxford and Lancaster Universities and the British Library...and the Bank of English corpus compiled at the University of Birmingham [...] (Partington, 1998, p. 4).

The BNC is a 100,000,000 word corpus. However, in the present study, the versions available to the general public at *www.lextutor.ca* were used instead. In this site, the BNCW sample contains 1,007,000 words while the BNCS one is composed of 965,000 words. The former contains texts from different sources such as books and periodicals. These texts belong to different domains as for example fiction, economy, business, travel, etc, and come from different places of publication in the United Kingdom. The latter is composed of transcriptions taken from different contexts including meetings, conversations, presentations, etc, and from different domains as well. These transcriptions were obtained from speakers of both genders, of all ages and from different regions and social classes. It must be borne in mind that these two corpora intend to be representative of the English language as a whole; their use should be supplemented with genre-related corpora in the study of texts of specific genres or English domains.

Firth argues that the study of a language can be achieved by studying concrete texts. However, since these texts belong to what he calls "restricted languages", like the language of trade, of technology, etc, corpus linguistics findings should apply to the type of language in question (Luis Quereda Rodriguez Navarro, n.d.). For the analysis of the collocations of words in specific contexts, such as in Business English texts for instance, corpora built from texts that contain specific types of English should be used instead (Partington, 1998). Sinclair (1991) suggests that using general corpora to study infrequent words or genre-related words be avoided and recommends their use to study those words that occur in a wide variety of genres as, in this study, modal verbs. In effect, the invention of the computer has allowed researchers to collect large amounts of data that have neutralised the representativeness of specific types of English and genres and, as a consequence, to provide corpora which are meant to be "[...] a reasonable sample of one state of a language [...]" (Sinclair 1991, p. 24) as it is the case of the BNCW and of the BNCS. In fact, "corpus linguistics...mainly focuses on repeated and typical uses that do not only hold in one text, but are found across a number of texts in a corpus" (Mahlberg, 2007a, p. 221).

In addition, the BNCS and the BNCW provide collective language data without being representative of anybody's individual priming of the language. As Hoey explains, "[...] the existence of a priming for an individual cannot be demonstrated directly from corpus evidence, because a corpus represents no one's experience of the language" (Hoey, 2007a, p. 9). Hoey (2007a) goes on to say that "[...] a corpus, even a general corpus, can only point indirectly to the relative likelihood of a language user being primed in a particular way" (Hoey, 2007a, p. 9). In fact, when he says he uses the *Guardian* to collect data, he explains that "What such a corpus does do is permit a more detailed account of how a person might be primed by regularly reading the *Guardian* [...]" (Hoey, 2007a, p. 9). Sinclair (2007) puts it like this:

> [...] although a corpus cannot be primed, the individuals whose communicative experiences form the texts that make up the corpus are primed to behave as they do, and so the corpus is a record both of the routine and regular primings and the instances that go against the anticipated primings (Sinclair, 2007, p. 2).

In contrast, our corpus of ESL textbooks is not a sample of natural language use. This corpus was obtained by converting four ESL textbooks officially approved by the MELSQ in 2007 to be used at secondary school cycle two as from 2008 into text format. It consists of 254,237 words, and what makes it different from the two corpora described above, the BNCS and the BNCW, is the fact that while the latter contain texts produced by a vast number of speakers and writers with no pedagogical designs, the ESL textbook corpus is not really a collection of 'naturally occurring' language: it contains texts produced by a limited number of contributors with a didactic intention. It is 'author generated' for pedagogical purposes, and genre-related features, i.e. those of ESL textbooks as such, are more likely to occur. Therefore, noticeable differences between the corpus of ESL textbooks produced in Québec and the BNCW and the BNCS are expected to happen.

The four ESL textbooks selected for analysis have several features in common. They have been conceived within a task-based approach to learning and are all multi-skill. They are all divided into units which centre around a teen-related topic and which consist of a string of small tasks that culminates in a big project in which the learner applies what he/she has learnt throughout. This final project is a writing and/or a speaking task that results from the steady amount of input that the learner received after doing a good number of activities presented at different stages of each unit. These activities include plenty of reading, very few listening materials, a lot of "express-your-opinion" instances, and are accompanied by grammar explanations. All grammar points are meant to inform the learner with a view to adding accuracy to his/her writing and speech. If the main goal of these ESL textbooks had to be defined, it could be said that it is to teach ESL teen learners to perform a variety of tasks in English and, likewise, to achieve effective oral and written communication in English in a wide range of contexts. More information of the corpus containing these textbooks is forthcoming in the *Methodology* section.

2.2 Corpus Linguistics Methodology and the Study of Collocations

When the decision to compare the frequencies of modal verbs and of modal verb patterns in an ESL textbook corpus with those in a corpus of written English and in a corpus of spoken English was made, it became necessary to design a method to carry out the comparison. This resulted in an exploration of corpus linguistics and of the use of a corpus-driven approach to study collocations from both a technical and an epistemological point of view.

As regards the technical aspect, it is essential to refer to the methodology used to study corpora. The practice of corpus linguistics is relatively old. Tognini-Bonelli (2001) explains that the study of corpora can be traced back to a century ago, when Bréal observed language data to discover the laws that govern changes in meaning. Nonetheless, at that time, corpus analysis had certain limitations: collocations had to be collected, selected and classified by hand. Recently, corpus linguistics has been defined as "[...] a way of investigating language by observing large amounts of naturally-occurring, electronically-stored discourse, using software which selects, sorts, matches, counts and calculates." (Hunston and Francis, 2000, p. 15). Teubert (2007) says that "Corpus linguistics presupposes not only corpora, but also a methodology to extract relevant data. Corpora are principled collections of texts in electronic form, and the methodology employs computer programs, to detect and extract data" (Teubert, 2007, p. 89). Indeed, it is since the advent of computer engineering that it has been possible to process long texts, selecting collocations and producing concordances (Sinclair, 1991, p. 27). For Sinclair a concordance is a "[...] collection of the occurrences of a word-form, each in its own textual environment" (Sinclair, 1991, p. 32), or "[...] an index to the words in a text." (Sinclair, 1991, p. 170). Partington calls it a Key Word in Context Concordance (KWIC), which is "[...] a list of unconnected lines of text, which have been summoned by the concordance program from a computer corpus [...]" (Partington, 1998, p. 9). Concordancers allow the identification and quantification of collocations in a variety of different corpora which can then be analysed at the syntagmatic level. More specifically, in the present study, they allowed the identification and quantification of modal verbs and of modal verb patterns in a corpus of ESL textbooks and in a corpus of written English and a

corpus of spoken English. That required, therefore, the analysis of large amounts of linguistic data. It was then necessary to select a tool capable of performing the tasks Hunston and Francis (2000) and Teubert (2007) enumerate above, i.e. detecting, selecting, sorting, matching, counting and calculating all the modal verb forms as well as their collocates to detect patterns. To accomplish these tasks, it was decided to use a concordancer that is available online at *www.lextutor.ca*. This concordancer allowed the performance of all of the operations listed above. These operations will be described in detail in the *Methodology* section.

The second aspect of corpus linguistics methodology that has had an influence on the analytical choices made in the current study is its epistemological foundation. Hoey (2005) explains that traditional generative grammarians have focused on distinguishing grammatical from ungrammatical sentences and that they have supported their theories with invented examples which are sometimes difficult to imagine in an authentic context. The author argues that

> [...] corpus linguists in contrast have derived their goals from John Sinclair and his associates and in part from what concordancing software currently makes feasible. These linguists have typically seen their goal as the uncovering of recurrent patterns in the language, usually lexical but increasingly grammatical. They have not been much concerned with the single linguistic instance but with probability of occurrence, and their data have been always authentic. They have been concerned with fluency in language rather than creativity, and corpus models have been designed to account for the normal and the natural occurring. (Hoey, 2005, p. 152).

Sinclair (1991) expresses this contrast between the approach of traditional generative grammar and that of corpus linguistics by distinguishing the *open-choice principle* from the *collocational principle*. This distinction has had an important effect on Second Language Acquisition, defined as the study of "[...] the way in which people learn a language other than their mother tongue, inside or outside of a classroom [...]" (Ellis 2003, p. 3). The author explains that "virtually all grammars are constructed on the open-choice principle" (Sinclair 1991, p. 110). The author exemplifies this principle by referring to the tree structure whose

nodes are choice points that can be filled by virtually any word from a lexicon that satisfies their restraints. The study of collocations using concordancing software, however, has led Sinclair to propose the *idiom* or *collocational principle* as opposed to the *open-choice principle*. Sinclair argues that language production is a process in which the language user fills slots with words from the lexicon relying on a large number of pre-constructed phrases which are choices that can be combined (Partington, 1998, p. 19). These recurrent combinations of two or more words which "[...] seem to appear frequently in each other's company" (Hoey, 2005, p. 2) is a property of language and it is what Hoey calls collocations. Based on the analytical aspects of corpus linguistics stated in these paragraphs, it can be said that the approach of the present study will be descriptive. Indeed, following the collocational principle proposed by Sinclair, it can be argued that the recurrent combinations of modal verbs with their collocates in a corpus of ESL textbooks and in two corpora, one of spoken English and another one of written English, will allow the discovery of patterns which will be described instead of assessed in terms of their grammaticality.

2.3 Priming Patterns

In the previous section, the *collocational principle*, through which language users rely on a large number of pre-constructed phrases, was presented. In Chapter 1, a possible reason why learners learn these phrases was proposed: Hoey (2007a) suggests, through the theory of "priming", why speakers use these pre-fabricated phrases:

> "[...] each time we encounter a word (or syllable or combination of words), we subconsciously keep a record of the context and co-text of the word, so that cumulatively as we re-encounter the word (or syllable or combination of words) we build up a record of its collocations (Hoey, 2007a, p. 7-8).

Nevertheless, the author says it is not only lexical collocations that prime, but also "[...] the grammatical patterns a word appears in" (Hoey, 2007a, p. 8). The author says that these patterns contribute, among other elements that speakers encounter like sounds, words and phrases, to the construction of a grammar, and that, as a consequence, these "[...] grammars exist as a product of our primings. Each of us, presumably to different extents and with different outcomes and different degrees of regularity, constructs a grammar...out of the

primings we have [...]" (Hoey, 2007b, p. 31). Stubbs (2007b) adds that recurrent phrasal constructions are the result of the combination of "partly fixed lexical core plus other variable items" (Stubbs, 2007b, p. 163). In the present study, an analysis of these combinations will allow the identification and comparison of modal verb pattern regularities across corpora. It is true that the frequencies of modal verb patterns in the corpus of ESL secondary school cycle two textbooks produced in Québec in 2007 will be compared with these frequencies in a corpus of written English and in a corpus of spoken English. Nonetheless, it will also be possible to identify similarities and differences in the frequencies of modal verb patterns between the corpus of written English and the corpus of spoken English as well.

Another important issue regarding pattern priming through ESL textbooks needs to be addressed: the association between the frequency of occurrence of patterns and their priming. Sinclair, who prefers the term "adjustment" to that of "priming", says that there "[...] is a direct connection between experience and expectation, and as repeated instances crop up in further encounters, the adjustment is proportionate to the frequency of the events" (Sinclair, 2007, p. 1-2). Several studies have focused on the relation between pattern priming and the frequency of occurrence of patterns. In the article "*Frequency of Basic English Grammatical Structures*", Roland, Dick and Elman (2007) show that in many studies of distributional frequencies, "[...] language processing is closely tied to a user's experience, and that distributional frequencies of words and structures play an important (though not exclusive) role in learning". This link between the frequency of patterns and the way that language users and learners are primed to use them is referred to by Stubbs (2007a) in this way:

> "If a pattern occurs over and over again, in the language use of many different speakers, then it cannot be dismissed as mere performance. The frequent occurrence of lexical or grammatical patterns in a large text collection is good evidence of what is typical and routine in language use. Frequency in the corpus is observable evidence of probability in the system" (Stubbs, 2007a, p. 130).

The link between priming and frequencies, which has just been discussed, gave rise to a series of questions (see *Research Questions* in Chapter 1). The first one was which modal verb patterns ESL textbooks prime the ESL learner to use and what frequencies these patterns have. The second question referred to the frequencies these patterns also have in both spoken

and written English. Finally, the comparison of these frequencies in the three corpora was addressed. The following section will account for the modal verbs that have been chosen for study.

2.4 Modal Verbs

Modal verbs are used to express possibility, probability, necessity, certainty, etc. Hoye (1997) says that the notions of possibility, probability and necessity "[...] derive from the fact that human beings often categorize their attitudes and experience in terms of the way things might or must be, or might have been or must have been other than they actually are or were" (Hoye, 1997, p. 40). Nevertheless, these are not the only notions that convey modality. Stubbs defines modality as

> [...] the ways in which a language is used to encode meaning such as degrees of certainty and commitment, or alternatively vagueness and lack of commitment, personal beliefs versus generally accepted or taken for granted knowledge. (Stubbs, 1996, p. 202)

He explains that it can be encoded in "[...] noun and adjective morphology, in the verbal group, in modal verbs and in logical and pragmatic connectors." (Stubbs, 1996, p. 197). Focusing on fully lexicalized parts of speech such as nouns, adjectives, adverbs or verbs would not work in the present study because the selection of any sample of words from any of these categories would likely depend on the researcher's choice and on the specific topics chosen by the authors of the textbooks that compose the textbook corpus. In contrast, modal verbs are lexically bleached and, as such, relatively topic-independent.

Nevertheless, deciding what constitutes a modal verb and what does not is not an easy task. Perkins (1983) explains that many linguists have defined modal verbs as "[...] pre-eminent vehicles for the expression of modality in English, since the modals are the only modal expressions which constitute a reasonably well-defined class" (Perkins, 1983, p. 25). This, it was thought, would reduce subjectivity when the time came to select the verbs that would be analysed in the present study. Nonetheless, when the selection of modal verbs started, this closed class did not prove to be so well-defined. It is true that, as Palmer argues, modality is "[...] a semantic term ... to refer to the meanings of the modals"

(Palmer, 1979, p. 4): *can* expresses ability, capability and permission, *could* is used to express possibility, uncertainty, ability, *must* shows obligation and deduction, *may* and *might* convey possibility, uncertainty or permission, *should* is used for expectations, recommendations, unlikelihood, *will* expresses a promise, a spontaneous decision, an offer, a prediction or may be even used to ask for a favour, etc. Every modal verb form conveys meaning. Nevertheless, the sole concept of modal verbs as verbs that convey modality did not suffice to define them. For example, *have to* may express obligation or necessity, but it is not considered a modal verb by many linguists. A second criterion had to be considered to be able to select the modal verbs that would be included in the present study, and this was syntax.

English modal verbs evolved from Old English preterite present verbs, which had the same syntactic behavior as main verbs (Krug, 2000). According to Krug, they later developed to become a unique class:

> This gradual focusing of their category status consisted on the one hand in retaining characteristics that were previously available to all verbs, such as NOT negation or inversion in questions. On the other hand, the central modals also developed new distinctive properties by losing their nonfinite and tense forms, or their tensed forms lost largely their potential to refer to past time [...] (Krug, 2000, p. 45).

Some of these retained features Krug refers to are used by H.E. Palmer and by Chomsky in their definition of modal verbs: they allow for Subject-Verb inversion, have the negative form with -n't, are used emphatically when stressed in affirmations in spoken language, are used to avoid repetition of verbs, take no *s* for the third person singular, have no non-finite forms and cannot co-occur (Palmer, 1979). Nonetheless, some of these syntactic criteria were still not too convincing because they do not apply to all the verbs that are usually regarded as modal verbs. No samples of *mightn't* have been found in the corpora selected for this study, for instance.

Neither semantic nor syntactic criteria being sufficient to identify modal verbs, two more criteria were explored. The first one was intuition. Palmer says that "[...] native speakers are aware of the modals as a set "[...] and that "[...] the modals have a great deal in common

semantically [...]" (Palmer, 1979, p. 11). However, intuition seemed less reliable than the previous two criteria, meaning and syntax, since there would be as many systems of classifications of modal verbs as English speakers. The second and last criterion was untidiness. Indeed, deciding which verb can be a member of this closed class is not a black-and-white decision:

> "[...] it would be quite wrong to assume that the English systems or those of any other languages follow any absolute set of logical rules or fit into a rigid logical framework. For logical systems are idealized systems, while natural languages are notoriously untidy. What logic they have is likely to be fragmentary and inconsistent. (Palmer, 1979, p. 7).

