A Corpus-based study of *perhaps*, *maybe*, *probably*, and *possibly* in native and non-native English speech

University of Eastern Finland
Philosophical Faculty
School of Humanities
English Language and Culture
Sonja Moalla, 291301

ITÄ-SUOMEN YLIOPISTO – UNIVERSITY OF EASTERN FINLAND

Tiedekunta –Faculty	**Osasto – School**
Philosophical Faculty	School of Humanities

Tekijä – Author
Sonja Moalla

Työn nimi – Title
A Corpus-based study of *possibly*, *perhaps*, *probably*, and *maybe* in native and non-native speech

Pääaine – Main subject	**Työn laji - Level**		**Päivämäärä - Date**	**Sivumäärä – Number of pages**
English Language and Culture	Pro gradu -tutkielma	X	20.07.2020	67
	Sivuainetutkielma			
	Kandidaatin tutkielma			
	Aineopintojen tutkielma			

Tiivistelmä – Abstract

The aim of this thesis is to study four epistemic adverbs *perhaps*, *maybe*, *possibly*, and *probably* in native and non-native English speech. In particular, the focus is to observe the differences and similarities in the use of the four epistemic adverbs between these two groups of speakers. The comparison of the occurrences of *perhaps*, *maybe*, *possibly*, and *probably* between native and non-native speech puts the understanding of pragmatics across. The following research questions are used to reach the aim:

1.To what extent does the frequency of *perhaps*, *maybe*, *probably*, and *possibly* differ between native and non-native speakers of English in the spoken corpora?

2.How do native and non-native English speakers use *perhaps*, *maybe*, *probably*, and *possibly* preceding and following the verb in the spoken language?

3.What are the most common verbs that co-occur with *perhaps*, *maybe*, *probably*, and *possibly* in MICASE and VOICE?

4.What kind of gender-based differences occur in the use of *perhaps*, *maybe*, *probably*, and *possibly* in MICASE and VOICE?

This thesis combines both quantitative and qualitative approaches to analyze the corpus. The primary materials used for the corpus searches are *The Vienna-Oxford International Corpus* (VOICE) and *The Michigan Corpus of Academic Spoken English* (MICASE). The quantitative analysis was carried out by analyzing the normalized frequencies and log-likelihood values for the four epistemic adverbs *perhaps*, *maybe*, *possibly*, and *probably*.

Based on the results of the first research question, the log-likelihood values between the two corpora show significant differences. For the second research question, only *perhaps* and *probably* show a significance difference between the two corpora. For the third and fourth questions, the majority of the adverbs show a significance difference between the native and non-native speech. The frequencies were normalized to 1,000,000 words in the first and fourth research questions, and to 1,000 words in the second and third research questions. The findings show that there are differences in the use of the four adverbs in the two corpora.

Avainsanat – Keywords
maybe, *perhaps*, *possibly*, *probably*, MICASE, VOICE, corpus, native, non-native

Contents

1. Introduction

Linguistic variation is central to the study of language use. In fact, different speakers use the same language and its resources differently. This variation might occur not only in grammatical and lexical levels, but speakers may also make choices concerning pronunciation, morphology, and the like (Tagliamonte 2006: 10). To express a level of doubt in the English language, one can use different expressions, verbs, modals, and adverbs. To notice this variation is the first simple act to conduct a variation analysis which is the core of this thesis. The variation analysis could be carried out using corpora which is a collection of linguistic data that is compiled from written or spoken material, usually using computer-assisted methods and techniques. This field in linguistics, corpus linguistics, can be applied to an increasing number of linguistic research questions and therefore has become a commonly used methodology in different areas of disciplines (Esimaje et al 2019: 1).

One important application of corpus linguistics is to study variations of a language. In fact, linguistic variation has been the interest of many professionals from different disciplines. One area that has been examined deals with modality, which is a system of linguistic options used to express the speaker's general intention. Modality is a way to express different degrees of necessity, obligation, intention, and probability (Aarts 2011: 275). The emphasis on the linguistic description of modality has mainly been on the meaning and use of modal verbs or modals (Leech 1971; Palmer 1979). Nevertheless, modal adverbs and epistemic modal adverbs have been addressed less commonly in the literature (Biber et al 1999; Tucker 2001).

Various approaches have been taken to study modality. Synchronic corpus studies have exposed the different frequencies of modals and semi-modals auxiliaries across registers. For instance, Biber et

al. (1999: 486) compare the distribution of different modals and semi-modals in the registers of news, fiction, conversation, and academic prose using *The Longman Spoken and Written English Corpus.* Their results, which are better described in Chapter 2, show that modals and semi-modals are most common in conversation and least common in news and academic prose. Similarly, Tucker (2001) conducts a corpus-based analysis on what he calls "secondary" modal resources which are lexical in nature and include *possibly* and *probably*. He focuses on the epistemic adverb *possibly* but also comments on *perhaps* and *maybe*. Examples (1) and (2) illustrate the way the epistemic adverb *possibly* is used in the native and non-native speech respectively:

(1) She thinks *possibly* by the end of this week. (MICASE: MTG999ST015)
(2) To counter- attack *possibly* before, before our general assembly. (VOICE: POmtg403: 508)

According to Davies (2003: 47) there is a consistent difference between native and non-native speech when it comes to linguistic features and communicative strategies. Thus, the central focus of the thesis is to observe the similarities and differences in the use of four epistemic adverbs: *perhaps*, *maybe*, *probably*, and *possibly* between native and non-native English speakers. The two spoken corpora used for this comparison are *The Vienna-Oxford International Corpus* (VOICE) which provides more than 1 million words of spoken English by non-native speakers in different context; the second corpus is *The Michigan Corpus of Academic Spoken English* (MICASE) which compromises almost 2 million words of English spoken by American native speakers. Examples (3) – (6) illustrate the use of modal adverbs under scrutiny in different environment in a sentence.

(3) *maybe* they wanted to end it on, a more happy note. (MICASE: LEL140SU074)
(4) we *perhaps* should directly just given em the letter. (MICASE: STP355SU010)
(5) they're going to change it *probably* for the whole semester. (VOICE: EDcon496:539)
(6) i (.) don't know but *possibly* you all e:r studied for the exam today. (VOICE: EDsed301:108)

English's position as an international language makes is a common choice for native and non-native speakers to use it in different contexts (Seidlhofer 2011: 7). The increased usage of English introduces similarities and differences between these two groups of speakers. It is therefore interesting to study whether and how *perhaps*, *maybe*, *probably*, and *possibly* work in the native and non-native speech. The topic of my thesis has not been explored in the non-native English context; thus, it offers a potential new insight for English linguistics. The research will be investigated using VOICE and MICASE which have been chosen for their suitability to conduct a variation study between the native and non-native speech.

The research questions of this thesis can be expressed as follows:

1) To what extent does the frequency of *perhaps*, *maybe*, *probably*, and *possibly* differ between native and non-native English speakers in the spoken corpora?

2) How do native and non-native English speakers use *perhaps*, *maybe*, *probably*, and *possibly* preceding and following verbs in the spoken language?

3) What are the most common verbs that co-occur with *perhaps*, *maybe*, *probably*, and *possibly* in MICASE and VOICE?

4) What kind of gender-based differences occur in the use of *perhaps*, *maybe*, *probably*, and *possibly* in MICASE and VOICE?

The structure of this thesis is organized as follows. In the first Chapter, I have introduced the topic and defined the aim and the research questions. Chapter 2 is divided in two subsections and contains the theoretical background of the study. Chapter 3 defines English as a Lingua Franca and provides an overview of the previous findings related to it. Chapter 4 explores the relation that exists between

language and gender. Chapter 5 includes the methods and materials that are used in this thesis. Chapter 6 presents the quantitative findings of the four research questions. Section 6.1 presents the frequencies of the occurrences of the four epistemic adverbs in native and non-native speech. Section 6.2 presents the frequencies of the positions of the four adverbs in native and non-native English speech. Section 6.3 details the frequencies of the most common verbs that occur next to *perhaps*, *maybe*, *probably*, and *possibly* in MICASE and VOICE. Section 6.4 reports on the gender-based differences in the use of the four epistemic adverbs in the native and non-native English speech. Lastly, chapter 7 discusses the data in relation to the theory reviewed and concludes with some implications.

2. Theoretical background

This chapter contains the theoretical background of the study. Section 2.1 defines modality, particularly, the three types of modality: the epistemic, deontic, and dynamic. It also accounts on the positions and frequencies of the modal adverbs. Section 2.2 presents the previous research related to this subject.

2.1.Modality

Numerous divisions have been proposed on the number and types of modal subdomains. To state the various definitions of modality, however, is beyond the scope of this current investigation. In this thesis, a narrow definition of modality presented by Aarts (2011: 275) will be applied. According to him, modality is a concept that is concerned with semantic notions such as "possibility", "probability" , "necessity", "obligation", "permission", "intention", and "ability". These terms involve a situation of non-factuality: a situation that does not imply a straightforward fact. Aarts (2011: 275) identifies three types of modality in English: deontic modality, epistemic modality, and dynamic modality. This distinction is similarly expressed by Collins (2009: 11).

Perhaps, *maybe*, *possibly*, and *probably* are adverbs that belong to epistemic modality, which is one of the two subdivisions of modality. In fact, most authors agree that modality is subdivided into root modality and epistemic modality (Aarts 2011). Root modality is further divided into deontic and dynamic modality. This study follows the division of Aarts (2011), as illustrated in Figure 1 below.

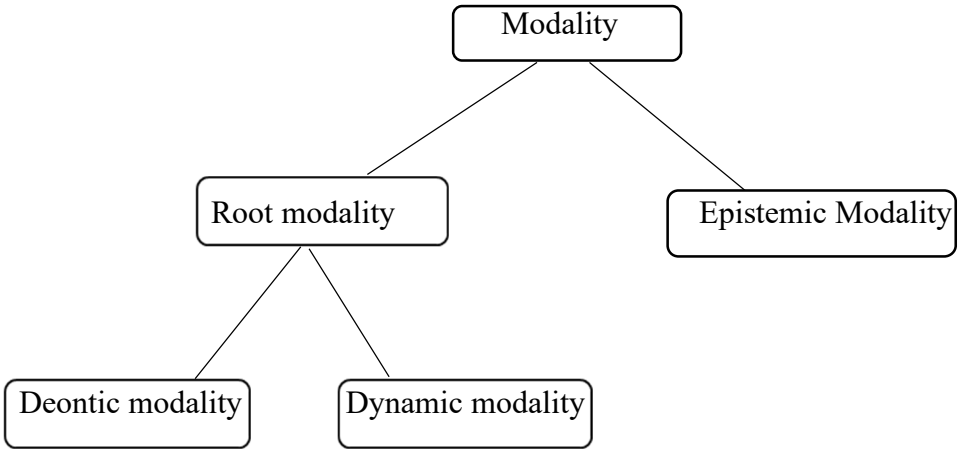

Figure 1. *The relation between different types of modality* (Aarts 2011: 275).

The concepts of epistemic modality, deontic modality, and dynamic modality and their differences will be discussed further in the following section.

2.1.1. Types of modality

There are three types of modality in English which are deontic, dynamic, and epistemic (Aarts 2011: 275). Deontic modality refers to acts, not propositions. Historically, deontic modality is derived from the Greek for "binding", hence, it is a question of imposing obligation or prohibition, giving permission, and the like (Huddleston 2002: 178). Aarts (2011: 276) defines deontic modality as being "…concerned with getting people to do things or (not) allowing them to do things". Examples (7) and (8) illustrate this idea:

(7) They *must* come in now. (Palmer et al. 2003: 7)
(8) They *may/can* come in now. (Palmer et al. 2003: 7)

Deontic modality deals with the potentiality of the event indicated in the proposition. Example (7) is associated with notion of obligation, whereas example (8) is associated with permission. These two

examples above involve the issuing of directives and they show that the control is external to the subject. Therefore, with deontic modality, the ability comes from the permission given externally (Palmer et al. 2003: 7).

The second modality discussed in this section is the dynamic modality. It is concerned with the ability or the will of the subject and therefore, it is not subjective as the other modalities (Palmer 1990: 36). Dynamic modality is related to the potentiality of the event signaled by the proposition. However, as opposed to deontic modality, with dynamic modality the control comes from the subject's own internal ability (Palmer et al. 2003: 7). A similar description is mentioned by Nuyts (2001: 25) as he describes dynamic modality as an agent oriented. The following examples illustrate this idea:

(9) They *can* run very fast. (Palmer et al. 2003: 7)
(10) I *will* help you. (Palmer et al. 2003: 7)

Example (9) shows that it is in the subject's ability to run fast. Example (10) highlights the speaker's willingness to help, thus suggests that the control is within the subject's own internal ability.

The last type of modality explained in this section is the epistemic modality which is a concept derived from the Greek for "knowledge". It reveals the state of lack of knowledge of the speaker (Huddleston 2002: 178). Epistemic modality allows the speaker to express their attitudes towards the factuality of the situation, -i.e., it communicates the speaker's judgement on how likely the proposition expressed is true (Collins 2009: 21). Epistemic modality refers to "matters of knowledge or belief on which basis the speakers express their judgements about states of affairs, events or actions" (Hoye 1997: 42). The same idea is expressed by Biber et al. (1999: 854) stating that epistemic stance markers are defined as being concerned with the speaker's judgement about the information of the proposition.