If languages are untidy and it is impossible to establish a rigid framework to define modal verbs, then what remained to be done was to look into different classifications of modal verbs in prescriptive grammars (those which base their selection of modal verbs on pre-established rules) and in descriptive ones (those which base their selection of modal verbs on data gathered from corpora, among other authentic sources). All grammars consulted agree that *can*, *could*, *may*, *must*, *might*, *will*, *would*, *shall*, *should* are modal verbs, but there was some disagreement concerning *dare*, *need*, *ought to* and *used to*. Unlike most modal verbs, *dare* and *need* can sometimes take *-s* for the third person singular, *ought to* requires the full infinitive and *used to* may require the auxiliary *do*. Each of these verbs will be discussed below to be able to list then the modal verbs that have been included in the present study.

Dare behaves like a modal verb only in negative and in interrogative sentences, while in affirmative contexts, it behaves like an ordinary verb (Thomson and Martinet, 1980). After studying the occurrences of the modal verb *dare* in corpora, Partington (1998) concludes that it "is used, in the vast majority of corpus examples, in negative or interrogative contexts. Exceptions are fixed expressions such as *I dare say* and *I dare you (to)*." (Partington, 1998, p. 87). The same applies to the verb *need*: seldom is it used as an auxiliary in the affirmative. Thus both *need* and *dare* can "[...] be used as modals (with no *to*), fronting in questions and taking the negative directly, e.g. *Dare/need he go?*, He *daren't/needn't go.*" (Dixon, 1992, p. 172). Thus, it was decided to add *dare* and *need* to the list of modal verbs under scrutiny in interrogative and in negative contexts only.

As regards *used to*, Greenbaum and Quirk (1973) consider it a modal verb, specifying that it "[...] always takes the *to*-infinitive and occurs only in the past tense." (Greenbaum and Quirk, 1973, p. 37). However, Dixon explains that *used to* "[...] could be regarded as an aberrant member of the MODAL type." since "[...] it generally requires *do* in questions and negation, e.g. *Did he use (d) to do that?*, *He didn't use (d) to do that* (although some speakers say *Used he to do that?* and *He use(d)n't to do that*)." (Dixon, 1992, p. 171). Consequently, *used to* was omitted from the list of modal verbs of this study.

Finally, the last dubious case: *ought to*. It is seen as an alternative to *should*: "Where *should* appears, *ought to* can also be used" (Vince, 1994, p. 59). Dixon argues that "It is hard to discern any semantic difference between *should* and *ought to*, these two modals being in most contexts substitutable one for the other [...]" (Dixon, 1992, p. 171). However, the author says that *should* is preferred in negative sentences since *ought* and *to* are not likely to be separated. Indeed, in a corpus of about four million words of the concordancer in *www.lextutor.ca* that includes samples from the BNC used in this research, five occurrences of *ought not to* and two occurrences of *oughtn't to* were found, while there were 150 occurrences of *should not* and 213 occurrences of *shouldn't*. *Ought to* has been left out in this research.

As a result, the following modal verbs have been kept: *can*, *could*, *may*, *must*, *might*, *will*, *would*, *shall*, *should*, *need*, and *dare* (these last two only in interrogative and in negative sentences).

After discussing the definition and the selection of modal verbs for the purpose of this study, two issues need to be addressed: firstly, the use of modal verbs in ESL/EFL contexts and, secondly, ESL textbooks as priming agents of spoken and written English.

2.5 Modal Verbs in EFL/ESL Use

There is evidence that shows that ESL speakers, on average, use modal verbs less often than English native speakers do. Mason (1994) conducted a study in the United Kingdom in which eight Panjabi-speaking subjects aged five to nine were asked to retell six model stories. Forty-five of these stories were compared with nine stories produced by L1 interlocutors. Not only did the L1 subjects produce longer narratives, but also used more modal verbs (on average eight modal verbs were reported per narrative in L1 retellings while five were reported in L2 ones). Besides, further research has indicated that L2 learners tend to use modal verbs in contexts in which L1 speakers would normally use others (Hinkel 1995; Montero, Watts and Garcia-Carbonell, 2007; Debbie, 2009).

Debbie (2009) explains that native speakers are inclined to use *could* and *would* for requests due to the fact that the use of these modals is perceived to be more tentative, more polite and less assertive than the use of present tenses. The study the author conducted was meant to measure Brunei's non natives' use of the modal verbs *can, could, will* and *would* in requests for correction action in complaints in order to see whether Brunei ESL speakers prefer using past modal verbs as natives would do in this context. Data on the use of these modal verbs was collected from two groups. The first group consisted of 91 letters of complaint written by local Bruneians to the local English newspaper. The second group was composed of a questionnaire given to 88 non native undergraduates from first to third year at the local university. This questionnaire included a collection of sentences containing the modals *can, could, will* and *would* picked out from the 91 letters mentioned above. These modals had been removed from the sentences and the university students had to fill the gaps with either of the options given: *can* or *could, will* or *would*. The writer came to the following conclusions: in the first group, there was a significant preference of *can* over *could*, while there was no significant preference of *will* over *would*. However, in the second group, there was a significant preference of *will* over *would* and a definite preference of *can* over *could*. When the undergraduate students of the second group were asked why they had used present modal verbs instead of past modals, they mentioned grammatical reasons such as reported speech, direct speech, future tenses, and grammar in general. The writer concludes that the

learning of English modal verbs in the ESL classroom in Brunei focuses on form and structure rather than on its pragmatic intent.

A study conducted by Montero, Watts and Garcia-Carbonell (2007) amongst Spanish EFL university students shows EFL learners' preference for *can* and *will* rather than *could* or *would*. EFL students of computer science at a technical university in Spain were asked to subscribe to discussion forums on their field. A corpus of 878 messages was built and modal verbs were isolated in the context in which they were written. Then, they were sorted out into different semantic clusters (obligation and necessity, ability and possibility, epistemic possibility, volition and prediction and hypothetical modality) and the results were compared with those obtained from four corpora using the same methodology. These four corpora were the CMC (the computer conferencing system students and teachers use worldwide at the British Open University), the Lancaster-Oslo/Bergen (LOB) corpus of written texts, the London-Lund corpus of spoken texts and Piqué et al.'s corpus of research articles in two domains: medicine and biology. The authors found that there were significant differences between the Spanish corpus and the other four corpora to express epistemic possibility, volition and prediction. The most frequent modal verb amongst Spanish speakers was *will*. While it accounted for 32.62 % of all the occurrences of modal verbs in the Spanish corpus, in the corpus of research articles it did for 6.09 % and in the other three corpora its percentage of occurrence never got higher than 19 %. *May* and *might* accounted for 7.53 % of all the occurrences of modal verbs in the Spanish corpus, whereas in the corpus of research articles they represented 45.18 % of all modal verbs and in the written and spoken corpora they accounted for more than 15 % of all modal verb occurrences. It was only in the CMC that the percentage of occurrence of *may* and *might* was similar to that of the Spanish corpus (10 %), but, since the CMC is a corpus that contains worldwide exchanges between university students and professors, it is difficult to establish which occurrences were produced by native speakers or by ESL/EFL students. The authors conclude that Spanish speakers prefer using *can* in contexts where native speakers would normally use *may* or *might* because, in Spanish, permission and possibility is expressed by using the verb *poder* (the Spanish word for *can*). This, according to the authors, makes it difficult for EFL Spanish students to grasp the difference between *may*, *might* and *can*. As seen above, the study

conducted by Debbie (2009) concludes that ESL teaching has a bearing on the frequency of the modal verbs that ESL learners normally use. In contrast, the study by Montero, Watts and Garcia-Carbonell (2007) addresses the question of interference of Spanish ESL students' L1 in EFL modal verb frequencies. However, it is not always EFL/ESL learners' L1 or ESL teaching that influences the way modal verbs are used. Hinkel (1995) shows that ESL learners' cultural background also determines what modal verbs they use.

Hinkel (1995) compared ESL learners' use of modal verbs with English native speakers'. To this end the writer compiled 455 essays written by ESL learners and 280 essays produced by native speakers over a period of five years. The first group was composed of the following nationalities: Chinese (n= 195), Vietnamese (n= 33), Japanese (n= 87), Korean (n= 72), Indonesian (n= 63). All of them had high TOEFL scores. These students had had, on average, 12.6 years of English instruction and had lived in the United States for 2.4 years, except for the Vietnamese, who had been in that country for 6.1 years. The second group included native speakers brought up in the west of the United States. All students were asked to write about the following topics: the family, responsibilities, patriotism, traditions, education, racism, politics and relationships. The writer shows that the Asian non native speakers, who had been brought up within Confucian, Taoist and Buddhist constructs, used *must* and *should* when they dealt with the family, friendship and traditions, issues in which the notions of obligation and necessity are usually expressed by Asian learners. Native speakers, on the other hand, used the verb *need* in contexts where responsibility and necessity are imposed. The author shows that those non native ESL advanced students who had been exposed to L2 values and culture had not replaced the use of *must* and *should* with *need*. The writer finally suggests that modal verbs should be taught by referring to L1 cultural and social constructs.

These studies indicate that ESL speakers use modal verbs less frequently than native speakers do when performing certain tasks such as story-telling, that they prefer using present modal verb forms in situations where native speakers would normally use past modal verb forms to sound more polite, and that they would rather use certain modal verbs in contexts where native speakers would certainly use others. These studies also suggest that these findings may derive from the learner's culture, L1 and ESL education. As a consequence,

since ESL textbooks are priming agents in ESL education, they are expected to have an important effect on modal verb frequencies in ESL learners' language production. One of the research questions of this study was whether the frequencies of past and present modal verb forms in the textbook corpus were different from their frequencies in spoken English and in written English. It would also be interesting to know what the frequencies of *can* and *will* in the textbook corpus, in the BNCS and in the BNCW are, and how they compare to the frequencies of *could* and *would*.

Besides, as stated in the introduction and in Chapter 1, ESL textbooks may tend to teach the learner to speak written English, which indicated that another variable in addition to the three stated above (the learner's culture, ESL education and L1) had to be approached. That was the mode of communication, i.e. whether ESL textbooks prime the learner to use written English or spoken English.

2.6 ESL Textbooks as Priming Agents of Written and Spoken English: Modal Verbs

When dealing with the differences between spoken and written English within the context of ESL learning, Lewis points out that "[...] fluent speech consists largely of rapidly produced short phrases, rather than formally correct 'sentences'", and that "many of these phrases are relatively fixed, prefabricated lexical items" (Lewis, 2000a, p. 174). The predominance of fast short phrases in fluent speech is reflected in the frequency of short contracted forms of modal verbs in spoken English. In the frequency lists of spoken and of written English built by Leech, Rayson and Wilson (2001) and based on the 100,000,000 word BNC, the frequency of *'ll* is 3,066 tokens per million words in spoken English, while in written English it is 455. Besides, there are 1,194 tokens of the modal *'d* in spoken English but 256 in written English. Lewis goes on to explain that writing adds another element to fluency: accuracy and complexity. "Speech, naturally richer in semi-fixed expressions and multi-word adverbials, contains comparatively few of the verb + (adjective) + noun combinations which learners need if they are to write essays or reports [...]" (Lewis, 2000b, p. 186).

Given the preference of complex language in written English, it can be assumed that complex modal verb patterns will be more frequent in written English than in spoken English. Complex modal verb patterns such as *modal verb + be + past participle* (*MVb + be + past pple; "should be done"*) are expected to have a higher frequency in written English than in spoken English. However, the simple modal verb pattern *modal verb + infinitive* (*MVb + inf; "should do"*) is expected to be the most numerous in both. Hunston and Francis (2000) give the example of *would*, which has "[...] the pattern MODAL inf simply because it belongs to the word-class 'modal'" in the same way that "[...] *from* has the pattern PREP n because that is a pattern that all prepositions have" (Hunston and Francis, 2000, p. 202-203). In addition to the prevalence of fast short semi-fixed phrases in spoken English and to the complexity of patterns in written English, there are other factors that distinguish modal verb frequencies in spoken English from written English and that may also have an effect on the way ESL textbooks prime the ESL learner to use modal verbs.

First, modal verbs are more frequent in spoken English than in written English. The frequency lists by Leech, Rayson and Wilson (2001) show that the spoken corpus contains 19,543 modal verb tokens per million words, while the written one has 13,635 tokens. Second, there are differences in the frequencies of individual modal verbs between written English and spoken English: *can*, for instance, is more frequent in spoken English, while *may* is definitely more common in written English (Leech, Rayson and Wilson (2001)). Third, there are 8,403 tokens of past modal verbs per million words in spoken English and 6,073 in written English (Leech, Rayson and Wilson (2001)). Nonetheless, when the percentage of past modal verbs out of the total number of modal verbs per million words in each corpus was calculated, it was concluded that past modal verb forms accounted for 43 % of the total number of modal verb tokens in spoken English, whereas, in written English, they did for 44.5 %. This indicates that even though past modal verbs are more numerous in the spoken corpus, their frequency is proportionally close to the one in written English.

Comparing the frequencies of modal verbs and of modal verb patterns in the English as a second language textbooks approved by the MESLQ in 2007 to be used in the second cycle of secondary schools of Québec as from 2008 with their frequencies in the BNCW and in the

BNCS will allows us to see if these textbooks prime the learner to use modal verbs and modal verb patterns by means of frequencies similar to the ones in spoken English or in written English. Nevertheless, it is necessary to point out that the ESL textbooks that compose the corpus constructed for the purpose of this study were written to cater for a wide range of needs. Had a corpus of textbooks of English for oral communication been chosen, the frequencies of modal verbs and of their collocates in it might be expected to resemble their frequencies in spoken English. Conversely, if a corpus of textbooks to teach English correspondence had been used, it would not come as a surprise if the frequencies of modal verbs and of their collocates in it were similar to those in written English. What makes the learning goals of general ESL textbooks different from those of English for Specific Purposes is the fact that the differences between spoken English and written English goals are likely to be blurred owing to the fact that general ESL textbooks need to teach learners to communicate in both modes. The document entitled *English as a Second Language: Core Program, Enriched Program* produced by the MELSQ, introduces three competencies that students need to acquire at secondary school cycle two: interact orally in English, understand texts and write and produce texts. This can be done, the document explains, by using authentic texts (by texts they refer to any type of spoken, written or visual communication) that reproduce "[...] natural speech or writing as used by native speakers of English." (MELSQ, 2008, p. 53). They can be popular texts, as in the case of movies or cartoons, literary, such as novels or plays, or information-based like dictionaries or textbooks. The document goes on to say that these texts should facilitate the learning of functional language, of vocabulary, of language conventions and of form. The use of available resources such as grammar references and textbooks helps secondary school cycle two ESL students experiment with and reflect on the language by using *form-focused activities*: "Focus on form refers to communicative teaching that draws students` attention to the structure of the English language within the context of the interactive classroom" (MELSQ, 2008, p.40). This leads inevitably to one of the biggest challenges in textbook writing: what words to teach and in which patterns to teach them. ESL textbooks might not be able to prime all the modal verb forms in all the modal verb patterns that native speakers use. The selection of modal verb forms and patterns will necessarily depend on the mode of communication textbook writers

want to teach, whether spoken, written, or both, as it is the case of the ESL textbooks to be used in the second cycle of secondary schools of Québec (MELSQ, 2008). As a result, the writers that made the textbooks that compose the textbook corpus are thought to have made choices regarding the language they wanted the learner to prime. It is possible that they decided to include dialogues that contain fast speech as it happens in authentic spoken English and, as a consequence, to use contracted modal verb forms and simple modal verb patterns. It is also possible that they opted to select or produce texts that contain full modal verb forms and complex modal verb patterns which are more likely to appear in written English, or perhaps a combination of both.

Therefore, owing to these challenges, it goes without saying that the purpose of this study is far from criticising the choice of language made by the textbook designers of the books that compose the textbook corpus of the present study. On the contrary, a descriptive approach to textbook language will give insight into the way ESL learners are primed to speak and write English as far as the frequencies of modal verbs and of modal verb patterns are concerned.

CHAPTER III

METHODOLOGY

In Chapter 1, it was stated that the objective of this research is to know whether the ESL textbooks approved by the MELSQ in 2007 and meant to be used by secondary school cycle two students as from 2008 prime the learner to use modal verbs and modal verb patterns by means of frequencies similar to the ones in spoken English or in written English. It was also said that this would be accomplished by comparing the frequencies of modal verbs and of modal verb patterns in these textbooks with their frequencies in a corpus of written English (BNCW) and in a corpus of spoken English (BNCS). This chapter describes the methodology used to carry out these two tasks in two parts. The first one focuses on the construction of the textbook corpus that contains the above-mentioned ESL textbooks as well as on the way modal verbs were sorted out in the three corpora: the textbook corpus, the BNCS and the BNCW. The second part centres on the methodology used to obtain and analyse the data.