Thus, epistemic stance markers convey the meaning of certainty, reliability, and limitation in the realm of the speaker's judgement. In addition, Lakoff (1973: 1) states that modal adverbs such as *perhaps* and *possibly* can also function as operators. In this thesis, epistemic modality follows the definition of Biber et al. (1999). Examples (11) – (13) show the speaker's limited level of certainty (Biber et al. 1999: 854):

(11) *Maybe* it's true, *maybe* it isn't.
(12) And *perhaps* the soul thrived on its sufferings.
(13) No it's alright I'll *probably* manage with it.

The adverbs *maybe*, *perhaps*, and *probably* used in the examples above communicate the feeling of doubt of the proposition in the clause. In fact, the three speakers in these examples communicate a feeling of uncertainty and a limited judgement of the situation.

Epistemic modality can be used in different linguistic forms such as epistemic phrases (*I think*, *I know*), adverbs (*probably*, *possibly*), nouns (*interpretation*), lexical verbs (*seem*, *assume*, *suggest*) and adjectives (*possible*, *probable*, *likely*). As seen in Figure 2 below, epistemic modality is realized by a number of English adverbials that cover seven main semantic categories which are the following: place, time, manner, degree, additive/restrictive, linking and stance. This thesis deals only with the category stance which is further divided into epistemic, attitude and style (Biber et al. 1999: 552). Figure 2 illustrates the categorization of the adverb and its different use.

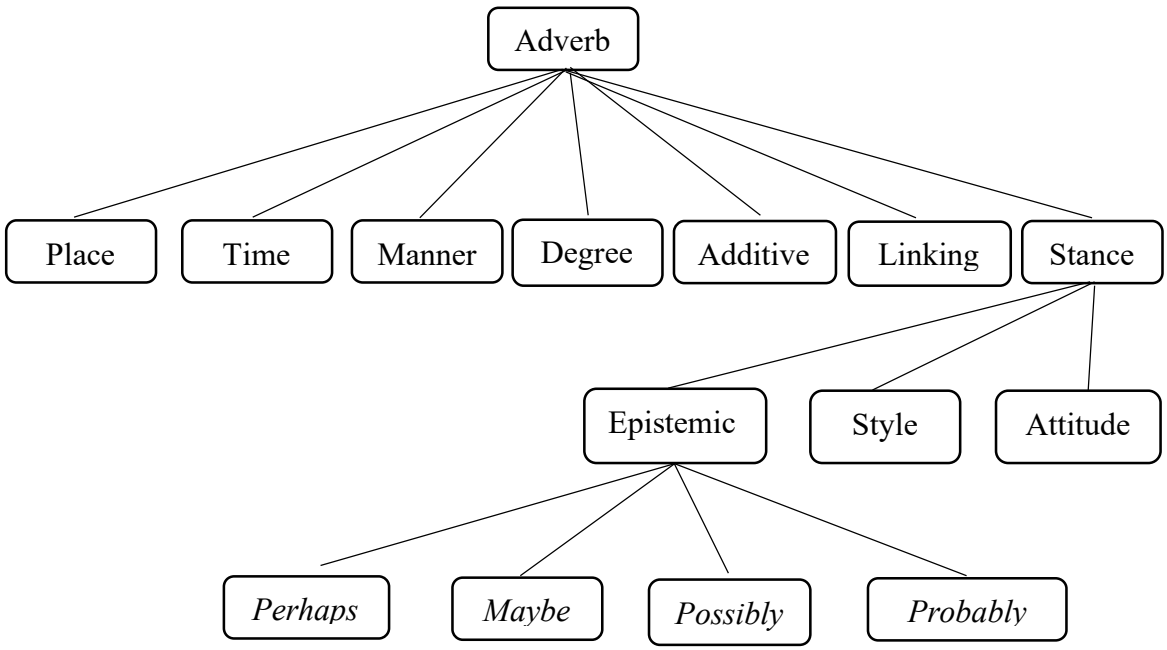

Figure 2. *Semantic subclasses of adverbs* (Biber et al. 1999: 552).

Perhaps, maybe, *probably*, and *possibly* belong to the category of epistemic stance that provides an assessment of the speaker's level of doubt. Biber et al. (1999: 966) identify stance adverbs as a linguistic feature used to express "individual attitudes, thoughts and feelings of the speaker". As the current study is mainly concerned with epistemic modals: *perhaps*, *maybe*, *probably*, and *possibly*, the focus will be restricted to the meaning of stance adverbials.

2.1.2. Frequencies and positions of epistemic modals

This subsection provides information on the frequencies and positions of adverbs. According to Biber et al. (1999: 765), stance adverbials that include *perhaps*, *maybe*, *possibly*, and *probably* are more common in conversation compared to other registers. This more frequent use of stance adverbials in the spoken language is related to the personal interactions in this register. In fact, stance adverbials

act as a device to help speakers communicate their thoughts, judgement, and attitudes (Biber et al. 1999: 767).

Biber et al. (1999: 854) investigate the subcategories of stance adverbials which are epistemic, attitude and style. Their results show that epistemic adverbials are more common than style and attitude stance adverbials in four domains: conversation, fiction, news, and academic prose (ibid: 859). These results are presented in Figure 3.

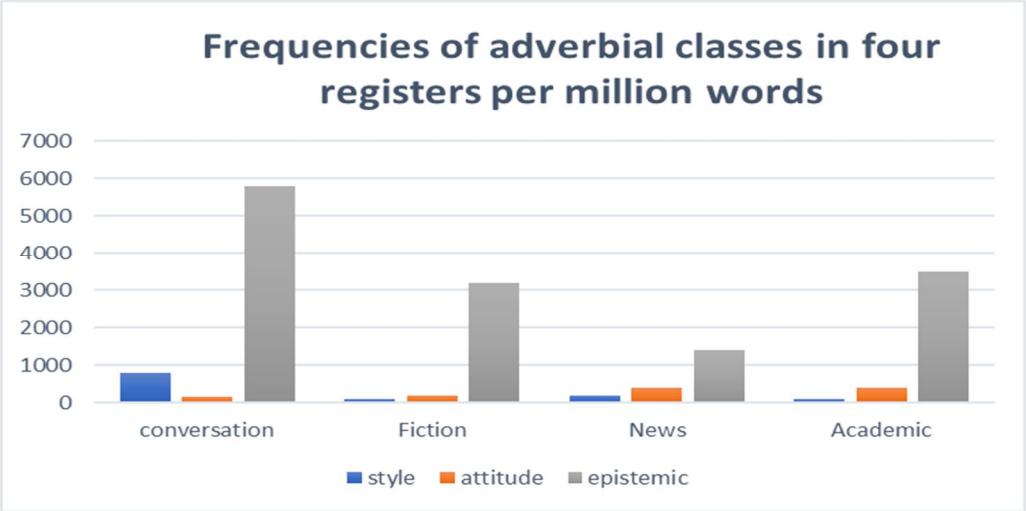

Figure 3. *Frequency of stance adverbials across registers (*Biber et al. 1999: 854).

Figure 3 shows that epistemic stance adverbials, which are colored in grey, are used the most in conversation. In fact, stance adverbials occur nearly 6,000 times per million words in conversation. While, style and attitude adverbials together account for around 1,000 times per million words. The high frequencies of epistemic adverbials in conversation show that they are part of the general communicative characteristics of the speaker. The following examples show the different ways speakers use the epistemic adverbs in conversations:

(14) i think *maybe* one-fifteen is a little bit too elementary for you. (MICASE: ADV700JU023)

(15) um, you can ask my mom, she *probably* would know. (MICASE: SVC999MX104)

(16) i think it's also (.) *perhaps* it's w- it's also wise just to s- to to send this. (VOICE: PBmtg414:886)

The high frequencies of epistemic adverbials in conversations reflect the speaker's feeling of doubt (Biber et al. 1999: 859). The use of *maybe*, *probably*, and *perhaps* in the examples above mark the hesitance and uncertainty of the speakers.

As seen in Figure 3 above, academic prose is the second most common domain that attests the use of epistemic modals with around 3,500 times per million words. Fiction comes in the third place with more than 3,000 instances per million words. Finally, the news accounts for 1,500 instances usage of epistemic stance adverbials per million words (Biber et al. 1999: 859). The most noteworthy conclusion in Biber et al. (1999) that relates to the present work is the high use of epistemic stance adverbials such as *maybe*, *perhaps*, *probably*, and *possibly* in conversations.

In addition to the frequency, one characteristic of epistemic adverbials is their position in a clause. The adverbs have the reputation of being able to take different positions in a clause (Nuyts 2001: 57). This significant nature makes its different positions interesting to study. In fact, same type of adverb might occur in different positions within the same language. The aim of this study is to account for this variation between the native and non-native English speech.

Huddleston and Pullum (2002: 575) distinguish three main positions for adverbs. First, the front position which refers to the adverb situated before the subject. Second, the end position refers to occurrence of the adverb after the verb. Third, the central position refers to the adverb being located between the subject and the verb. The use of *probably* and *contentedly* in the examples (17) – (21) better illustrate this distinction proposed by Huddleston and Pullum (2002: 575).

(17) *Probably* they would watch TV for hours. Front

(18) They *probably* would watch TV for hours. Pre-auxiliary central

(19) They would *probably* watch TV for hours. Post auxiliary central

(20) They would watch TC *contentedly* for hours. End

(21) They would watch TV for hours *contentedly*. End

Similarly, Biber et al. (1999: 770) distinguish four different adverbial positions. In their distinction, they refer to four positions as the following initial, medial, final, and "other speaker main clause". Initial position of the adverb occurs before the subject or other obligatory element of the clause as seen in the example below:

(22) *Maybe* second and third generation will maintain, at least some kind of closer identification with India. (Biber et al. 1999: 770)

The medial position includes all the positions between obligatory initial and final clauses elements. In this position the adverb can be located between the subject and the verb.

(23) Glenda *maybe* used a more, um a norm- a normal, type of writing. (Biber et al. 1999: 770)

The final position of the adverb occurs after all obligatory elements of the clause. The only exception is the presence of another final adverb in the clause as follows:

(24) Well, that's just so they don't have to write out those coordinates *probably*. (Biber et al. 1999: 770)

Lastly, Biber et al (1990: 771) describes "other speakers' main clause" as a position that occurs in conversations when one of the speakers seeks clarification, thus he adds an adverbial to the other person's utterance, for example:

a) I mean you don't have to pay *for those*. (Biber et al: 1990: 771)
b) Oh *for the films*. (Biber et al: 1990: 771)

Another distinction was proposed by Hoye (1997: 148) which I follow in my thesis. Hoye distinguishes up to seven different positions for the adverbial as illustrated in the table below.

Table 1. *Adverbial positions* (Hoye 1997: 148).

Position	***Examples***
Initial	*Possibly* they may have been sent to London.
Initial- medial	They *possibly* may have been sent to London.
Medial	They may *possibly* have been sent to London.
Medial-medial	They may have *possibly* been sent to London.
End- medial	They may have been *possibly* sent to London.
Initial- end	They may have been sent *possibly* to London.
End	They may have been sent to London *possibly*.

The positions that I will focus on in my thesis are the initial-medial, medial, medial-medial, end-medial, and initial-end positions. i.e., the positions of the epistemic adverbs that precede and follow the verb or the auxiliary verb.

The choice of selecting these particular positions to study is grounded on two different reasons. The first reason lies in the fact that the literature accounts for the high distribution of epistemic stance

adverbs in the middle position (Biber et al. 1999: 872). The second reason is the challenges I faced to define the boundaries of a sentence in the spoken corpora. These reasons are better explained in Chapter 6.

Maybe, *probably*, *perhaps*, and *possibly* enjoy considerable mobility in a clause compared to adverbials from other classes (Biber et al. 1999: 872). Biber et al. (1999: 872) investigate the position of stance adverbials in four registers academia, conversation, news, and fiction. According to their findings, all four registers display a preference for stance adverbials in the medial position as presented in Table 2.

Table 2. *The position of stance adverbials in conversation* (Biber et al. 1999: 872).

Position	***Percentage***
Initial position	*15%*
Medial position	*50%*
Final position	*35%*

In conversation, one reason of the high frequencies of stance adverbials in the medial position might be explained by the speaker's tendency to soften suggestions. This idea is illustrated in the following examples that are taken from Biber et al (1999: 874):

(25) Well I was thinking we could *perhaps* take the Blagden Hall now that's open.
(26) And then you could *perhaps*- tinkle the ivories.

The examples above show that the use of the epistemic adverb *perhaps* has a goal. In these examples, *perhaps* is used as a device suggesting that the speaker does not want to appear assertive and rather prefer to soften the suggestion.

2.2. Previous Research

In this section, I present a corpus-based study by Kärkkäinen (2003). In her study, she analyzes the use of epistemic stance in conversational American English using *Santa Barbara Corpus of Spoken American English*. This corpus consists of spontaneous American English audiotaped recordings from different social and ethnicity groups around the United States. Although her focus was mainly on the use of 'I think', she investigated the frequencies of epistemic phrases, adverbs, modals, and semi-modals using spoken language data comprising 5,402 intonation units.

The results presented in Kärkkäinen's (2003: 37) show that the epistemic adverbs are the second most common epistemic stance used in *Santa Barbara Corpus of Spoken American*. The results are demonstrated in Table 3 below.

Table 3. *Frequency of epistemic stance in conversation* (Kärkkäinen 2003: 37).