3.1 Corpora

While the BNCS and the BNCW were described in the *Theoretical Background* section, the constitution of the textbook corpus requires explanation.

The list of schoolbooks approved by the MELSQ in 2007 for use in 2008 and beyond contained four secondary school cycle two ESL textbooks fully accepted for use and available to the public at the time this research was started. This relatively limited number of approved textbooks, attributable to a recent reform in the school system that textbook

publishers had yet to respond to fully, eliminated potential sampling problems. All the available officially approved ESL textbooks for the upper secondary school level were integrated into a corpus, the textbook corpus. These books are: *Connecting through English Cycle Two* (328 pages), *Express Yourself Cycle Two* (198 pages), *Prime Time Cycle Two* (212 pages) and *Quest Cycle Two* (272 pages).

Once copies of the four books had been obtained, they were converted into text-only format using an automatic recognition system. Since they contained a lot of graphs, tables, diagrams, flow charts, etc, meant to illustrate a variety of language points, lots of words got jumbled on almost every page or were even lost during the conversion. As a consequence, careful proofreading of the contents of each converted page was required, followed by double checking for mistakes to make sure that no word was missing and that each word was in the right place. In the end, about 1000 pages were converted, resulting in a corpus of 254,237 tokens.

After the raw data had been collected, the textbook corpus was split into files of about 50,000 words each in order to fit them in the chosen concordancer in *www.lextutor.ca.* Next, the concordancer was used to search for the occurrences of each modal verb. All modal verb tokens were then sorted into as many files as modal verb forms found in the corpus, i.e. each file contained all the occurrences of only one modal verb form. For instance, there was one file grouping all the occurrences of the full affirmative form *will*, another one containing all the occurrences of the contracted affirmative form *'ll*, a third one including the occurrences of the full negative form *will not* and one grouping the occurrences of the contracted negative form *won't*. The same was done to group the rest of the modal verb forms. Next, all the word forms that were not modal verbs were identified and deleted from each list of occurrences, as for example *can* meaning *tin*, *may* meaning the month or the proper name *May*, *might* as *power*, *will* as *desire* or *mental power*, *'d* as the contraction of *had*, *need* and *dare* as nouns or finite verbs (all the occurrences of *daresay* were also deleted since it may be considered a compound finite verb). The same identification, sorting and vetting operations were conducted on the BNCS and on the BNCW, an easier undertaking since modal verb forms were automatically grouped by the corpus-based concordancer in *www.lextutor.ca.*

3.2 Analysis

Once each modal verb form in each corpus had a computer file that contained all its occurrences, the frequencies of the full and contracted forms of each modal verb form (affirmative and negative) in the three corpora (the BNCW, the BNCS and the textbook corpus) were calculated. To ensure comparability of results, the raw figures were converted to occurrences per million words. It was then possible to look for similarities and differences in the frequency of occurrence of all the targeted modal verb forms in the three corpora and then of the affirmative forms, of the negative forms, of the full forms, of the contracted forms, of the present forms and of the past forms separately. The reason for analysing each modal verb form separately was to provide a detailed description of modal verb frequencies.

The comparison of each group of modal verb forms across the three corpora was carried out by using the tools in *www.lextutor.ca* to calculate paired samples *t*-tests and correlations. The former enabled us to see whether the mean of the frequency of occurrence of all the modal verb forms and of each modal verb form (affirmative form, negative form, full form, contracted form, present form and past form) in the textbook corpus was significantly different from the ones in the other two corpora: the BNCW and the BNCS. The latter allowed us to see the relationship of the distribution of modal verb forms in the three corpora. Finally, these results were supplemented by comparing the figures in the table containing the frequencies of all the modal verb forms per million words so as to explore further differences across the three corpora. These analyses enabled us to answer all the research questions linked to the first objective.

Collocational patterns were then established by calculating the frequencies of the collocates following but not preceding each modal verb token in the textbook corpus. Following contexts were used in order to limit the study to the verb phrase and maintain both validity and reliability. Since modal verbs are integral parts of the verb phrase, it was logical to look for contextual variability within verb phrases only. After having identified all the patterns in which all modal verb tokens appeared in the textbook corpus, the frequency per million words of each pattern was calculated. Then, the percentage of occurrence of each

modal verb pattern out of the frequency of occurrence of modal verbs per million words was calculated. For example, there are 9,425 modal verb tokens per million words that belong to the pattern *MVb + inf* in the textbook corpus. Since there are 14,439 modal verb tokens per million words in the same corpus, those 9,425 occurrences of the pattern *MVb + inf* account for 65.3 % of all the modal verb patterns in this corpus. Only those patterns that had a frequency that was higher than one per cent out of the total number of occurrences of modal verbs per million words in this corpus were retained for analysis, i.e. *modal verb + infinitive* (*MVb + inf; "should do"*), *modal verb + be + past participle* (*MVb + be + past pple; "should be done"*), *modal verb + personal pronoun* (*MVb + pers.pron; "should I"*), *modal verb + adverb+ infinitive* (*MVb + adv + inf; "should never do"*), and *modal verb + have + past participle* (*MVb + have + past pple; "should have done"*). Then the frequencies per million words as well as the percentage of occurrence of these five modal verb patterns in the BNCW and in the BNCS were also calculated. This procedure allowed us to see the distribution of modal verb patterns in the three corpora. After that, the comparison of the mean frequency of each modal verb pattern across the three corpora was made by calculating the paired samples *t*-tests using the figures in Appendixes A to J, i.e. the figures representing the frequencies of modal verb patterns per million words verb by verb. The analyses of these results were supplemented with the discussion of the frequencies of these patterns containing individual modal verb forms presented in the tables in the same appendixes. This helped to see in detail what patterns the ESL textbooks that make up the textbook corpus intend the ESL learner to prime when using specific modal verb forms and if the frequencies of these patterns are similar to their frequencies in written English or in spoken English. This way, the research questions related to the second objective were answered.

CHAPTER IV

FINDINGS

As explained in Chapter 3, the frequencies of the targeted modal verb forms per million words in the BNCW, the BNCS and the textbook corpus will be presented first. These figures will be followed by the calculation of the *t*-scores and of the correlations of the frequencies of all the modal verb forms together and then of the affirmative, negative, full, contracted, present and past modal verb forms separately across the three corpora. This will be done in the first section of this chapter, entitled *Frequency of Occurrence*. This information will be used to answer all the research questions related to the first objective, which is to know whether the ESL textbooks approved by the MELSQ in 2007 and meant to be used by secondary school cycle two students as from 2008 prime the learner to use modal verbs by means of frequencies similar to the ones in spoken English or in written English.

The second part, *Analysis of Pattern Differences between Corpora*, will look at the complexity of the verbal patterns in which the modal verbs occur. As explained in Chapter 3, *t*-scores will be calculated to know whether the mean of the frequency of occurrence of each modal verb pattern across the three corpora is statistically different. These results will answer the research questions connected with the second objective, which is to know whether the ESL textbooks approved by the MELSQ in 2007 and meant to be used by secondary school cycle two students as from 2008 prime the learner to use modal verb patterns by means of frequencies similar to the ones in spoken English or in written English. Finally, a table summarising all significant differences in both parts will be presented at the end.

4.1 Frequency of Occurrence

As already explained in the introduction to this chapter, the frequencies of all the modal verb forms per million words in the three corpora will be presented first. This will be followed by the calculation of the correlations and of the *t*-scores of these frequencies across the three corpora. The same statistical analyses will be done when dealing with the frequencies of the affirmative and negative modal verb forms in the second part, with the frequencies of the present and past modal verb forms in the third part and with the frequencies of the contracted and full modal verb forms at the end of this section.

4.1.1 Analysis of All Modal Verb Forms

The number of occurrences of all the modal verb forms in the three corpora was totalled and then converted to a frequency rate per million words in each corpus (see Table 4.1).

Table 4.1
Occurrence of All Modal Verb Forms per Million Words

Modal Verbs	BNC Written	BNC Spoken	Textbook Corpus
Can	2,004	3,210	3,324
Cannot	258	81	192
Can't	200	1,216	354
Total	**2,462**	**4,507**	**3,870**
Could	1,341	1,405	1,089
Could not	135	17	74
Couldn't	107	401	137
Total	**1,583**	**1,823**	**1,300**
May	973	407	696
May not	49	49	70
Total	**1,022**	**456**	**766**
Might	443	642	444
Might not	16	20	19
Total	**459**	**662**	**463**

Modal Verbs	BNC Written	BNC Spoken	Textbook Corpus
Shall	211	285	15
Shall not	9	2	4
Shan't	3	17	0
Total	**223**	**304**	**19**
Should	1,286	976	869
Should not	68	23	62
Shouldn't	17	142	47
Total	**1,371**	**1,141**	**978**
Must	803	619	475
Must not	35	5	86
Mustn't	7	21	47
Total	**845**	**645**	**608**
Will	2,878	1,735	4,236
'll	393	3,111	74
Will not	134	33	157
Won't	92	557	157
Total	**3,497**	**5,436**	**4,624**
Would	2,185	2,324	1,620
'd	165	1,177	11
Would not	131	30	78
Wouldn't	69	503	94
Total	**2,550**	**4,034**	**1,803**
Dare	0	6	4
Dare not	2	0	4
Daren't	0	1	0
Total	**2**	**7**	**8**
Need	0	1	0
Need not	17	0	0
Needn't	1	9	0
Total	**18**	**10**	**0**

Table 4.2 presents the correlations of all the modal verb forms across the three corpora. No significant differences have been reported: there is significant correlation between the three corpora ($r > .66$, $p < .0001$).

Table 4.2
Correlation Matrix for All Modal Verb Forms

	Textbook Corpus	BNC Written
Textbook Corpus	-	
BNC Written	**.93***	-
BNC Spoken	**.66***	**.71***

***P* < .0001**

To check for significant differences in the mean frequency of the targeted modal verb forms between corpora, paired sample *t*-tests were conducted. Table 4.3 indicates that the mean frequency of all the modal verb forms in the textbook corpus is not statistically different from their mean frequency in written English or in spoken English ($p > .22$).

Table 4.3
***T*-Test Results for the Frequencies of All Modal Verb Forms**

	$\bar{x}a - \bar{x}b$	*df*	*t*	*p* Value.
BNC Written - BNC Spoken	-151.30	64	-.76	**.22**
BNC Written - Textbook Corpus	-12.33	64	-.06	**.47**
BNC Spoken - Textbook Corpus	138.96	64	.62	**.26**

So far, it may seem that the analysed ESL textbooks prime the learner to use modal verbs by means of frequencies similar to those in written and in spoken English. The analysis of each group of modal verb forms will give a closer look into modal verb frequencies.

4.1.2 Analysis of Affirmative and Negative Modal Verb Forms.

The same statistical analyses were performed to compare the frequencies of the affirmative and of the negative modal verb forms. In the *Methodology* section, it was explained that the reason for analysing each form separately was to provide a detailed description of modal verb frequencies.

Table 4.4 shows the correlations of the affirmative modal verb forms in the three corpora. The results indicate that there are three significant levels. A high correlation can be observed between written English and the textbook corpus (r = .92, p < .0001). There is also correlation between written English and spoken English. However, it is not as strong as the one between the textbook corpus and written English (r = .56, p < .05). No significant correlation between the distribution of affirmative modal verb forms in spoken English and in the textbook corpus has been reported (r = .54, p > .05).

Table 4.4
Correlation Matrix for the Affirmative Modal Verb Forms

	Textbook Corpus	BNC Written
Textbook Corpus	-	
BNC Written	.92**	-
BNC Spoken	**.54**	.56*

***p* < .05 ****p* < .0001**

Table 4.5 shows the t-scores for the affirmative modal verb forms. No significant differences can be observed (p > .27). The results presented in Tables 4.4 and 4.5 indicate that it is not the mean frequency of the affirmative modal verb forms per million words in each of the three corpora that is significantly different. It is the frequencies of individual affirmative modal verb forms in the spoken English corpus-textbook corpus comparison that are significantly different, which is reflected in the correlation between these two corpora. These differences in the frequencies of individual affirmative modal verb forms are expected to be accounted for in the analyses of the frequencies of the past, present, full and contracted modal verb forms and in the discussion of Table 4.1 in Chapter 5.

Table 4.5
***T*-Test Results for the Frequencies of the Affirmative Modal Verb Forms**

	x̄a - x̄b	*df*	*t*	*p* Value.
BNC Written - BNC Spoken	-247.38	24	-.62	**.27**
BNC Written - Textbook Corpus	-13.46	24	-.03	**.48**
BNC Spoken - Textbook Corpus	233.92	24	.49	**.31**

As far as the distribution of the negative modal verb forms is concerned, Table 4.6 provides their correlations in the three corpora. There is significant correlation in the three comparisons, but the one between BNCW and BNCS ($r = .47$, $p < .05$) is less strong than the correlation between the textbook corpus and BNCW ($r = .83$, $p < .0001$) and between the textbook corpus and BNCS ($r = .81$, $p < .0001$).

Table 4.6
Correlation Matrix for the Negative Modal Verb Forms

	Textbook Corpus	BNC Written
Textbook Corpus	-	
BNC Written	**.83****	-
BNC Spoken	**.81****	.47*

***p* < .05 ****p* < .0001**

The *t*-score results are shown in Table 4.7 below. No significant differences can be reported between the three corpora either ($p > .10$).

Table 4.7
***T*-Test Results for the Frequencies of the Negative Modal Verb Forms**

	x̄a - x̄b	*df*	*t*	*p* Value.
BNC Written - BNC Spoken	-88.85	38	-1.27	**.10**
BNC Written - Textbook Corpus	-11.60	38	-.45	**.32**
BNC Spoken - Textbook Corpus	77.25	38	1.09	**.14**

4.1.3 Analysis of Present and Past Modal Verb Forms

In the *Theoretical Background*, the fact that ESL speakers tend to use present modal verb forms in contexts where past modal verbs would be preferred by native speakers was discussed. At the same time, it was argued that the percentage of past modal verb forms in spoken English and in written English is very close. One of the research questions was whether the frequencies of past modal verb forms in the textbook corpus were different from their frequencies in the spoken English corpus and in the written English one. Therefore, present and past modal verb forms were analysed separately.

The correlations between present modal verb forms in the three corpora are shown in Table 4.8. There is significant correlation between the three corpora, but the one between the textbook corpus and written English ($r = .96$, $p < .0001$) is stronger than the correlation in the textbook corpus-spoken English comparison ($r = .42$, $p < .05$) and in the spoken English-written English one ($r = .50$, $p < .05$).

Table 4.8
Correlation Matrix for the Present Modal Verb Forms

	Textbook Corpus	BNC Written
Textbook Corpus	-	
BNC Written	**.96****	-
BNC Spoken	.42*	.50*

***p* < .05 ****p* < .0001**

The *t*-scores (see Table 4.9) show no statistically significant differences between the textbook corpus and the written English corpus or between the textbook corpus and the spoken English corpus as far as the mean frequency of present modal verb forms is concerned ($p > .32$).

Table 4.9
***T*-Test Results for the Frequencies of the Present Modal Verb Forms**

	x̄a - x̄b	*df*	*t*	*p* Value.
BNC Written - BNC Spoken	-104.50	38	-.45	**.32**
BNC Written - Textbook Corpus	-25.30	38	-.10	**.46**
BNC Spoken - Textbook Corpus	79.20	38	.29	**.38**

Table 4.10 presents the correlation of past modal verb forms. It shows that it is significant in the three comparisons ($p < .0001$, $r > 87$).

Table 4.10
Correlation Matrix for the Past Modal Verb Forms

	Textbook Corpus	BNC Written
Textbook Corpus	-	
BNC Written	**.99***	-
BNC Spoken	**.87***	**.88***

***p* < .0001**

Table 4.11 indicates that the frequencies of the past modal verb forms are not significantly different in any of the three comparisons ($p > .16$). In conclusion, tense is not a variable that seems to indicate any significant difference in the frequencies of modal verb forms in the three corpora.