	Percentage	**Frequency**
Epistemic phrase	*60.7%*	*242*
Epistemic adverbs	*18.8%*	*75*
Epistemic modals and semi-modals	*17.8%*	*71*

Although Kärkkäinen's (2003: 37) findings indicate that epistemic adverbs do not represent a majority in American conversations, the epistemic adverbs *maybe* and *probably* were two among the most common epistemic adverbs that occurred in *The Santa Barbara Corpus of Spoken American English*. The findings show that epistemic phrases occur the most in the spoken language, whereas epistemic modals and semi-modals occur in the third place.

According to Kärkkäinen (2003: 45), the sentential adverbs are almost equal to modal auxiliaries' frequencies. Her results show that some of the most commonly used epistemic items in the database are the following adverbs: *maybe*, *probably*, *apparently*, *of course*, and *definitely.* These results are in alignment with Biber et al's (1999: 869). However, there was a minor difference in Kärkkäinen's (2003) and Biber et al' (1999) results regarding the order of frequency of the epistemic adverbs. According to the results presented by Biber et al (1999: 869), the order from the most used epistemic adverb to the least used was as follows: *probably*, *maybe*, of course, *certainly*, *definitely*, and *perhaps* in the American English database. Table 4 illustrates the most common epistemic markers found in the database of Kärkkäinen (2003: 37) regardless of their grammatical classes. As seen in the table below, unlike the results found by Biber et al (1999: 869), the epistemic adverb *maybe* occurs more often than *probably* in the American English corpus.

Table 4. *The list of most common epistemic markers* (Kärkkäinen 2003: 37).

Epistemic marker	***Frequency***
I think	*46*
s/he said'	*34*
I don't know	*28*
Maybe	*26*
I said	*26*
I don't know + compl.	*25*
I guess	*20*
I though	*18*
Probably	*17*

The list of epistemic markers is longer, but this is an illustration of the nine most frequent instances in Kärkkäinen's (2003) data. This list includes *maybe* and *probably*, two of the four adverb*s* that will be studied in this thesis. The list shows that the epistemic adverbs *possibly* and *perhaps* are not the most common in the spoken American data.

In this regard, it is interesting to find out the way the four epistemic adverbs under scrutiny are used in the non-native speech compared to the native speech. My thesis is an attempt to observe the differences and similarities in the use of *perhaps*, *maybe*, *possibly*, and *probably* in these two groups of speakers.

3. English as a lingua franca (ELF)

This study seeks to provide a comparison on the use of the four epistemic adverbs *perhaps*, *maybe*, *possibly*, and *probably* between native speakers, in particular, American English, and non-native speakers. This comparison is interesting as nowadays there are more non-native speakers of English compared to native speakers (Jenkins et al. 2017: 7). In fact, three out of four English users are non-native speakers (Crystal 2012: 68). As this study compares these two groups of speakers, it is noteworthy to explain the terms English as a lingua franca (hereafter ELF) and native speakers.

The term lingua franca is normally used to mean a contact language. Although sometimes, it could be used to refer exclusively to English, it includes any vehicular language between speakers who do not share a first language (Jenkins et al. 2017: 7). ELF is a relatively new field of study and has two widespread definitions. The first refers to a language contact by people for whom is it not a first language and thus, it excludes the native speakers. The second definition considers the native speakers as part of the mix (Jenkins 2017: 8). I follow the second definition which is also adopted by Seidlhofer (2011: 7). In this definition, Seidlhofer does not exclude native speakers of English but believes they consist a minority, thus it conforms to the description found in the non-native corpus (VOICE). In fact, it is mentioned in the VOICE description that speakers who use English as a first language only represent seven per cent of all the speakers recorded. Having said this, the use of VOICE as the corpus that represents the non-native speakers in my thesis is appropriate.

In addition to defining ELF speakers, it is important to define the term native speaker. A native speaker is defined as a person who acquires the first language during childhood (Davis 2003: 210).

A native speaker has intuition about the use of grammar, use the language creatively, able to produce fluent discourse, and knows the differences between their own speech and that of the standard form of the language (ibid).

For the past two decades or so, most references to ELF grammar presents the non-standard features (Jenkins et al. 2017: 248). In other words, the majority of the research related to ELF grammar has been carried out to examine the non-standard features in ELF at different levels including phonology, pragmatics, and lexicogrammar (Seidlhofer 2005: 340).

Seidlhofer (2004 quoted here from Jenkins et al. 2017: 249) report on the following non-standard features in the ELF data. The list of studies related to the feature spotting includes the following:

- Dropping the third person present tense -s (e.g., he work, instead of he works).
- Confusing the relative pronouns *who* and *which.*
- Omitting definite and indefinite articles where they are required in the native speech and place them where they do not occur. For example, “Most people do not understand *the* Japanese society” (Peterson 2019: 164).
- Incorrect use of tag questions (e.g., isn't it? or no? instead of, shouldn't they?)
- Inserting prepositions that are not required, as in (we have to study about ...).
- Overuse certain verbs of high semantic generality (such as *do*, *have*, *make*, *put*, *take*).

In addition, Deterding (2013: 107) mentions that ELF speakers have innovative patterns of grammatical usage. ELF speakers tend to care about getting their message across rather than worrying about grammar. For instance, in ELF, non-countable nouns are often expressed with a plural -s such as *furnitures, advices*, *informations*, *equipments* (Deterding 2013: 108). Although the list is longer, this is just an overview of the list of findings that refers to the non-standard features in ELF.

It is believed that listing the non-standard features can be used to provide targeted training especially in pronunciation (Jenkins 2002 quoted here from Peterson 2019: 168). However, some researchers have criticized the list of features that distinguish ELF users from the native speech (Peterson 2019: 141). In fact, Peterson does not agree with concepts such as "standard English" and "good English" and considers them a socially constructed truth (Peterson 2019: 179). In this regard, ELF is not viewed as a copy of the native speech but rather as rich variety that has developed independently (Seidlhofer 2001: 138). It is also suggested that non-standard English features are not limited to ELF users but are also found in the native speech (Ranta 2019: 111).

Having said this, the aim in my thesis is not to compare the native and non-native English speakers with the assumption that the native speech is the more competent of the two. This study is rather an attempt to observe the way these two groups of speakers vary in their usage of the four epistemic adverbs under scrutiny. In addition, *perhaps*, *maybe*, *possibly*, and *probably* have not been studied in ELF, thus, one aim of my study is an attempt to understand the ELF use of the epistemic adverbs.

4. Language and Gender

Language and gender are related in a non-straightforward manner. In fact, it is suggested that language use differs according to gender (Sunderland 2006: 5; Mesthrie et al. 2011: 219). Studying the relation between gender and language is part of sociolinguistics that examines the way social factors affect language use (Lindquist and Levin 2018: 156). These factors include information about the speaker, for instance, education, income, age, gender. The gender-related results in language use could enrich the findings related to pragmatic and discourse functions that will be tackled in the discussion section.