Table 4.11
***T*-Test Results for the Frequencies of the Past Modal Verb Forms**

	x̄a - x̄b	*df*	*t*	*p* Value.
BNC Written - BNC Spoken	-141.41	22	-.49	**.31**
BNC Written - Textbook Corpus	118.25	22	.46	**.32**
BNC Spoken - Textbook Corpus	259.66	22	1.01	**.16**

4.1.4 Analysis of Full and Contracted Modal Verb Forms

Several differences between spoken English and written English were discussed in the *Theoretical Background*. One of these differences was the fact that modal verb contractions are a feature that distinguishes spoken English from written English. This fact was based on the lists drawn up by Leech, Rayson and Wilson (2001), according to which the frequencies of *'d* and of *'ll* are much higher in the spoken English corpus than in the written English corpus of the 100,000,000 word BNC. As a result, the correlations and the *t*-scores of the contracted modal verb forms and of the full modal verb forms across the three corpora are expected to be indicators of possible differences in modal verb frequencies between the textbook corpus and the spoken English and the written English ones.

Table 4.12 shows the correlations of the contracted modal verb forms. While the correlations between the textbook corpus and BNCW and between the textbook corpus and BNCS are not significant ($p > .05$), the spoken English corpus correlates significantly with the written English corpus ($r = .98$, $p < .0001$).

Table 4.12
Correlation Matrix for the Contracted Modal Verb Forms

	Textbook Corpus	BNC Written
Textbook Corpus	-	
BNC Written	**.39**	-
BNC Spoken	**.27**	.98*

***p* < .0001**

Table 4.13 presents the *t*-score results. The mean frequency of these forms in the textbook corpus is statistically different from their mean frequency in the spoken English corpus ($p < .05$) but not from the one in the written English corpus ($p = 40$). Indeed, the table shows that there are more contracted modal verb forms in the spoken English corpus than in the textbook corpus and than in the written English corpus, whereas their frequencies in the textbook corpus and in the written English corpus are very close.

Table 4.13
***T*-Test Results for the Frequencies of the Contracted Modal Verb Forms**

	x̄a - x̄b	*df*	*t*	*p* Value.
BNC Written - BNC Spoken	-554.63	20	-1.96	**.03***
BNC Written - Textbook Corpus	12.09	20	.25	.40
BNC Spoken - Textbook Corpus	566.72	20	2.01	**.02***

***p* < .05**

Table 4.14 indicates that, by contrast, the correlations of the full modal verb forms are significant in the three comparisons ($r > .85$, $p < .0001$).

Table 4.14
Correlation Matrix for the Full Modal Verbs Forms

	Textbook Corpus	BNC Written
Textbook Corpus		
BNC Written	**.93***	
BNC Spoken	**.85***	**.89***

***p* < .0001**

As opposed to the contracted modal verb forms, no significant differences in the mean frequencies of the full modal verb forms in the textbook corpus - BNCS ($p = .40$) or in the BNCW - BNCS ($p = .42$) comparisons have been reported.

Table 4.15
***T*-Test Results for the Frequencies of the Full Modal Verbs Forms**

	x̄a - x̄b	*df*	*t*	*p* Value.
BNC Written - BNC Spoken	50.36	42	.19	**.42**
BNC Written - Textbook Corpus	-24.55	42	-.08	**.46**
BNC Spoken - Textbook Corpus	-74.9	42	-.25	**.40**

4.2 Analysis of Pattern Differences between Corpora

The fact that fluency is an important feature of speaking and that complexity is an element that is usually incorporated in writing (Lewis, 2000b) was discussed in the *Theoretical Background*. Indeed, even though it was argued that *MVb + inf* was expected to be the most frequent pattern in both spoken and written English, the frequencies of more complex patterns such as *MVb + be + past pple* were supposed to be higher in written English.

As explained in the *Methodology* section, only the modal verb patterns that had a frequency that was higher than one per cent out of the total number of occurrences of modal verbs in the textbook corpus were going to be retained for analysis. The example that was provided was that of the pattern *MVb + inf*, whose 9,425 occurrences per million words account for 65.3 % of the 14,439 modal verb tokens per million words in the textbook corpus. These modal verb patterns are: *MVb + inf, MVb + be + past pple, MVb + pers.pron, MVb + adv + inf*, and *modal verb + have + past participle*. The frequencies of these modal verb patterns per million words across the three corpora can be observed in Table 4.16. The percentage of occurrence of these patterns out of the total number of occurrences of modal verbs in each of the three corpora may be observed in Table 4.17.

Table 4.16
Frequencies of Modal Verb Patterns per Million Words

Modal Verb Patterns	BNC Written	BNC Spoken	Textbook Corpus	Total
MVb + inf	**8,289**	**12,123**	**9,425**	29,837
MVb + pers.pron.	407	**2,042**	**1,807**	4,256
MVb + be + past pple	**2,312**	487	522	3,321
MVb + adv + inf	**1,078**	**1,027**	**673**	2,778
MVb + have + past pple	**575**	**757**	167	1,499
Total	**12,661**	**16,436**	**12,594**	**41,691**

Table 4.17
Frequencies of Modal Verb Patterns as a Percentage of the Total Number of Occurrences of Modal Verbs across the Three Corpora

Modal Verb Patterns	BNC Written	BNC Spoken	Textbook Corpus
MVb + inf	**59**	**63.7**	**65.3**
MVb + pers.pron.	2.9	**10.7**	**12.5**
MVb + be + past pple	**16.4**	2.6	3.6
MVb + adv + inf	**7.7**	**5.4**	**4.7**
MVb + have + past pple	**4.1**	**3.9**	1.2
Total	**90.1**	**86.3**	**87.2**

The *t*-scores for the modal verb patterns under consideration will indicate whether the differences in their mean frequencies between the three corpora are significant. Table 4.18 shows the t-scores for the pattern *MVb + inf*. As expected, this pattern is the most frequent in the three corpora and no significant differences have been reported ($p > .17$).

Table 4.18
***T*-Test Results for the Pattern MVb + inf**

	$\bar{x}a - \bar{x}b$	*df*	*t*	*p* Value.
MVb + inf				
BNC Written - Textbook Corpus	-37.86	58	-.26	**.39**
BNC Spoken - Textbook Corpus	89.93	58	.54	**.29**
BNC Written - BNC Spoken	-127.80	58	-.93	**.17**

Similar comments may be made about the pattern *MVb + adv + inf*. Table 4.19 indicates that the mean frequencies of this pattern in the three corpora are not significantly different ($p > .19$). It seems that the position of the adverb between the modal verb and the infinitive is not a distinguishing feature of BNCW, of BNCS or of the textbook corpus.

Table 4.19
***T*-Test Results for the Pattern MVb + adv + inf**

	x̄a - x̄b	*df*	*t*	*p* Value.
MVb + adv + inf				
BNC Written - Textbook Corpus	13.50	58	.86	**.19**
BNC Spoken - Textbook Corpus	11.80	58	.86	**.19**
BNC Written - BNC Spoken	1.70	58	.10	**.46**

Nonetheless, Table 4.20 reveals that there are significant differences in the mean frequency of the complex pattern *MVb + be + past pple* between the written English corpus and the textbook corpus (p = .016) and between the written English corpus and the spoken English one (p = .013). However, its mean in the textbook corpus is close to the one in spoken English (p = .44).

Table 4.20
***T*-Test Results for the Pattern MVb + be + past pple**

	x̄a - x̄b	*df*	*t*	*p* Value.
MVb + be + past pple				
BNC Written - Textbook Corpus	59.66	58	2.19	**.016***
BNC Spoken - Textbook Corpus	-1.16	58	- 0.15	.44
BNC Written - BNC Spoken	60.83	58	2.27	**.013***

***p* < .05**

Table 4.21 shows a similar trend in respect to the pattern *MVb + pers.pron.* Again the mean frequency of this pattern in the textbook corpus is close to the one in the spoken English corpus (p = .42). While its mean frequency in the spoken English corpus is significantly different from its frequency in the written English corpus (p = .03), it tends to be different in the written English corpus-textbook corpus comparison (p = .05).

Table 4.21
***T*-Test Results for the Pattern MVb + pers.pron**

	x̄a - x̄b	*df*	*t*	*p* Value.
MVb + pers.pron.				
BNC Written - Textbook Corpus	-46.66	58	-1.62	**.05****
BNC Spoken - Textbook Corpus	7.83	58	.20	.42
BNC Written - BNC Spoken	-54.50	58	-1.94	**.03***

***p* < .05 **p* < .10**

Table 4.22 shows an atypical pattern if compared with the previous four. The mean frequency of *MVb + have + past pple* in the textbook corpus is statistically different from its frequency in the written English corpus ($p = .04$) and in the spoken English corpus as well ($p = .02$). However, the mean of this pattern in spoken and in written English is not significantly different ($p = .29$).

Table 4.22
***T*-Test Results for the Pattern MVb + Have + past pple**

	x̄a - x̄b	*df*	*t*	*p* Value.
MVb + have + past pple				
BNC Written - Textbook Corpus	13.60	58	1.79	**.04***
BNC Spoken - Textbook Corpus	19.66	58	2.17	**.02***
BNC Written - BNC Spoken	-6.06	58	-.53	.29

***p* < .05**

The data regarding the frequencies of modal verb patterns show there are significant differences in the frequencies of the following patterns: *MVb + be + past pple*, *MVb + pers.pron* and *MVb + have + past pple*. These differences as well as the differences in the occurrence of modal verbs in the three corpora will be discussed in the following chapter. Nevertheless, as explained in the introduction to this chapter, since a lot of statistical information has been provided, a table that summarises all significant differences will be presented in Table 4.23 below.

Table 4.23
Summary of All Significant Differences

Correlation of the Affirmative Modal Verb Forms		
	Textbook Corpus	*p* Value
BNC Spoken	.54	> .05
Correlation of the Contracted Modal Verb Forms		
	Textbook Corpus	*p* Value
BNC Written	.39	> .05
BNC Spoken	.27	> .05
***T*-Test Results for the Frequencies of the Contracted Modal Verb Forms**		
	$\bar{x}a - \bar{x}b$	*p* Value
BNCW - BNCS	-554.63	.03
BNCS - Textbooks	566.72	.02
***T*-Test Results for the Pattern MVb + be + past pple**		
	$\bar{x}a - \bar{x}b$	*p* Value
BNCW - Textbooks	59.66	.016
BNCW - BNCS	60.83	.013
***T*-Test Results for the Pattern MVb + pers.pron**		
	$\bar{x}a - \bar{x}b$	*p* Value
BNCW - Textbooks	-46.66	.05
BNCW - BNCS	-54.50	.03
***T*-Test Results for the Pattern MVb + Have + past pple**		
	$\bar{x}a - \bar{x}b$	*p* Value
BNCW - Textbooks	13.60	.04
BNCS - Textbooks	19.66	.02

CHAPTER V

DISCUSSION

In Chapter 4, the occurrences of all the modal verb forms per million words in the textbook corpus, in the spoken English corpus and in the written English corpus were calculated. This data was used to analyse the differences in the frequencies of all the modal verb forms together and of the affirmative, negative, present, past, contracted and full modal verb forms separately across the three corpora. This was done by calculating the corresponding *t*-test results and correlations.

Secondly, those modal verb patterns whose frequency was higher than one per cent in the textbook corpus were presented. The frequencies of these patterns in the three corpora were compared by calculating *t*-test results.

Following the structure of the previous chapter, the frequencies of the modal verb forms targeted across the three corpora and the learning implications of these frequencies will be discussed first (Objective 1). This will be followed by the discussion of the frequencies of the verb structures in which these modal verbs occur and of their effect on ESL learning (Objective 2).

5.1 Discussion of Modal Verb Frequencies

Modal verb forms in the textbook corpus and in the written corpus correlate significantly. Table 4.2 showed an almost absolute positive relationship between the distribution of these forms in both corpora ($r = .93$, $p < .0001$). Besides, the paired sample *t*-test revealed no

significant differences in the mean frequency of all the modal verb forms in written English and in the textbook corpus (p = .47), (see Table 4.3). In addition, these tests confirmed no significant differences when the affirmative, negative, present, past, full and contracted forms were analysed separately, except for the correlation of the contracted modal verb forms, which was not significant (r = .39, p > .05), (see Table 4.12).

These findings would lead us to think that the selected ESL textbooks prime the learner to use modal verbs by means of frequencies similar to the ones in written English. However, there is a surprising finding. Differences in the correlations and in the mean frequencies of modal verb forms between the spoken English corpus and the textbook corpus tend not to be significant either. However, there are more differences in the textbook corpus-spoken English corpus comparison than in the textbook corpus-written English one.

The first difference is the correlation of the affirmative modal verb forms (r = .54, p > .05) (see Table 4.4). The second difference is the correlation of the contracted modal verb forms, as it happened in the written English-textbook corpus pair (r = .27, p > .05), (see Table 4.12). What is more, the mean frequencies of the contracted forms in the ESL corpus-spoken English comparison are significantly different (p = .02) (see Table 4.13). This indicates that a discussion of the differences in the frequencies of modal verb forms between the spoken English corpus and the textbook corpus should be oriented towards the affirmative modal verb forms and the contracted modal verb forms in particular.

Nonetheless, the research problem already discussed in Chapter 1 raised a number of questions with pedagogic implications, all of them listed in the section *Research Questions*. The first set of questions referred to modal verb frequencies. Do the selected ESL textbooks prime the ESL learner to use all the modal verbs included in this study? The second and third questions addressed the frequencies of these verbs and any possible emphasis on any modal verb form or lemma. The fourth question referred to the comparison of modal verb frequencies in the textbook corpus, in the written English corpus and in the spoken English one. Next, the frequencies of the past and of the present modal verb forms across the three corpora were approached. Are the frequencies of the past modal verb forms in the textbook

corpus lower than in spoken English and than in written English? These questions need to be answered and it is for this reason that the data in Table 4.1 will be discussed further to go deeper into the analyses. It needs to be stated that these questions will not be answered one by one. They will be approached in an integrative manner by comparing and contrasting the frequencies of the contracted modal verb forms and then the frequencies of the past and of the present modal verb forms in the corpus of spoken English and in the corpus of written English with their frequencies in the textbook corpus. This will result in further explorations based on the findings, as explained in the *Methodology* section.

5.1.1 Discussion of the Frequencies of Contracted Modal Verbs

In the *Theoretical Background*, several differences between spoken and written English were discussed. From the frequency lists prepared by Leech, Rayson and Wilson (2001) and based on the 100,000,000 word BNC, it could be established that there were 19,543 modal verb tokens per million words in the spoken English corpus, while the written English corpus contained 13,635 tokens. In Table 5.1, the frequencies of the affirmative and of the negative modal verb forms per million words in the three corpora of the present study, i.e. the BNCW, the BNCS and the textbook corpus, are shown.

Table 5.1
Frequencies of Affirmative and Negative Modal Verb Forms per Million Words in the Three Corpora

	BNC Written	BNC Spoken	Textbook Corpus
All Modal Verbs	**14,032**	19,025	**14,439**
Affirmative Foms	**12,682**	15,898	**12,857**
Negative Forms	**1,350**	3,127	**1,582**

As discussed in *The Research Problem*, one of the three important factors that differentiate L1 learning from L2 learning is exposure to a wider range of lexical items, words and clauses. Table 5.1 shows that the students who use the ESL textbooks that compose the textbook corpus as the sole or main source of priming will be exposed to a

lower frequency of modal verbs (n = 14,439) than those students who are educated in an oral environment (n = 19,025). In fact, the number of modal verbs in the textbook corpus is close to the one in the written English corpus (n = 14,032). This difference between spoken English and textbook English may be compensated by supplementing ESL textbooks with materials and activities that contain samples of spoken English, i.e. with teachers' lesson planning and intervention, particularly in the area of fast speech, where contracted forms are used instead. Indeed, a second difference between spoken and written English was contractions.

Contracted modal verb forms are more frequent in spoken English. As explained in the section *ESL Textbooks as Priming Agents of Written and Spoken English: Modal Verbs*, the frequencies of *'d* and of *'ll* are higher in spoken English than in written English. Indeed, Table 4.13 showed significant differences in the mean frequencies of the contracted modal verb forms. Table 5.2 shows the frequencies of the forms *'d* and *'ll* in the three corpora of the present study contrasted with the frequencies of the corresponding full forms *would* and *will*.