It is important form the outset to define the concept of gender which has traditionally been used as a synonym for sex. However, in sociolinguistics, these two terms do not mean the same thing. Gender is not simply limited to the biological or physiological sex, but is related to matters such sexual orientation, sexual identity and conceptualizations of masculinity and femininity (Mesthrie 2011: 218). Eckert et al. (2003: 10) distinguish between the two terms sex and gender and consider that "sex is a biological categorization based primarily on reproductive potential, whereas gender is the social elaboration of biological sex". In this thesis, due to the lack of information in the corpora, sex and gender will be used as synonyms and will refer to the biological categorization.

The link between gender and language have been discussed for a long time, yet the focus on this matter as field of research only extended during the 1960s and 1970s. Around that time, language and gender has become a lively field of research. This field of research was inspired by the feminist

movement and the way the use language mirrored the inequality between men and women in society during that time (Weatherall 2002: 2).

In my thesis, I will focus on the way native and non-native males and females use epistemic adverbs in conversations, more specifically the epistemic adverbs *perhaps*, *maybe*, *probably*, and *possibly*. There are several ways to study gender roles in language, one important way deals with discourse (Lindquist and Levin 2018: 156). One instance of this direction is put by Lindquist and Levin (2018: 156) as follows: "how men and women interact by means of language in different situations like everyday conversation or business meetings. Who speaks most? Who interrupts most frequently? Who gives more supportive feedback? Who laughs most often?". Studies based on discourse are frequently carried out using corpora, ideally marked up with information about the speakers. The aim in my thesis, is to find out which gender use the four epistemic adverbs under scrutiny more often. The finding to this question will be described in Section 6.4.

Previous studies have shown that men and women use language differently in many levels. Concerning the phonological variation, women tend to use higher levels of standard pronunciation of the -ing suffix (as in swimming), while men tend to use higher levels of non-standard -in' suffix (as in swimmin') (Mesthrie 2011: 223). Furthermore, women use more frequently the "r-ful" pronunciation, while men tend to use more often the r-less pronunciation as in [pa:k] for park (Labov 1966b quoted here in Mesthrie 2011: 223). From these observations, one might conclude that women use more hypercorrect grammar compared to men as expressed by Lakoff and Bucholtz (2004: 80).

Another difference in use of linguistic forms between men and women concerns the use of adjectives. In fact, women use more "empty adjectives" such as "cute" and "divine" compared to men (Lakoff 2004: 78). Lindquist and Levin (2018: 166) show that women use more color terms than men. The word pink, white and blue are used considerably more often by female writers (ibid). There are different possible explanations for this difference in language use, one interpretation could be linked to the fact that female writers deal with different topics, where the description of colors is more relevant (ibid).

In addition, Lakoff and Bucholtz (2004: 78) discuss gender-based differences and concludes that women use:

- more tag questions (such as "it's hot here, isn't it") (Lakoff and Bucholtz 2004: 78).
- less use of humor (Lakoff and Bucholtz 2004: 80).
- use of super polite forms such as "please" "thank you" (Lakoff and Bucholtz 2004: 80).
- use hypercorrect grammar compared to men (Lakoff and Bucholtz 2004: 80).

In an article called Gender Differences in Language Use: An Analysis of 14,000 Text Samples, Newman et al. (2008) examine the way males and females use language. The authors observe that women use more intensive verbs, more conjunctions, and more auxiliary verbs. Men, on the other hand, have been found to use more articles, more reference to location and swear words. In addition, it is suggested that women use a language style that reflects shyness and indicates lack of commitment or strong opinion (Eckert & McConnell 2003: 188). One example is the use of euphemisms, where women avoid swear words and replace them by words such as "piffle" or "heck".

Another gender-based difference in language use, which is closely related to my thesis, is the use of hedges. The most obvious similarity between the epistemic adverbs such as (*perhaps*, *maybe*, *possibly*, and *probably*) and hedges is that it allows the speakers to signal a level of caution in making assertion. In fact, women speech includes many instances of hedging such as "well", "you know", "kinda", and so forth (Lakoff and Bucholtz 2004: 79). The use of hedging reflects a feeling of uncertainty about what the speakers are saying. It might also be used to make a statement less harsh (ibid). The use of hedges does have their legitimate use; however, women seem to include them more often in their speech. As it is suggested that women use more hedges, it would be interesting to find out they way *perhaps*, *maybe*, *possibly*, and *probably* occur in women's and men's speech.

In this regard, Facchinetti et al. (2012) show that epistemic modals occur more in women's speech compared to men. The high use of epistemic modals among women does not necessarily mean that women are unassertive or uncertain as suggested by (Lakoff and Bucholtz 2004: 79). In fact, it might be quite the opposite. It is possible that women exploit the different functions of epistemic modal forms in their speech because of their sensitivity to interpersonal aspects of talk **(**Facchinetti 2012: 346). In fact, epistemic modals can be used as a device to signal that the speaker is searching for the right word, to avoid "playing the expert" and to negotiate a sensitive topic that is related to feelings and people (Facchinetti 2012 :337– 339).

In this chapter, I summarized some of the findings concerning gender-based differences in language use. The goal in my thesis is to observe if any differences arise among women and men in the use of *probably*, *possibly*, *perhaps*, and *maybe* in the two corpora. The results are presented in Section 6.4.

5. Material and Methodology

This section presents the materials and the methods used in this research. First, 5.1 describes the materials used in this thesis. It defines the notion of corpus and describes in detail the two corpora used which are VOICE and MICASE. Then, 5.2 illustrates the methods used and demonstrates the stages of the analysis included in this study.

5.1. Corpus

In my thesis, the use of corpora allows me to gain insights on the way the four epistemic adverbs *perhaps*, *maybe*, *possibly*, and *probably* occur in the native and non-native speech. In addition, it enables me to observe the frequency and the position of these four epistemic adverbs in MICASE and VOICE. In fact, to observe the differences that occur between these two groups of speakers, it is not possible to just ask a person how a human being communicates. The answer to that question will be considered as a subjective report that only shows how that person uses the language and how his mind reflects it (Gonzalez-Marquez 2007: 53). It is important, therefore, to use more objective methods to conduct a study. One important way that pervades language representation, processing, and language change is word frequencies which is conducted using corpora (Jannedy 2003: 3). In fact, the reliance on corpus data is essential for an adequate account of *possibly*, *perhaps*, *maybe*, and *probably* as it represents frequent real-life examples. As this research is corpus based, it is important to first gain a good understanding of the term corpus and corpus linguistics.

In the humanities, corpus is generally used to refer to a collection of texts that could be written and spoken (McEnery et al. 2006: 4). Another more detailed definition can be put as follows "a corpus (plural: corpora) is a collection of written or spoken material, occurring naturally, stored on computer, and typically used to carry out some kind of linguistic analysis" (Esimaje et al. 2019: 7). These two definitions entail that a corpus is not restricted to written texts but can also represent a spoken transcribed data which can be searched for different purposes.