Table 5.2
Frequencies of *'ll*, *'d*, *would* and *will* per Million Words in the Three Corpora

Modal Verbs	BNC Written	BNC Spoken	Textbook Corpus
'll	393	3,111	74
'd	165	1,177	11
Total 1	**558**	**4,288**	**85**
Will	2,878	1,735	4,236
Would	2,185	2,324	1,620
Total 2	**5,063**	**4,059**	**5,856**
Total 1 + 2	**5,621**	**8,347**	**5,941**

The figures in Table 5.2 clearly illustrate the extent to which the frequencies of the contracted affirmative modal verb forms in the textbook corpus are lower than in spoken English and in written English. Indeed, while their frequency in spoken English (n = 4,288) is about 50 times higher than in the textbook corpus (n = 85), their frequency in written English (n = 558) is almost 7 times higher. This indicates that the ESL students who use only the textbooks that compose the textbook corpus as a source of priming will be exposed to a much lower number of *'d* and of *'ll* forms than students educated in an L1 environment. Since it is not a slight difference that is being discussed, its pedagogical implications will be critical for two reasons.

First, since the textbook corpus consists of four textbooks, it would come as no surprise that the 11 occurrences of *'d* or the 74 occurrences of *'ll* were concentrated in one, two or three of these textbooks, which means some ESL learners will not come across either or both of these contracted modal verb forms in writing at all. Secondly, the lists by Leech, Rayson and Wilson (2001) showed that the contraction of both affirmative modal verb forms is a distinct element of spoken English. The learners who use these textbooks are in cycle two of secondary school and are supposed to be priming words that they could use in spoken and in written contexts as established by the *English as a Second Language: Core Program, Enriched Program* produced by the MELSQ. If, as discussed in Chapter 1, to sound native-like the ESL learner needs to use the language that native speakers actually use, the ESL textbooks of secondary school cycle two approved in 2007 in Québec are far from presenting the language the learner needs to this end as far as the modal verb forms *'ll* and *'d* are concerned.

By contrast, the textbook corpus contains a high number of the form *will* (n = 4,236) compared to written English (n = 2,878) and to spoken English (n = 1,735). However, the forms *will* and *'ll* together in the textbook corpus add up to a total figure of 4,310 (74 + 4,236) tokens, while in spoken English they amount to 4,846 (3,111 + 1,735). This means that the lemma *will* is not absent from the textbook corpus, but that the ESL textbooks that compose it prime students to use the full form, which is more frequent in written English.

Table 5.3 below supplies more data related to contractions, this time the contraction of the negative modal verb forms.

Table 5.3
Frequencies of the Contracted Negative Modal Verb Forms per Million Words in the Three Corpora

Modal Verbs	BNC Written	BNC Spoken	Textbook Corpus
Can't	200	**1,216**	354
Couldn't	107	**401**	137
Shan't	3	**17**	0
Shouldn't	17	**142**	47
Mustn't	7	21	**47**
Won't	92	**557**	157
Wouldn't	69	**503**	94
Daren't	0	1	0
Needn't	1	9	0
Total	**496**	**2,867**	**836**

Looking at the totals, the reader will see that, as it happens with the contracted affirmative modal verb forms, the frequency of the contracted negative forms is much higher in the spoken English corpus (n = 2,867) than in the textbook corpus (n = 836) and in the written English corpus (n = 496). A closer look into the frequencies of each modal verb form will help to establish further differences.

The ESL learner that uses any of the books that compose the textbook corpus will not come across any occurrence of the form *shan't*, very rare in written English but still present in spoken English. The same applies to the forms *needn't* and *daren't*, which appear in spoken English and/or in written English, but whose frequencies are still very low. It is interesting to note that the frequencies of most of the other contracted forms (*can't*, *couldn't*, *shouldn't*, *won't*, *wouldn't*) are higher in spoken English than in the textbook corpus and in written English. It is also noteworthy that the frequencies of *can't*, *couldn't*, *shouldn't*, *won't*, *wouldn't,* and *mustn't* are higher in the textbook corpus than in written English. The reason why the frequency of *mustn't* is higher in the textbook corpus than in written English and in spoken English may be due to the prescriptive nature of the linguistic and cultural content of textbooks. Indeed, textbooks are key agents in adolescents' upbringing and they are expected to teach them what they must not do to become responsible citizens and good language users.

The frequencies of the full negative forms will be presented below in Table 5.4 in order to contrast them with the frequencies of the contracted negative forms.

Table 5.4
Frequencies of the Full Negative Modal Verb Forms per Million Words in the Three Corpora

Modal Verbs	BNC Written	BNC Spoken	Textbook Corpus
Cannot	**258**	81	**192**
Could not	**135**	17	**74**
May not	49	49	**70**
Might not	**16**	**20**	**19**
Shall not	**9**	2	**4**
Should not	**68**	23	**62**
Must not	35	5	**86**
Will not	**134**	33	**157**
Would not	**131**	30	**78**
Dare not	**2**	0	**4**
Need not	**17**	0	0
Total	**854**	**260**	**746**

The total frequency of the full negative modal verb forms in the textbook corpus is higher (n = 746) than in spoken English (n = 260) but close to their frequency in written English (n = 854). Table 5.4 also shows that the frequency of each of the full negative modal verb forms in the textbook corpus is higher than in the spoken English corpus except for *might not*, where the frequency in both corpora is almost the same (19 and 20), and for *need not*, which is absent from both corpora. It is also important to point out the high frequency of *must not* in the textbook corpus (n = 86) compared with its frequency in the written English corpus (n = 35), perhaps as further evidence of the prescriptive language of textbooks as alluded to earlier.

Two important differences in modal verb frequencies between spoken English and written English have been discussed in this section. The first one is the total frequency of all the modal verb forms, which is higher in spoken English than in written English and in the textbook corpus (see Table 5.1). The second difference is the frequencies of the contracted modal verb forms, which are also higher in spoken English than in written English and in the textbook corpus (except for *mustn't*, whose frequency is higher in the textbook corpus). A third difference, which was also discussed in the *Theoretical Background*, still needs to be addressed: some modal verb forms are more frequent in spoken English than in written English. This third distinction will be analysed in the next section, where the frequencies of individual past and present modal verb forms are discussed.

5.1.2 Discussion of the Frequencies of Present and Past Modal Verbs.

Studies showing that ESL learners use the present forms *can* and *will* in situations in which native speakers normally use the forms *would*, *could*, *may* or *might* have been presented (Montero, Watts and Garcia-Carbonell, 2007; Debbie, 2009). These studies apply to specific contexts, nationalities and cultures. Their results cannot be generalised, but they have brought up the issue of tense in modal verb frequencies. Data from the 100,000,000 British National Corpus (Leech, Rayson and Wilson, 2001) discussed in the *Theoretical Background* section showed that, proportionally, the frequencies of past modal verbs in written English and in spoken English are very similar: 44.5 % in the former and 43 % in the latter (see *ESL Textbooks as Priming Agents of Written and Spoken English: Modal Verbs*). Table 5.5 below provides the percentage of past and present modal verbs in the three corpora.

Table 5.5
Percentage of Present and Past Modal Verbs out of the total number of modal verbs in the Three Corpora

	BNC Written	BNC Spoken	Textbook Corpus
Present Modal Verbs	57.50	60	68.50
Past Modal Verbs	42.50	40	31.50
Total	**100**	**100**	**100**

Table 5.5 shows that native speakers do not use past modal verbs more often than present modal verbs. Nonetheless, what can be observed is that the percentage of past modal verbs in the textbook corpus is lower than in written English and in spoken English. This may be so because, as discussed above, there are 11 occurrences of *'d* in the textbook corpus, which might have an effect on these results. But besides *'d*, the present study looks into the comparison of the frequencies of the other past modal verb forms. Table 5.6 shows the frequencies of all the past modal verb forms per million words.

Table 5.6
Frequencies of the Modal Verbs *Could, Should, Might* and *Would* per Million Words in the Three Corpora

Modal Verbs	BNC Written	BNC Spoken	Textbook Corpus
Could	1,341	1,405	1,089
Could not	135	17	74
Couldn't	107	401	137
Total 1	**1,583**	**1,823**	**1,300**
Might	443	642	444
Might not	16	20	19
Total 2	**459**	**662**	**463**
Should	1,286	976	869
Should not	68	23	62
Shouldn't	17	142	47
Total 3	**1,371**	**1,141**	**978**
Would	2,185	2,324	1,620
'd	165	1,177	11
Would not	131	30	78
Wouldn't	69	503	94
Total 4	**2,550**	**4,034**	**1,803**
Total 1 + 2 + 3 + 4	**5,963**	**7,660**	**4,544**

A closer look at all the past modal verb forms appearing in the textbook corpus as well as at their frequencies in written English and in spoken English shows that their frequencies in the textbook corpus are not as low as one may conclude regarding the studies cited above, which show that the ESL learner is primed mainly to use present modal verb forms. Nevertheless, Table 5.6 shows that the frequency of the lemma *would* is much higher in the spoken English corpus than in the textbook corpus and in the written English corpus (BNCW n = 2,550; BNCS n = 4,034; Textbook Corpus n = 1,803). The reason for this is that, as discussed, the frequency of *'d* is noticeably low in the textbook corpus and in written English

compared with spoken English. This means that the ESL learner will be primed to use past modal verbs (except for the *'d* form as already discussed) by the textbooks that make up the textbook corpus. Table 5.7 will point out any possible differences in connection with the frequencies of the present modal verb forms.

Table 5.7
Frequencies of the Modal Verbs *Can, May, Shall, Will, Dare, Must* and *Need* per Million Words in the Three Corpora

Modal Verbs	BNC Written	BNC Spoken	Textbook Corpus
Can	2,004	3,210	3,324
Cannot	258	81	192
Can't	200	1,216	354
Total 1	**2,462**	**4,507**	**3,870**
May	973	407	696
May not	49	49	70
Total 2	**1,022**	**456**	**766**
Shall	211	285	15
Shall not	9	2	4
Shan't	3	17	0
Total 3	**223**	**304**	**19**
Must	803	619	475
Must not	35	5	86
Mustn't	7	21	47
Total 4	**845**	**645**	**608**
Will	2,878	1,735	4,236
'll	393	3,111	74
Will not	134	33	157
Won't	92	557	157
Total 5	**3,497**	**5,436**	**4,624**

Modal Verbs	BNC Written	BNC Spoken	Textbook Corpus
Dare	0	6	4
Dare not	2	0	4
Daren't	0	1	0
Total 6	**2**	**7**	**8**
Need	0	1	0
Need not	17	0	0
Needn't	1	9	0
Total 7	**18**	**10**	**0**
Total 1 to 7	**8,069**	**11,365**	**9,895**

The occurrences of the forms of the lemmas *dare* and *need* are very rare or inexistent in the three corpora. Indeed, students will receive no samples of the lemma *need*, and eight occurrences of the lemma *dare* per million words, but, as with *'d*, it might happen that they all occur in only one or a few of the textbooks that form the textbook corpus and/or that they are used in sentences that illustrate grammar points. Students will also receive very little priming of the lemma *shall*: 19 occurrences compared to 304 in spoken English and to 223 in written English. This does not come as a surprise considering that, as explained in *Limitations of the study* in Chapter 1, *shall* is more frequently used in BrE than in AmE and that the textbooks contained in the textbook corpus have been made in North America. However, an important finding, which is closely related to the studies by Montero, Watts and Garcia-Carbonell (2007) and by Debbie (2009), is that the two most frequent modal verb forms in the textbook corpus are *can* and *will*. This is shown in Table 5.8 below.

Table 5.8
***Can* and *Will* as a Percentage of the Total Number of Modal Verbs in the Three Corpora**

Modal Verbs	BNC Written	BNC Spoken	Textbook Corpus
Can	14	17	23
Will	20.50	9	29
Total	**34.50**	**26**	**52**

Did the ESL speakers who participated in the studies conducted by Montero, Watts and Garcia-Carbonell (2007) and by Debbie (2009) use *will* and *can* instead of *would*, *might*, *could* or *may* because they were the modal verbs to which they had been exposed the most in their ESL learning process? It is not possible to know, but Table 5.8 shows that the frequencies of the modal verb forms *can* and *will* account for more than half of the modal verb frequencies in the textbook corpus. In contrast, in the spoken English corpus, these two modal verb forms account for about a quarter of its modal verb frequencies. In Table 5.9, the percentage of the frequencies of the forms *would* and *could* are presented.

Table 5.9
***Could* and *Would* as a Percentage of the Total Number of Modal Verbs in the Three Corpora**

Modal Verbs	BNC Written	BNC Spoken	Textbook Corpus
Could	9.50	7.50	7.50
Would	15.50	12	11
Total	**25**	**19.50**	**18.50**

Tables 5.8 and 5.9 show that *can*, *could*, *will* and *would* are more evenly distributed in the written English corpus and in the spoken English corpus than in the textbook corpus. The ratio of the percentage of occurrence of *would* and *could* to *can* and *will* in the textbook corpus is 1:2.81, while in the BNCW it is 1:1.38 and in the BNCS 1:1.33. However, compare these data with the addition of all the word forms of the lemmas *will* and *would* in Table 5.10 below.

Table 5.10
All Forms of the Lemmas *Can, Will, Could* and *Would* as a Percentage of the Total Number of Modal Verbs in the Three Corpora

Modal Verbs	BNC Written	BNC Spoken	Textbook Corpus
Can + Can't + Cannot	17.50	24	27
Will + 'll + Won't + Will not	25	28.50	32
Total	**42.50**	**52.50**	**59**
Could + Couldn't + Could not	11	9.50	9
Would + 'd + Wouldn't + Would not	18	21	12.50
Total	**29**	**30.50**	**21.50**

Now that all the word forms of the lemmas *can*, *will*, *could* and *would* have been added, the ratio for written English remains almost the same (BNCW 1:1.46), in spoken English the percentage of present forms tends to be higher compared with the percentage of past forms (BNCS 1:1.72), and in the textbook corpus there is still strong priming of the present forms compared with that of the past forms (Textbook corpus 1:2.74).

To summarise the discussions of the present section of this study, the occurrence of modal verbs, it is possible to say that, except for the correlation of the contracted modal verb forms, the statistical analyses show that there are no significant differences in modal verb frequencies between the textbook corpus and the written English corpus. In contrast, there are more significant differences between the textbook corpus and the spoken English corpus,

specifically in the correlation of the affirmative modal verb forms, and in the correlation and mean frequency of the contracted modal verb forms.

However, the detailed analyses of the frequencies of the contracted and of the full past and present modal verb forms presented in Table 4.1 have revealed some interesting findings. Firstly, the analysed ESL textbooks tend not to prime the learner to use the forms *'d* and *'ll*, which are really necessary for authentic spoken communication. Secondly, the total frequencies of the contracted and of the full negative forms in the textbook corpus are closer to the ones in the written English corpus. Thirdly, there are certain forms that the ESL textbooks that compose the textbook corpus will not prime the learner to use. These forms are *shan't*, *daren't*, *need*, *needn't*, *need not*. Fourthly, the ESL textbooks that make up the textbook corpus will prime the learner to use all past modal verb lemmas. However, the percentage of past modal verb forms is lower in the textbook corpus than in the spoken and in the written English corpus. Finally, an important finding is that, in the textbook corpus, the frequency of *could* and *would* is much lower than the frequency of *can* and *will*. This difference is more noticeable than in spoken and in written English, where the frequency of *can* and *will* is also higher than the frequency of *could* and *would*, but where the difference between them is smaller. A parallel was drawn between these last findings and the studies conducted by Montero, Watts and Garcia-Carbonell (2007) and by Debbie (2009), which indicate that *will* and *can* are the most frequently used modal verbs amongst ESL learners because they prefer using them in contexts where native speakers would normally use *would* and *could*.

5.2 Discussion of Modal Verb Pattern Differences in the Three Corpora.

The second set of questions dealt with modal verb patterns. The first question asked about the frequencies of these patterns in the textbook corpus. The second question was if these frequencies were similar to the ones in spoken English or in written English. The third one approached the issue of whether modal verb patterns in the textbook corpus tend to be simple as they are supposed to be in spoken English or whether they are more complex as they are expected to be in written English. Finally, the possible implications of these answers in ESL

learning were also addressed. To conclude, Willis' (1990) statement that there is an attempt to teach ESL learners to speak written language was challenged.

As explained in the *Methodology* section, the frequencies of the following patterns will be discussed: *MVb + inf*, *MVb + be + past pple*, *MVb + pers.pron*, *MVb + adv + inf*, and *MVb + have + past pple*. Each pattern will be approached separately in order to find out whether the frequencies of the modal verb patterns that the ESL textbooks that compose the textbook corpus prime the ESL learner to use are similar to their frequencies in spoken English or in written English. To this end the findings presented in Chapter 4 will be discussed first. They will then be supplemented with the discussion of the data included in the tables of Appendixes A to J.

The infinitive following the modal verb is the most frequent pattern in the three corpora. In the textbook corpus, 65.3 % of all modal verb tokens are followed by an infinitive, 59 % in the BNCW and 63.7 % in the BNCS (see Table 4.17). The difference in the mean frequencies of this pattern across the three corpora is not statistically significant (see Table 4.18). Even though the frequency of the pattern *MVb + adv + inf* is, on average, about ten times lower than that of the pattern *MVb + inf*, it is also very similar in the three corpora: 4.7 % of all modal verb tokens in the textbook corpus, 7.7 % in the BNCW and 5.4 % in the BNCS (see Table 4.17). The difference in the mean frequencies of this pattern across the three corpora is not statistically significant either (see Table 4.19).

Statistical results indicate that it is in complex patterns that differences between the textbook corpus, the written English corpus and the spoken English one are significant. In this regard, the pattern *MVb + be + past pple* is much more frequent in written English than in the textbook corpus and in spoken English: 3.6 % of all modal verb tokens in the textbook corpus belong to this pattern, 16.4 % in the BNCW and 2.6 % in the BNCS (see Table 4.17). As can be seen in Table 4.16, the frequency of this pattern in the textbook corpus is close to the one in spoken English. Table 4.20 shows that the textbook corpus pairs with the BNCS. Indeed, while the mean frequency of the pattern *MVb + be + past pple* in the BNCS and in the textbook corpus is statistically different from its frequency in the BNCW, there are no

significant differences between the textbook corpus and the BNCS. An inverse distribution can be observed in the interrogative pattern *MVb + pers.pron.*

Table 4.17 shows that the frequency of *MVb + pers.pron* is much lower in written English: 12.5 % of all modal verb tokens in the textbook corpus occur in this pattern, 2.9 % in the written English corpus and 10.7 % in the spoken English corpus. In addition, the *t*-test results indicate that, while the mean frequency of the pattern *MVb + pers.pron* in the BNCS is statistically different from the one in the written English corpus and tends to be different in the textbook corpus-written English corpus comparison (see Table 4.21), there are no significant differences between the textbook corpus and the corpus of spoken English.

Finally, the pattern *MVb + have + past pple*, whose frequency is atypical because it is lower in the textbook corpus than in the spoken English corpus and in the written English one: 1.2 % of all modal verb tokens in the textbook corpus are in this pattern, 4.1 % in written English and 3.9 % in spoken English (see Table 4.17). Table 4.22 reveals that, while the mean frequency of the pattern *MVb + have + past pple* in the textbook corpus is statistically different from its mean frequency in the spoken English corpus and in the written English corpus, there are no significant differences between the last two corpora.

This shows that, as far as the statistical information on the frequencies of modal verb patterns is concerned, the ESL textbooks selected for study do not prime the learner to use modal verb patterns by means of frequencies similar to the ones in written English. First of all, the mean frequency of the pattern *MVb + be + past pple* in the textbook corpus is significantly different from its mean frequency in the written English corpus. Secondly, the mean frequency of the pattern *MVb + pers.pron* in the same two corpora tends to be significantly different. As regards the mean frequency of the pattern *MVb + have + past pple* in the textbook corpus, it is statistically different from its frequency in the spoken English corpus and in the written English one. These findings indicate that the ESL textbooks that make up the textbook corpus prime the ESL learner to use, mainly, simple modal verb patterns.

As discussed in the *Theoretical Background*, complexity is a feature of written English. However, even though the ESL textbooks that compose the textbook corpus contain written text, they expose the ESL learner to a low frequency of the complex patterns *MVb + be + past pple* and *MVb + have + past pple*. Indeed, Table 4.17 indicates that these two patterns account for the following percentages in each of the three corpora: BNCW $\sum$ = 20.5 % (16.4 % + 4.1 %), BNCS $\sum$ = 6.5 % (2.6 % + 3.9%), Textbook Corpus $\sum$ = 4.8 % (3.6 % + 1.2 %). These data show that written English has the highest frequency of the modal verb patterns *MVb + be + past pple* and *MVb + have + past pple*, whereas the textbook corpus has the lowest frequency owing to the fact that it contains fewer samples of the pattern *MVb + have + past pple* than the spoken English corpus. It all seems to indicate that the frequencies of the five modal verb patterns selected for study in the textbook corpus are similar to their frequencies in the spoken English corpus except for the pattern *MVb + have + past pple*. The following sections will shed more light on the frequencies of modal verb patterns by discussing the figures taken from the tables in Appendixes A to J.

5.2.1 The Pattern MVb + Inf

The data in Appendixes A to J reveal that *MVb + Inf* is always the most frequent pattern that the ESL textbooks that make up the textbook corpus will prime the ESL learner to use with the affirmative modal verb forms, except for *dare*, which never occurs in this pattern as it never does in the spoken English corpus or in the written English corpus either (see Table 5.11). However, there are some differences between the textbook corpus, the written English corpus and the spoken English corpus that need to be discussed.

Table 5.11
Frequencies of the Pattern MVb + Inf per Million Words in the Three Corpora

Modal Verb + inf	BNC Written	BNC Spoken	Textbook Corpus
Can + inf	**925**	1,845	**1,986**
Cannot + inf	163	61	**126**
Can't + inf	156	865	**283**
Could + inf	745	783	625
Could not + inf	98	16	**60**
Couldn't + inf	93	272	**114**
May + inf	551	224	432
May not + inf	38	32	63
Might + inf	230	401	310
Might not + inf	8	11	12
Shall + inf	**129**	**119**	**11**
Shall not + inf	6	2	**4**
Shan't + inf	1	11	**0**
Dare + inf	**0**	**0**	**0**
Dare not + inf	**0**	**0**	**0**
Daren't + inf	**0**	**0**	**0**
Should + inf	511	568	**480**
Should not + inf	27	8	**39**
Shouldn't + inf	6	76	**20**
Must + Inf	476	384	346
Must not + inf	25	4	**47**
Mustn't + inf	7	9	**31**

Modal Verb + inf	BNC Written	BNC Spoken	Textbook Corpus
Will + inf	**1,956**	**1,083**	**3,209**
'll + inf	**337**	**2,496**	**63**
Will not + inf	104	26	**114**
Won't + inf	77	373	**122**
Would + inf	**1,369**	**1,293**	**810**
'd + inf	**123**	**831**	**4**
Would not + inf	79	16	**55**
Wouldn't + inf	49	314	**59**
Total	**8,289**	**12,123**	**9,425**

First, in section 5.1.2, the fact that *can* and *will* account for 52 % of all the modal verb occurrences in the textbook corpus was reported. In fact, the patterns *can* + *inf* and *will* + *inf* (see Table 5.11) are the biggest primings the ESL learner will receive from the ESL textbooks that belong to the textbook corpus as far as modal verb patterns are concerned. The frequencies of these two patterns are very high if compared with their frequencies in written English (as is the case of *can* + *inf* and *will* + *inf*) and in spoken English (as is the case of *will* + *inf* only). Indeed, the distribution of the patterns containing *can* shows that the ESL textbooks that compose the textbook corpus prime the learner to use this modal verb form in patterns whose frequencies are closer to the ones in the spoken English corpus than in the written English corpus (see Table 5.12 and Appendix A).

Table 5.12
Frequencies of the Patterns Containing the Modal Verb Form *Can* per Million Words in the Three Corpora

	BNC Written	BNC Spoken	Textbook Corpus
Can			
MVb + inf	925	**1,845**	**1,986**
MVb + have + past pple	0	**0**	**0**
MVb + be + past pple	516	**63**	**142**
MVb + pers.pron.	96	**769**	**630**
MVb + adv + inf	275	**187**	**216**
Total	**1,812**	**2,864**	**2,974**

In contrast, the frequency of *would + inf* in the textbook corpus is lower than in the written English corpus and than in the spoken English one. In addition, the frequencies of *'d + inf*, and *'ll + inf* in the textbook corpus are very low (see Table 5.11). Sixty-three occurrences of the pattern *'ll + inf* have been reported in the textbook corpus compared with 2,496 in spoken English, and four occurrences of *'d + inf* compared with 831 in spoken English as well (see Table 5.11).

With regard to the negative modal verb forms, *MVb + inf* is always the most frequent pattern in the textbook corpus (see Appendixes A to J) except for *shan't + inf*, which never occurs in it, and for *dare not + inf* and *daren't + inf* which, as *dare + inf*, are not present in any of the three corpora (see Table 5.11). Finally, there is an interesting observation to make. In the textbook corpus, there are more frequencies of *can't + inf*, *couldn't + inf*, *won't + inf* and *wouldn't + inf* than of the same pattern containing the corresponding full forms. The opposite applies to the patterns *shan't + inf*, *shouldn't + inf* and *mustn't + inf*, whose frequencies are lower than *shall not + inf*, *should not + inf* and *must not + inf* probably with a view to creating a stronger effect on the reader(see Table 5.11). This also relates to the prescriptive nature of textbooks alluded to in the first two sections of this chapter.

5.2.2 The Pattern MVb + Pers.Pron

As discussed in section 5.2, the frequency of the pattern *MVb + pers.pron* in the textbook corpus is close to its frequency in the spoken English corpus and much higher than in written English. Indeed, Table 5.13 shows that this happens particularly with the frequencies of the following patterns: *can + pers.pron*, *could + pers.pron*, *may + pers.pron*, *will + pers.pron* and *would + pers.pron* . This could be due to the fact that textbook writers invite the ESL learner to do tasks or to engage in conversation by means of questions, or even to the presence of dialogues in textbooks to illustrate oral communication. At the same time, there are two more differences between the spoken English corpus and the textbook corpus worth commenting on.

Firstly, the number of occurrences of the pattern *shall + pers.pron* in the textbook corpus (n = 4) is low if compared with its 137 occurrences in spoken English and 18 in written English. Definitely, the ESL textbooks that form the textbook corpus do not prime the learner to use this modal verb in the two most frequent patterns in which it occurs in the spoken English corpus and in the written English one: *shall + inf* (as discussed above, see Table 5.11) and *shall + pers.pron*. The reason for this could be the fact mentioned by Mathews (2003) and cited in Chapter 1 that *shall* is used more frequently in BrE than in AmE.

Table 5.13
Frequencies of the Affirmative Modal Verb Forms Followed by a Personal Pronoun per Million Words in the Three Corpora

Modal Verb + pers.pron.	BNC Written	BNC Spoken	Textbook Corpus
Can + pers.pron.	96	**769**	**630**
Will + pers.pron.	45	**168**	**271**
Should + pers.pron.	44	56	87
Could + pers.pron.	50	**182**	**173**
Would + pers.pron.	75	**315**	**547**
Shall + pers.pron.	**18**	**137**	**4**
Dare + pers.pron.	0	6	4
Must + pers.pron.	11	6	0
May + pers.pron.	20	**45**	**59**
Might + pers.pron.	5	7	4
Total	**364**	**1,691**	**1,779**

Secondly, with respect to the use of the contracted negative modal verb forms in this pattern, the fact that *can't + pers.pron*, *won't + pers.pron*, *shouldn't + pers.pron*, *couldn't + pers.pron* and *mustn't + pers.pron* occur in the spoken English corpus while they are inexistent or very rare in the textbook corpus can be observed in Table 5.14. It may be noted that *Wouldn't + pers.pron* is the only pattern whose frequency is higher in the textbook corpus than in the spoken English corpus and in the written English one.

Table 5.14
Frequencies of the Contracted Negative Modal Verb Forms Followed by a Personal Pronoun per Million Words in the Three Corpora

Modal Verb + pers.pron.	BNC Written	BNC Spoken	Textbook Corpus
Can't + pers.pron.	19	**150**	8
Won't + pers.pron.	4	**79**	0
Shouldn't + pers.pron.	4	**18**	0
Couldn't + pers.pron.	2	**32**	4
Wouldn't + pers.pron.	9	2	**16**
Shan't + pers.pron.	0	0	0
Daren't + pers.pron.	0	0	0
Mustn't + pers.pron.	0	**9**	0
Total	**38**	**290**	**28**

This sub-section indicates that the four ESL textbooks that compose the textbook corpus will rarely prime the learner to ask questions using the pattern *contracted negative modal verb form + pers.pron*, which occurs in spoken English. Conversely, they will prime the learner to ask questions using the pattern *affirmative modal verb form + pers.pron*, particularly *can + pers.pron*, *could + pers.pron*, *may + pers.pron*, *will + pers.pron* and *would + pers.pron* as discussed above.

5.2.3 The Pattern MVb + Adv + Inf

Table 5.15 reports no big differences in the frequencies of *MVb + adv + inf* between the textbook corpus, the written English corpus and the spoken English one. It tends to occur with affirmative modal verb forms in the three corpora. In addition, wherever there are a few or no occurrences of *MVb + adv + inf* in the textbook corpus, there are also a few or no occurrences of the same pattern in the other two corpora.

However, an interesting finding is that *MVb + adv + inf* is the second most frequent pattern containing the forms *'d* and *'ll* in the textbook corpus, and both in the written English and in the spoken English ones. Nevertheless, this cannot be considered an important source of priming in the textbook corpus since figures are very low. This shows that native speakers are primed to use the infinitive or an adverb followed by an infinitive after the contracted affirmative modal verb forms in spoken English and in written English (see Table 5.16).

Table 5.15
Frequencies of the Pattern MVb + Adv + Inf per Million Words in the Three Corpora

Modal Verb + adv + inf	BNC Written	BNC Spoken	Textbook Corpus
Can + adv + inf	**275**	**187**	**216**
Cannot + adv + inf	7	2	4
Can't + adv + inf	3	60	24
Could + adv + inf	**117**	**99**	**63**
Could not + adv + inf	5	0	0
Couldn't + adv + inf	4	22	0
May + adv + inf	**77**	**26**	**39**
May not + adv + inf	2	1	4
Might + adv + inf	**36**	**59**	**12**
Might not + adv + inf	3	4	0
Shall + adv + inf	**19**	**8**	**0**
Shall not + adv + inf	2	0	0
Shan't + adv + inf	0	0	0
Dare + adv + inf	**0**	**0**	**0**
Dare not + adv + inf	0	0	0
Daren't + adv + inf	0	0	0
Should + adv + inf	**31**	**25**	**63**
Should not + adv + inf	5	0	0
Shouldn't + adv + inf	1	4	0
Must + adv + inf	**35**	**13**	**12**
Must not + adv + inf	0	0	4
Mustn't + adv + inf	0	0	0

Modal Verb + adv + inf	BNC Written	BNC Spoken	Textbook Corpus
Will + adv + inf	**237**	**87**	**173**
'll + adv + inf	30	**207**	8
Will not + adv + inf	6	0	4
Won't + adv + inf	2	7	0
Would + adv + inf	**147**	**86**	**40**
'd + adv + inf	20	**117**	7
Would not + adv + inf	11	0	0
Wouldn't + adv + inf	3	13	0
Total	**1,078**	**1,027**	**673**

Table 5.16
Patterns of the Contracted Affirmative Modal Verb Forms

	BNC Written	BNC Spoken	Textbook Corpus
'll + inf	**337**	**2,496**	**63**
'll + have + past pple	0	2	0
'll + be + past pple	2	22	0
'll + pers.pron.	1	44	0
'll + adv + inf	**30**	**207**	**8**
'd + inf	**123**	**831**	**4**
'd + have + past pple	6	75	0
'd + be + past pple	5	22	0
'd + pers.pron.	0	16	0
'd + adv + inf	**20**	**117**	**7**
Total	**524**	**3,832**	**82**

5.2.4 The Pattern MVb + be + Past pple

As discussed in section 5.2, *MVb + be + past pple* is a pattern that is typical of written English. Nevertheless, the statistical results presented in Chapter 4 reported it was not a frequent pattern in the textbook corpus. Table 5.17 reveals that its frequency is always higher in the written English corpus than in the textbook corpus except for the following patterns: *can't + be + past pple*, *shouldn't + be + past pple* and *will not + be + past pple*. Nonetheless, differences in the frequencies of these three patterns are so slight and figures so low that they do not compensate for the vast number of occurrences of this pattern containing other modal verb forms in the written English corpus. Therefore, it is not possible to speak of any strong priming of this pattern in the ESL textbooks that make up the textbook corpus.