The use of electronic corpora began around 1960, when computers started becoming powerful to some degree. The use of corpus has become a commonly used approach to answer different questions related to linguistics (Esimaje et al. 2019: 1). In fact, recent advances in corpora linguistic paved the way to obtain and analyze data using quantitative and qualitative approaches in a more efficient way. The former includes extracting the frequency of occurrence of a feature to be studied as it is the case in this thesis. In other words, it is used to observe the prevalence of some linguistic features in the corpus. The latter provides an in-depth understanding of different types of features that occur within the sample (Gonzalez-Marquez 2007: 36). The quantitative approach was mainly used in this thesis to determine the frequencies and positions of the four epistemic adverbs *perhaps*, *maybe*, *possibly*, and *probably*. In addition, the qualitative approach was used to provide more information concerning the third research question.

There are different types of corpora, such as spoken, written, monolingual, multilingual specialized. In my study, the two corpora used MICASE and VOICE are spoken. In this regard, it is noteworthy to mention that the compilation of written corpora is easier than spoken corpora. The former only requires the assembling of written texts in electronic format. For the spoken corpora, however, a much

greater effort is needed. The compilation of spoken corpora includes the recording of the conversations occurring between the individuals and later transcribing them. The transcription process brings some challenges as it requires choices that are not obvious. This challenge could be a valid reason behind the large number of written corpora compared to the spoken corpora.

Spoken language is the most common way of communication. Yet, spoken corpora are not as common as written corpora as their collection is time consuming and expensive to record and transcribe (Kennedy 1998: 20). Although spoken corpora are not as frequent as the written corpora, VOICE and MICASE compromise over one-million-word each. The number of words that each corpus compromises provide an opportunity to conduct a contrastive analysis between native and non-native speakers of English. This comparison makes it possible to perceive the way the four epistemic adverbs *possibly*, *perhaps*, *maybe*, and *probably* are utilized in the spoken language between native and ELF speakers.

VOICE is a perfect fit for this study as it provides 1,023,082 words of spoken English by non-native speakers in different contexts. The VOICE project works with the definition of ELF which include a marginalized minority of all speakers recorded. The audio material was recorded in 2010 and took around 110 hours and 35 minutes of recorded and transcribed interactions to complete this corpus. The speakers in this corpus have 49 different first languages including non-Europeans which will be treated in this paper as one entity of ELF speakers. VOICE is freely available for academic research and found online at: http://voice.univie.ac.at./

The second corpus that will be used in this thesis is MICASE. It is a collection of transcripts of American English spoken language developed at English Language Institute, University of Michigan, United States. Around 190 hours material was recorded and orthographically transcribed to form 1,848,364 word-corpus. MICASE comprises data from different locations across the university and different speech events such as lectures, classroom discussions, lab sections and seminars. This collection of transcripts is freely available for academic research and found online at: https://quod.lib.umich.edu/cgi/c/corpus/corpus?page=home;c=micase;cc=micase

5.2 Methods

To answer the research questions, MICASE and VOICE were analyzed quantitatively with the help of computer software AntConc. It is a freeware concordance program developed by Laurence Anthony and offers a variety of corpus query tools which are useful to process the corpora. AntConc is found online at: https://www.laurenceanthony.net/software.html. In order to sketch an overall picture of the frequency of the epistemic adverbs: *perhaps*, *maybe*, *probably* and *possibly* used in the two corpora, **word list** was used to count all the words in the corpus.

Following this, I investigated the middle and initial-end positions of the four epistemic adverbs, i.e. the adverbs preceding and following the verbs. For this procedure, I used AntConc which provides a better illustration of the positions of the four epistemic adverbs. To identify the positions of the four epistemic adverbs next to the verbs, I used the **concordance plot tool**. For instance, I searched for the epistemic adverb *maybe*, and I limited the search to one step to the left and right. In other words, only the words that occur just next to *maybe* were highlighted and thus allowed a better visualization

of the occurrence of *maybe* next to verbs. I did the same procedure for the three remaining adverbs which made the calculation of the number of adverbs that precede and follow the verbs easier.

To answer the third research question, I have selected four among the most common verbs used in English which are **to be**, **think**, **know**, **have**, and examined their occurrences next to the four epistemic adverbs under scrutiny. The verbs include all the forms including (is, 's, was', were, etc.). To be able to do so, I have selected *probably*, *maybe*, *possibly*, and *perhaps* using AntConc and I have examined all the verbs that co-occur with them. The list of verbs was long; however, I only wrote down the occurrences of the chosen verbs next to the four adverbs. For instance, I searched for *possibly* on AntConc and looked for its occurrence (one step to the left and right) with verbs **to be**, **think**, **know**, and **have**. As a result, I ended up with the sum of the number of instances of each of the four verbs next to *perhaps*, *maybe*, *possibly*, and *probably*.

Lastly, I have answered the fourth question related to gender-based differences in the native and non-native English speech. First, I began the search using MICASE which was much easier than VOICE. In MICASE the search of the gender was done automatically by the search program on the web. MICASE has an option that provides a statistical summary of the corpus. The summary includes, statistics of the academic division, the total number of tokens by academic role and the total number of tokens by gender. Therefore, the frequencies of the use of *perhaps*, *maybe*, *possibly*, and *probably* are presented by gender in MICASE. In VOICE, however, I had to go through all the instances of the four epistemic adverbs and check the gender of the speaker separately. The information concerning the gender of the speaker in VOICE is illustrated in the figure below. By clicking on S4, as circled in red in Figure 4, a small table that includes different information concerning the age, mother tongue, and gender of the speaker appears.