Table 5.17
Frequencies of the Pattern MVb + Be + Past pple per Million Words in the Three Corpora

Modal Verb + be + past pple	BNC Written	BNC Spoken	Textbook Corpus
Can + be + past pple	**516**	63	142
Cannot + be + past pple	**64**	9	24
Can't + be + past pple	4	**13**	**8**
Could + be + past pple	**191**	54	8
Could not + be + past pple	**18**	0	8
Couldn't + be + past pple	1	**3**	0
May + be + past pple	**179**	11	28
May not + be + past pple	1	**2**	0
Might + be + past pple	**53**	20	16
Might not + be + past pple	0	0	0
Shall + be + past pple	**22**	3	0
Shall not + be + past pple	0	0	0
Shan't + be + past pple	0	0	0

Modal Verb + be + past pple	BNC Written	BNC Spoken	Textbook Corpus
Dare + be + past pple	0	0	0
Dare not + be + past pple	0	0	0
Daren't + be + past pple	0	0	0
Should + be + past pple	**511**	81	55
Should not + be + past pple	**30**	10	16
Shouldn't + be + past pple	1	**7**	**4**
Must + be + past pple	**158**	13	35
Must not + be + past pple	**9**	1	0
Mustn't + be + past pple	0	0	0
Will + be + past pple	**358**	76	146
'll + be + past pple	2	**22**	0
Will not + be + past pple	13	2	**16**
Won't + be + past pple	1	**6**	0
Would + be + past pple	**163**	55	12
'd + be + past pple	5	**22**	0
Would not + be + past pple	**12**	3	4
Wouldn't + be + past pple	0	**11**	0
Total	**2,312**	**487**	**522**

5.2.5 The Pattern MVb + Have + Past pple

Statistical evidence showed the pattern *MVb + have + past pple* is less frequent in the textbook corpus than in spoken and in written English. Table 5.18 indicates that native speakers are mostly primed to use this pattern with the modal verb forms *could*, *may*, *might*, *should*, *must*, *would* and *'d* in spoken English, whereas in written English they are primed to use it with the same modal verb forms except for *'d*. In fact, the most frequent pattern both in the written English corpus and in the spoken English corpus is *would + have + past pple*.

Table 5.18 also shows that, apart from *'d*, the frequencies of the pattern *MVb + have + past pple* containing the above mentioned modal verb forms are also the highest in the textbook corpus. The ESL learner, however, will receive very little priming to use this pattern: its occurrence with these verbs in the textbook corpus is not as frequent as in spoken or as in written English.

Table 5.18
Frequencies of the pattern MVb + Have + Past pple per Million Words in the Three Corpora

Modal Verb + have + past pple	BNC Written	BNC Spoken	Textbook Corpus
Can + have + past pple	0	0	0
Cannot + have + past pple	0	0	0
Can't + have + past pple	1	1	0
Could + have + past pple	**59**	**77**	**43**
Could not + have + past pple	7	1	0
Couldn't + have + past pple	7	13	8
May + have + past pple	**60**	**29**	**20**
May not + have + past pple	6	5	0
Might + have + past pple	**70**	**56**	**12**
Might not + have + past pple	3	5	0
Shall + have + past pple	1	1	0
Shall not + have + past pple	0	0	0
Shan't + have + past pple	1	0	0
Dare + have + past pple	0	0	0
Dare not + have + past pple	2	0	4
Daren't + have + past pple	0	1	0
Should + have + past pple	**42**	**99**	**20**
Should not + have + past pple	2	1	0
Shouldn't + have + past pple	2	12	4

Modal Verb + have + past pple	BNC Written	BNC Spoken	Textbook Corpus
Must + have + past pple	**65**	**116**	**20**
Must not + have + past pple	1	0	0
Mustn't + have + past pple	0	0	0
Will + have + past pple	21	10	8
'll + have + past pple	0	2	0
Will not + have + past pple	2	0	0
Won't + have + past pple	1	2	0
Would + have + past pple	**198**	**218**	**24**
'd + have + past pple	6	**75**	0
Would not + have + past pple	12	6	4
Wouldn't + have + past pple	6	27	0
Total	**575**	**757**	**167**

Following the presentation and the discussion of the results concerning the frequencies of modal verbs and of modal verb patterns in the textbook corpus, in the written English corpus and in the spoken English corpus, conclusions referring to the objectives formulated in this study will be drawn. Conclusions about modal verb frequencies in the textbook corpus will be presented first, followed by those commenting on the frequencies of modal verb patterns. At the end, future research possibilities will be suggested.

5.3 Conclusions: Objective 1. The Frequency of Occurrence of Modal Verbs

The first research question posed in the *Research Questions* section in Chapter 1 regarding modal verb frequencies was whether all the modal verbs chosen for analysis in the present study occur in the textbook corpus. The results reported in this study showed that the selected ESL textbooks will not prime their readers to use *shan't*, *daren't*, *need*, *needn't* or

need not. The presence of all the other modal verb forms has allowed the comparison of modal verb frequencies in the three corpora.

The next two questions asked whether the ESL textbooks that belong to the textbook corpus prioritise certain modal verbs, i.e. whether they prime some modal verbs more frequently than others, and whether modal verb frequencies in the textbook corpus are close to those in the written English corpus or in the spoken English one. The study results indicated that learners in oral English environments would be exposed to a higher frequency of modal verbs than those who use the ESL textbooks under consideration as the sole or main source of learning. A link could be established between this finding and the study by Mason (1994) (see Chapter 2), which shows that native speakers produce more modal verbs in narrations than L2 ones. A reason for this higher production of modal verbs amongst L1 speakers could be the higher frequency of modal verbs to which they are exposed in oral environments compared to L2 speakers in an L2 context. The difference in modal verb frequencies between the textbook corpus and spoken English lies mainly in a much higher number of contracted modal verb forms in the latter. The comparison of modal verb frequencies in the textbook corpus and in spoken and in written English revealed that there are two specific forms which are abundant in spoken English and which ESL learners will barely be primed to use by the textbooks that make up the textbook corpus: *'d* and *'ll*. The ESL learner will be primed to use the full forms *will* and *would* instead. However, while the frequency of *would* in the textbook corpus is lower than in written and in spoken English, *will* and *can* are the two modal verb forms that the ESL textbooks approved by the MELSQ in 2007 and meant to be used by secondary school cycle two students as from 2008 will prime the ESL learner to use the most. This coincides with the findings of the studies by Watts and Garcia-Carbonell (2007) and by Debbie (2009), which indicate that ESL learners prefer using *can* and *will* to *could* and *would*. As regards the frequencies of both the contracted and the full negative modal verb forms in the textbook corpus, the study shows they are closer to the ones in the written English corpus.

The following question was whether the frequencies of past modal verb forms in the textbook corpus were different from their frequencies in spoken English and in written English. The study showed that, even though the total frequency of past modal verb forms in the textbook corpus was lower than in the corpus of written English and in the corpus of spoken English, statistical results indicated that these differences were not significant. Nevertheless, the percentage of occurrence of the two most frequent modal verb lemmas *will* and *can* out of the total number of modal verbs was compared with the percentage of occurrence of the modal verb lemmas *could* and *would*. The ratio of the percentage of all the forms of the lemmas *could* and *would* to all the forms of the lemmas *can* and *will* in written English was 1:1.46, in spoken English 1:1.72, and in the textbook corpus 1: 2.74. This shows that the percentage of occurrence of *can* and *will* out of the total number of modal verbs is much higher than that of *could* and *would* in the textbook corpus than in written and in spoken English. Cultural differences may provide an explanation for this. In chapter 1, the issue raised by the study by Biber (1987), which shows that American speakers consider British speakers more polite, formal and proper than they themselves are, was brought to light. The research conducted by Precht (2003), which indicates that this could possibly happen because the former tend to use lexical verbs instead of modal verbs much more often than British people do, was presented as a possible explanation for these stereotypes. The fact that the targeted ESL textbooks were produced in North America could be one of the reasons why they include less tentative or more assertive language.

It is then possible to answer the most important question. Do the targeted ESL textbooks prime the learner to use modal verbs by means of frequencies similar to the ones in written English or in spoken English? Statistical evidence and its discussion showed that these textbooks tend to prime the ESL learner to use modal verbs in frequencies similar to those in the written English corpus. Thus, it may be said that Willis' (1990) assertion that there is an attempt for ESL textbooks to teach ESL learners to speak written language is true. Nevertheless, this assertion needs to be verified by the conclusions about the frequencies of modal verb patterns.

5.4 Conclusions: Objective 2. Modal Verb Patterns

The first research question posed in the *Research Questions* section in Chapter 1 regarding the frequencies of modal verb patterns was what patterns the ESL textbooks that compose the textbook corpus prime the ESL learner to use and what frequencies they have. Five patterns that have a frequency which is higher than one per cent out of the total number of occurrences of modal verbs in the textbook corpus have been found: *MVb + inf* , *MVb + be + past pple*, *MVb + pers.pron*, *MVb + adv + inf* and *MVb + have + past pple*. The next question asked whether the frequencies of these patterns in the textbook corpus are similar to those in spoken English and/or in written English.

As expected, the ESL textbooks that compose the textbook corpus prime students to use the infinitive after modal verbs mostly. The pattern *MVb + inf* accounts for almost 65.3 % of all the modal verb patterns in the textbook corpus. The study revealed that the mean frequency of this pattern in the textbook corpus is not significantly different from its mean frequency in either the spoken English corpus or the written English one. The same applies to the pattern *MVb + adv + inf*, whose mean frequencies in the three corpora are not significantly different. As regards the patterns *MVb + be + past pple* and *MVb + pers.pron*, statistical results showed that their mean frequencies in the textbook corpus are significantly different from their frequencies in the written English corpus but similar to the ones in the spoken English corpus. It may be said, then, that the textbooks in question prime the ESL learner to use modal verb patterns in frequencies similar to those in spoken English. However, the study shows that the last pattern under scrutiny, *MVb + have + past pple*, has a very low frequency in the textbook corpus if compared with its frequency in the written English corpus and in the spoken English one. It is now possible to answer the last question: if modal verb patterns in the textbook corpus tend to be simple as it is the case in spoken English or if they are more complex as in written English. This research indicates that the ESL textbooks that belong to the textbook corpus prime learners to use simple modal verb patterns, even simpler than in spoken English. The fact is that both the complex patterns *MVb + be + past pple* and *MVb + have + past pple* account for 20.5 % of all the modal verb patterns in the written English corpus, for 6.5 % in the spoken English corpus and for 4.8 %

in the textbook corpus. If it is true, as Sinclair and Hoey argue, that the higher the frequency of patterns speakers are exposed to, the stronger the speakers' primings of these patterns are, it may be concluded that the ESL textbooks that compose the textbook corpus will prime the learner to use modal verb patterns in frequencies similar to those in spoken English, except for the pattern *MVb + have + past pple*, whose frequency is lower in textbooks.

The analyses of each pattern across modal verbs gave more insight into their frequencies. They showed that the textbooks that make up the textbook corpus will prime the ESL learner to use the patterns *can + inf* and *will + inf* instead of the patterns *'d + inf* and *'ll + inf*. In contrast, the contracted negative forms followed by an infinitive in the textbook corpus are more frequent than the corresponding full negative forms, aside from *must*, *shall* and *should*, whose full negative forms might be more numerous in an attempt to create a stronger effect on the learner.

With regard to the interrogative pattern *MVb + pers.pron*, the study shows that the textbook corpus aligns with the spoken English corpus in the case of the affirmative modal verb forms. However, the ESL learner receives very little priming to use *shall + pers.pron* or a contracted negative modal verb form before a personal pronoun. Therefore, the targeted ESL textbooks prime students to use this pattern with affirmative modal verb forms only.

As for the pattern *MVb + adv + inf*, no big differences between the textbook corpus and the spoken English corpus and the written English one have been reported in this study. Even though this is the second most frequent pattern containing the forms *'d* and *'ll* across the three corpora, it is not an important source of priming for the ESL learner who uses the textbooks in question because figures are very low in the textbook corpus.

The data showing the pattern *MVb + be + past pple* containing individual modal verb forms confirmed the high frequency of this pattern in written English compared with the textbook corpus. When it was said that the frequencies of the patterns *can't + be + past pple*, *shouldn't + be + past pple*, and *will not + be + past pple* were higher in the textbook corpus than in the written English corpus, it was pointed out that this was not significant because

figures were very low and because they did not compensate for the large number of occurrences of this pattern in written English.

Finally, the analysis of the pattern *MVb + have + past pple* containing specific modal verb forms confirmed this is the least frequent pattern in the textbook corpus. The study shows that the occurrences of this pattern with the verbs *could*, *may*, *might*, *should*, *must* and *would* in the written English corpus and in the spoken English corpus, and with *'d* in the spoken English corpus only, are the most frequent. Aside from *'d*, the modal verb forms *could*, *may*, *might*, *should*, *must* and *would* are also the ones that occur the most frequently in this pattern in the textbook corpus. However, their frequencies in the textbook corpus are much lower than their frequencies in written and in spoken English.

5.5 Conclusion

The ESL textbooks in question prime the ESL learner to use modal verbs in frequencies which are similar to those in the written English corpus. They do not prime the learner to use the contracted form of the affirmative modal verbs, which are frequently used in oral communication but much less frequently in written English. In addition, the two modal verbs these textbooks prime the learner to use the most are *can* and *will*, which coincides with the studies by Watts and Garcia-Carbonell (2007) and by Debbie (2009), which indicate that ESL learners prefer using these two verbs to *could* and *would*. However, in the textbook corpus, modal verbs are used in simple patterns, as they are mostly used in spoken English.

This implies that the teachers that use these textbooks as the main component of the ESL course or courses that they teach will need to supplement the syllabus with a variety of authentic reading and listening materials that reflect the use of the contracted form of the affirmative modal verbs in oral contexts and that serve as further input to prime the learner to use them. Supplementary materials will also be required to prime the learner to master the use of modal verbs in complex patterns. Likewise, textbook writers should look for frequent collocations in existing written or spoken English corpora or both (depending on the aims of the course), and, at the same time, rely on authentic sources such as real business

presentations, letters or everyday conversations to make sure that the language the learner encounters in textbooks is not only native-like and fluent, but also adequate to the register and the context in question, be it as informal as small talk or as formal as business correspondence.

5.6 Further Research Possibilities

This research opens some possibilities for further study. The same methodology could be used to find out whether the trends that have been identified concerning modal verb frequencies and modal verb patterns in this study are also true of other collections of upper intermediate ESL materials. Are modal verb frequencies in other upper intermediate ESL textbooks similar or different from modal verb frequencies in the textbook corpus? And are they similar to modal verb frequencies in spoken English or in written English? Do other ESL textbooks contain a similar frequency of contracted modal verb forms? Can a similar simplification of modal verb patterns be reported in other ESL textbooks? Do other ESL textbooks pair with the spoken English corpus as regards the frequencies of modal verb patterns? Do other upper intermediate textbooks present a smaller, a bigger or a similar proportion of past modal verbs?

A second research possibility could be the study of other categories of speech, such as adverbs or adjectives, to see if the conclusions drawn concerning the use of modal verbs in the present corpus of ESL textbooks apply as well. Which are the most frequent adjectives or adverbs in the textbook corpus, and what is their frequency in the spoken English corpus and in the written English corpus? Do adjectives and adverbs in the textbook corpus collocate similarly or differently from written and spoken English? Does language simplification apply in this regard?

Another interesting research line would be to focus on meaning and compare the communicative functions of modal verbs in written English and in spoken English with the ones they perform in any corpus of ESL textbooks. Is the use of modal verbs simplified as well? Can we speak of an overuse of *can* in textbooks to express ability compared with a

wider variety of uses in spoken and/or in written English? These are some of the fascinating options that the study of textbook language allows. The doors it opens are endless and the possibilities immense.

APPENDIX A: CAN

Table A
Occurrences of the Patterns Containing any of the Forms of the Modal Verb *Can* per Million Words in the Three Corpora

	BNCW	BNCS	Textbook Corpus
Can			
MVb + inf	925	1,845	1,986
MVb + have + past pple	0	0	0
MVb + be + past pple	516	63	142
MVb + pers.pron	96	769	630
MVb + adv + inf	275	187	216
Cannot			
MVb + inf	163	61	126
MVb + have + past pple	0	0	0
MVb + be + past pple	64	9	24
MVb + pers.pron	0	1	0
MVb + adv + inf	7	2	4
Can't			
MVb + inf	156	865	283
MVb + have + past pple	1	1	0
MVb + be + past pple	4	13	8
MVb + pers.pron	19	150	8
MVb + adv + inf	3	60	24
Total	**2,229**	**4,026**	**3,451**

APPENDIX B: COULD

Table B
Occurrences of the Patterns Containing any of the Forms of the Modal Verb *Could* per Million Words in the Three Corpora

	BNCW	BNCS	Textbook Corpus
Could			
MVb + inf	745	783	625
MVb + have + past pple	59	77	43
MVb + be + past pple	191	54	8
MVb + pers.pron	50	182	173
MVb + adv + inf	117	99	63
Could not			
MVb + inf	98	16	60
MVb + have + past pple	7	1	0
MVb + be + past pple	18	0	8
MVb + pers.pron	0	0	0
MVb + adv + inf	5	0	0
Couldn't			
MVb + inf	93	272	114
MVb + have + past pple	7	13	8
MVb + be + past pple	1	3	0
MVb + pers.pron	2	32	4
MVb + adv + inf	4	22	0
Total	**1,397**	**1,554**	**1,106**

APPENDIX C: MAY

Table C
Occurrences of the Patterns Containing any of the Forms of the Modal Verb *May* per Million Words in the Three Corpora

	BNCW	BNCS	Textbook Corpus
May			
MVb + inf	551	224	432
MVb + have + past pple	60	29	20
MVb + be + past pple	179	11	28
MVb + pers.pron	20	45	59
MVb + adv + inf	77	26	39
May not			
MVb + inf	38	32	63
MVb + have + past pple	6	5	0
MVb + be + past pple	1	2	0
MVb + pers.pron	0	0	0
MVb + adv + inf	2	1	4
Total	**934**	**375**	**645**

APPENDIX D: MIGHT

Table D
Occurrences of the Patterns Containing any of the Forms of the Modal Verb *Might* per Million Words in the Three Corpora

	BNCW	BNCS	Textbook Corpus
Might			
MVb + inf	230	401	310
MVb + have + past pple	70	56	12
MVb + be + past pple	53	20	16
MVb + pers.pron	5	7	4
MVb + adv + inf	36	59	12
Might not			
MVb + inf	8	11	12
MVb + have + past pple	3	5	0
MVb + be + past pple	0	0	0
MVb + pers.pron	0	0	0
MVb + adv + inf	3	4	0
Total	**408**	**563**	**366**

APPENDIX E: SHALL

Table E
Occurrences of the Patterns Containing any of the Forms of the Modal Verb *Shall* per Million Words in the Three Corpora

	BNCW	BNCS	Textbook Corpus
Shall			
MVb + inf	129	119	11
MVb + have + past pple	1	1	0
MVb + be + past pple	22	3	0
MVb + pers.pron	18	137	4
MVb + adv + inf	19	8	0
Shall not			
MVb + inf	6	2	4
MVb + have + past pple	0	0	0
MVb + be + past pple	0	0	0
MVb + pers.pron	1	0	0
MVb + adv + inf	2	0	0
Shan't			
MVb + inf	1	11	0
MVb + have + past pple	1	0	0
MVb + be + past pple	0	0	0
MVb + pers.pron	0	0	0
MVb + adv + inf	0	0	0
Total	**200**	**281**	**19**

APPENDIX F: DARE

Table F
Occurrences of the Patterns Containing any of the Forms of the Modal Verb *Dare* per Million Words in the Three Corpora

	BNCW	BNCS	Textbook Corpus
Dare			
MVb + inf	0	0	0
MVb + have + past pple	0	0	0
MVb + be + past pple	0	0	0
MVb + pers.pron	0	6	4
MVb + adv + inf	0	0	0
Dare not			
MVb + inf	0	0	0
MVb + have + past pple	2	0	4
MVb + be + past pple	0	0	0
MVb + pers.pron	0	0	0
MVb + adv + inf	0	0	0
Daren't			
MVb + inf	0	0	0
MVb + have + past pple	0	1	0
MVb + be + past pple	0	0	0
MVb + pers.pron	0	0	0
MVb + adv + inf	0	0	0
Total	**2**	**7**	**8**

APPENDIX G: SHOULD

Table G
Occurrences of the Patterns Containing any of the Forms of the Modal Verb *Should* per Million Words in the Three Corpora

	BNCW	BNCS	Textbook Corpus
Should			
MVb + inf	511	568	480
MVb + have + past pple	42	99	20
MVb + be + past pple	511	81	55
MVb + pers.pron	44	56	87
MVb + adv + inf	31	25	63
Should not			
MVb + inf	27	8	39
MVb + have + past pple	2	1	0
MVb + be + past pple	30	10	16
MVb + pers.pron	0	0	0
MVb + adv + inf	5	0	0
Shouldn't			
MVb + inf	6	76	20
MVb + have + past pple	2	12	4
MVb + be + past pple	1	7	4
MVb + pers.pron	4	18	0
MVb + adv + inf	1	4	0
Total	**1217**	**965**	**788**

APPENDIX H: MUST

Table H
Occurrences of the Patterns Containing any of the Forms of the Modal Verb *Must* per Million Words in the Three Corpora

	BNCW	BNCS	Textbook Corpus
Must			
MVb + inf	476	384	346
MVb + have + past pple	65	116	20
MVb + be + past pple	158	13	35
MVb + pers.pron	11	6	0
MVb + adv + inf	35	13	12
Must not			
MVb + inf	25	4	47
MVb + have + past pple	1	0	0
MVb + be + past pple	9	1	0
MVb + pers.pron	0	0	0
MVb + adv + inf	0	0	4
Mustn't			
MVb + inf	7	9	31
MVb + have + past pple	0	0	0
MVb + be + past pple	0	0	0
MVb + pers.pron	0	9	0
MVb + adv + inf	0	0	0
Total	**787**	**555**	**495**

APPENDIX I: WILL

Table I
Occurrences of the Patterns Containing any of the Forms of the Modal Verb *Will* per Million Words in the Three Corpora

	BNCW	BNCS	Textbook Corpus
Will			
MVb + inf	1,956	1,083	3,209
MVb + have + past pple	21	10	8
MVb + be + past pple	358	76	146
MVb + pers.pron	45	168	271
MVb + adv + inf	237	87	173
'll			
MVb + inf	337	2,496	63
MVb + have + past pple	0	2	0
MVb + be + past pple	2	22	0
MVb + pers.pron	1	44	0
MVb + adv + inf	30	207	8
Will not			
MVb + inf	104	26	114
MVb + have + past pple	2	0	0
MVb + be + past pple	13	2	16
MVb + pers.pron	0	0	0
MVb + adv + inf	6	0	4
Won't			
MVb + inf	77	373	122
MVb + have + past pple	1	2	0
MVb + be + past pple	1	6	0
MVb + pers.pron	4	79	0
MVb + adv + inf	2	7	0
Total	**3,197**	**4,690**	**4,134**

APPENDIX J: WOULD

Table J
Occurrences of the Patterns Containing any of the Forms of the Modal Verb *Would* per Million Words in the Three Corpora

	BNCW	BNCS	Textbook Corpus
Would			
MVb + inf	1,369	1,293	810
MVb + have + past pple	198	218	24
MVb + be + past pple	163	55	12
MVb + pers.pron	75	315	547
MVb + adv + inf	147	86	40
'd			
MVb + inf	123	831	4
MVb + have + past pple	6	75	0
MVb + be + past pple	5	22	0
MVb + pers.pron	0	16	0
MVb + adv + inf	20	117	7
Would not			
MVb + inf	79	16	55
MVb + have + past pple	12	6	4
MVb + be + past pple	12	3	4
MVb + pers.pron	3	0	0
MVb + adv + inf	11	0	0
Wouldn't			
MVb + inf	49	314	59
MVb + have + past pple	6	27	0
MVb + be + past pple	0	11	0
MVb + pers.pron	9	2	16
MVb + adv + inf	3	13	0
Total	**2,290**	**3,420**	**1,582**

REFERENCES

Biber, Douglas. 1987. « A Textual Comparison of British and American Writing». *American Speech*, vol. 62, no 2, p.99-119.

Bougie, Paul, Teresa Capparelli, Marie Claire Ford and Vicky Giannas Evridiki. 2007. *Quest : English as a Second Language, Secondary Cycle Two.* Montréal: Graficor, 272 p.

Colgan, Margaret-Anne, Monique Mainella and Michael O'Neill. 2007. *Prime Time: English as a Second Language, Secondary Cycle Two, Year One.* Coll. « Prime Time». Anjou: Les Éditions CEC, 212 p.

Debbie, G.E. HO. 2009. « Exponents of Politeness in Brunei English ». *World Englishes*, vol. 28, no 1, p. 35-51.

Dixon, Robert M.W. 1992. *A New Approach to English Grammar, on Semantic Principles.* Oxford: Clarendon Press, 398 p.

Ellis, Rod. 2003. *Second Language Acquisition.* Oxford : Oxford University Press, 147 p.

Gagné, Suzanne, Martine Picard and Elisa Shenkier. 2007. *Connecting through English: English as a Second Language : Secondary Cycle Two, Year One.* Coll. «Connecting through English». Laval: Grand Duc, 328 p.

Hinkel, Eli. 1995. « The Use of Modal Verbs as a Reflection of Cultural Values ». *TESOL Quarterly*, vol. 29, no 2, p. 325-343.

Hoey, Michael. 2007a. «Lexical Priming and Literary Creativity». In *Text, Discourse and Corpora: Theory and Analysis*, sous la dir. de Wolfgang Teubert et Michaela Mahlberg, p. 7-29. New York: Continuum.

_____. 2007b. «Grammatical creativity: a Corpus Perspective». In *Text, Discourse and Corpora: Theory and Analysis*, sous la dir. de Wolfgang Teubert et Michaela Mahlberg, p. 31-56. New York: Continuum.

_____. 2005. *Lexical Priming: A New Theory of Words and Language.* New York: Routledge, 202 p.

_____. Lexical Priming and the Properties of Text. On line. <http://www.monabaker.com/tsresources/LexicalPrimingand thePropertiesofText.htm>

Hoye, Leo. 1997. *Adverbs and Modality in English.* Coll. «English Language Series». No 21. Harlow: Addison Wesley Longman, 322 p.

Hunston, Susan, and Gill Francis. 2000. *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English.* Coll. «Studies in Corpus Linguistics», no 4. Amsterdam: John Benjamins, 288 p.

Krug, Manfred G.2000. *Emerging English Modals: A Corpus-Based Study of Grammaticalization* . Coll. «Topics in English Linguistics». No 32. Berlin: Mouton de Gruyter.Addison , 332 p.

Leech, Geoffrey, Paul Rayson and Andrew Wilson. 2001. *Word Frequencies in Written and Spoken English: based on the British National Corpus.* London: Longman, 320 p.

Lewis, Michael. 2000a. « Learning in the Lexical Approach ». In *Teaching Collocation: Further Developments in the Lexical Approach,* sous la dir. de Michael Lewis, p. 155-185. Hove: Language Teaching Publications.

_____. 2000b. « Materials and Resources for Teaching Collocation ». In *Teaching collocation: Further Developments in the Lexical Approach,* sous la dir. de Michael Lewis, p. 186-204. Hove: Language Teaching Publications.

Lyung, Magnus. 1990. *A Study of TEFL Vocabulary: Acta Universitatis Stockholmiensis 78* (Stockholm, 1990). Stockholm: Almqvist & Wiksell; cité dans Elena Tognini-Bonelli, *Corpus Linguistics at Work,* Amsterdam, 2001, p. 40.

Mahlberg, Michaela. 2007a. «Corpus Stylistics: Bridging the Gap between Linguistic and Literary Studies». In *Text, Discourse and Corpora: Theory and Analysis*, sous la dir. de Wolfgang Teubert et Michaela Mahlberg, p. 219-242. New York: Continuum.

Mahlberg, Michaela. 2007b. «Lexical Items in Discourse: Identifying Local Textual Functions of Sustainable Development». In *Text, Discourse and Corpora: Theory and Analysis*, sous la dir. de Wolfgang Teubert et Michaela Mahlberg, p. 191-218. New York: Continuum.

Mason, Margaret Ruth. 2007. « The Development of Narrative Skills and the Evaluative Use of Modal Verbs in the Narratives of Young Non-Native Speakers of English». *International Journal of Applied Linguistics*, vol.4, no 1, p. 79-99.

Matthews, Richard. 2003. « Modal Auxiliary Constructions TAM and Interrogatives» In *Modality in Contemporary English*, sous la dir. de Roberta Facchinetti, Manfred G. Krug et Frank Robert Palmer , p. 47- 70. Berlin: Walter de Gruyter.

Montero, Begoña, Frances Watts and Amparo García Carbonell. 2007. « Discussion Forum Interactions: Text and Context ». *System*, no 4, p. 566-582.

Palmer, Frank R. 1979. *Modality and the English Modals.* Coll. «Longman Linguistics Library ». No 23. London: Longman, 196 p.

Parks, Philippa, and Tanja Vaillancourt. 2007. *Express Yourself: English as a Second Language : First Year of Secondary, Cycle Two.* Coll. «Express Yourself». Montréal: Éditions du Renouveau Pédagogique, 198 p.

Partington, Alan. 1998. *Patterns and Meanings: Using corpora for English Language Research and Teaching.* Coll. «Studies in Corpus Linguistics», no 2. Amsterdam: John Benjamins, 162 p.

Perkins, Michael R. 1983. *Modal Expressions in English.* Coll. «Open Linguistics Series ». London: Frances Pinter, 186 p.

Precht, Kristen. 2003. « Great vs. Lovely: Grammatical and Lexical Stance Differences in American and British English. ». In *Corpus Analysis: Language Structure and Language Use*, sous la dir. de Charles Meyer et Pepi Leistyna, p. 133-152. Amsterdam: Rodopi.

Québec, Ministère de l'Éducation, du Loisir et du Sport du Québec. 2008. *English as a Second Language: Core Program, Enriched Program.* Québec: Les Publications du Québec, 56 p.

Quereda Rodriguez-Navarro, Luis. El Modelo Lingüístico de J.R. Firth. On line <http://www.ugr.es/~lquereda/modelo_firth.htm>

Quirk, Randolph, and Sidney Greenbaum. 1973. *A University Grammar of English.* London: Longman, 484 p.

Roland, Douglas, Frederick Dick and Jeffrey Elman. 2007. « Frequency of Basic English Grammatical Structures: A Corpus Analysis ». *Journal of Memory and Language*, no 57, p. 348-379.

Sinclair, John. 1991. *Corpus, Concordance, Collocation.* Oxford : Oxford University Press, 179 p.

Sinclair, John. 2007. «Introduction». In *Text, Discourse and Corpora: Theory and Analysis*, sous la dir. de Wolfgang Teubert et Michaela Mahlberg, p. 1-5. New York: Continuum.

Stubbs, Michael. 1996. *Text and Corpus Analysis : Computer-Assisted Studies of Language and Culture.* Coll. «Language in Society». No 23. Cambridge: Blackwell, 267 p.

_____. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics.* Malden: Blackwell, 267 p.

_____. 2007a. «On Texts, Corpora and Models of Language». In *Text, Discourse and Corpora: Theory and Analysis*, sous la dir. de Wolfgang Teubert et Michaela Mahlberg, p. 126-161. New York: Continuum.

_____. 2007b. «Quantitative Data on Multi-Word Sequences in English: the Case of the Word *World*». In *Text, Discourse and Corpora: Theory and Analysis*, sous la dir. de Wolfgang Teubert et Michaela Mahlberg, p. 163-189. New York: Continuum.

Teubert, Wolfgang. 2007. «Natural and Human Rights, Work and Property in the Discourse of Catholic Social Doctrine». In *Text, Discourse and Corpora: Theory and Analysis*, sous la dir. de Wolfgang Teubert et Michaela Mahlberg, p. 89-126. New York: Continuum.

Thomson, Audrey Jean, and Agnes V. Martinet. 1980. *A Practical English Grammar.* Oxford: Oxford University Press, 369 p.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work.* Coll. «Studies in Corpus Linguistics», no 6. Amsterdam: John Benjamins, 223 p.

Vince, Michael. 1994. *Advanced Language Practice.* Oxford: Heinemann, 264 p.

Willis, Dave. 1990. *The Lexical Syllabus : A New Approach to Language Teaching.* London: Collins, 136 p.