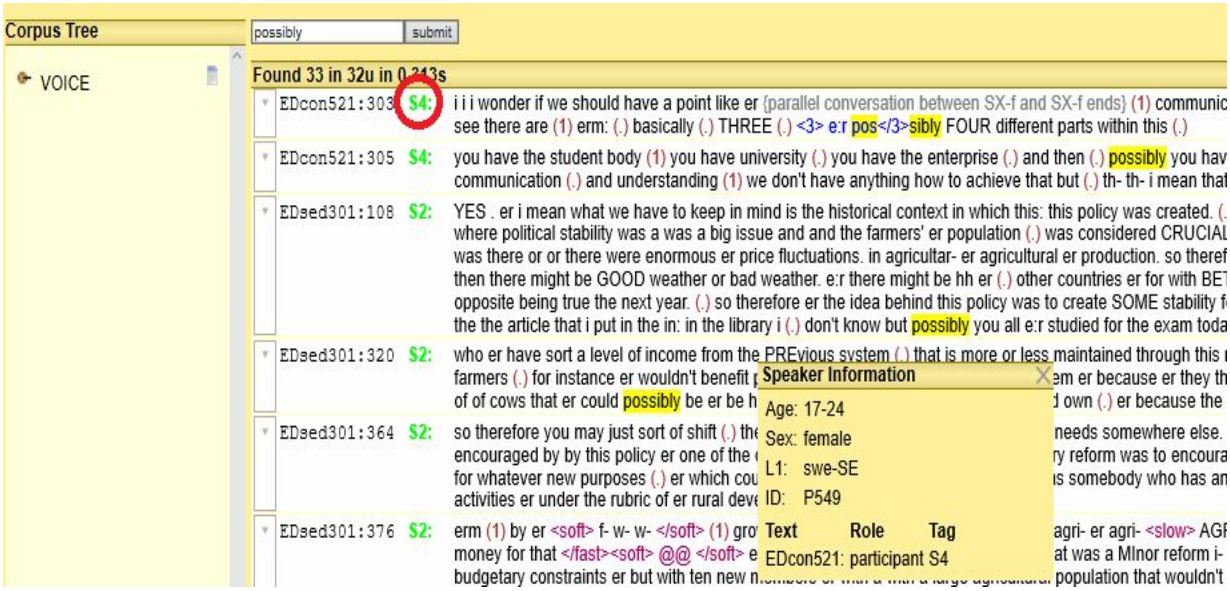

Figure 4: Information concerning the speaker's gender (VOICE)

Following the calculation of the number of occurrences of *perhaps*, *maybe*, *possible*, and *probably* in the two corpora VOICE and MICASE, I had to perform a statistical test to determine whether the difference between the results obtained from the two corpora is considered as statistically significant or not. Therefore, I have used log-likelihood (LL) with the help of a log-likelihood calculator available online: http://ucrel.lancs.ac.uk/llwizard.html (Rayson n.d.). The log-likelihood calculator performs statistical tests automatically, the only two information that are required to fill in the table are the frequency of words and the corpus size. It is important to use the raw frequencies of the epistemic adverbs and the exact size of the corpora. The log-likelihood calculator uses a contingency table (Table 5) to calculate the log-likelihood values.

Table 5. *Contingency table for calculating log-likelihood* (Rayson n.d.).

	Corpus 1	Corpus 2	Total
Frequency of word	a	b	a+b
Frequency of other words	c-a	d-b	c+d-a-b
Total	c	d	c+d

Table 6. *Log-likelihood calculator results for maybe*

Item	O1	%1	O2	%2	LL
Maybe	1538	0.08	0	0.00+	1355.02

O1 refers to the observed frequency in Corpus 1, O2 refers to the observed frequency in Corpus 2. The + indicates overuse in O1 relative to O2, and the - indicates underuse in O1 relative to O2. The significance is evaluated from the LL-value. The higher the LL, the more significant the difference between the frequencies is. The percentiles for significance are as follows:

- 95th percentile; 5% level; $p < 0.05$; critical value = 3.84
- 99th percentile; 1% level; $p < 0.01$; critical value = 6.63
- 99.9th percentile; 0.1% level; $p < 0.001$; critical value = 10.83
- 99.99th percentile; 0.01% level; $p < 0.0001$; critical value = 15.13 (Rayson n.d.).

Table 7 indicates that the log-likelihood values are at $p < 0.05$ for *possibly*, $p < 0.0001$ level for *perhaps* and *probably*, $p < 0.01$ for *maybe*. Therefore, the frequencies for the four epistemic adverbs range between significant and highly significant which indicate that there is a difference in the use of these epistemic adverbs in the two corpora. I have also conducted log-likelihood tests on the remaining research questions to see whether the results are statistically significant in the native and non-native speech. The results for the log-likelihood tests are presented in the next chapter.

Table 7. *Log-likelihood values for perhaps, maybe, possibly, and probably*

	VOICE	***MICASE***	***Log-likelihood***
possibly	*32*	*50*	*- 4.87; p < 0.05*
perhaps	*342*	*93*	*- 212.22; p < 0.0001*

probably	*515*	*606*	*- 461.69; p < 0.0001*
maybe	*1757*	*832*	*+ 9.74; p < 0.01*
total	*2646*	*1581*	*+ 366.53; p < 0.0001*

Another important detail that needed to be considered is the different text lengths of the two compared corpora; therefore, it was important to normalize the frequencies and adjust the raw counts. This procedure is called data normalization which means transforming all variables in the data to a specific range. Taking this measure into consideration is crucial to get an accurate account of the relative frequencies of the epistemic modals in the two corpora. The formula used to normalize the frequency of the four adverbs under scrutiny follows the example provided by Ippolit (2013: 34) which is the following: Frequency per million words (pmw) = number of instances/ number of words X 1,000,000. With (/) stands for division sign, (X) stands for multiplication sign and (=) stands for equals sign.

In this study, the results of the first and fourth research questions, concerning the frequencies of the four adverbs and the gender-based difference, were normalized to 1,000,000 words. In addition, the results were rounded for the sake of clarity. Considering the second and third research questions, that examine the positions of the four epistemic adverbs under scrutiny and their occurrences next to the verbs, the results were normalized to 1,000 words. The normalization of the frequencies was done using The Grammar Lab found online: http://www.thegrammarlab.com/?p=160. This website simplifies the procedure of normalizing the frequencies. The Grammar Lab includes a table where only three information need to be filled in: word count, corpus size and the chosen normalized frequency.

6. Quantitative Analysis

These sections present the quantitative portion of my thesis. These sections are based on the comparison of the native and non-native use of the four epistemic adverbs in MICASE and VOICE. Section 6.1 describes the frequencies of the four epistemic adverbs under scrutiny in the two corpora MICASE and VOICE. Section 6.2 reports the differences and similarities in the use of *perhaps*, *maybe*, *probably*, and *possibly* in the middle and initial-end positions by native and non-native English speakers. Section 6.3 reports on the most common verbs that occur next to the epistemic adverbs in MICASE and VOICE. Finally, section 6.4 describes the gender-based differences in the use of the four epistemic adverbs under scrutiny.

6.1 Frequencies of *perhaps*, *maybe*, *probably*, and *possibly* in VOICE and MICASE

This section presents the overall frequencies of the four epistemic adverbs found in the two corpora VOICE and MICASE after being normalized per 1,000,000 words. The results reported in Table 8 below display the frequencies of the four epistemic adverbs: *perhaps*, *maybe*, *possibly*, and *probably* used in the spoken language by both native and non-native English speakers. The first interesting conclusion reported in Table 8 is that non-native English speakers use the four epistemic adverbs more often than native speakers. In fact, there are 2,646 instances of the four adverbs recorded by non-native speakers compared to 1,581 by native speakers. Another noteworthy conclusion is related to the order of preference of the four epistemic adverbs in MICASE and VOICE. Both native and non-native speakers favor the use of these adverbs in the following order: *maybe*, *probably*, *perhaps*, and *possibly*, with *maybe* being the most frequent epistemic adverb and *possibly* the least frequent. Despite this order of preference, the frequencies of the adverbs differ in the two corpora.

Table 8. *Normalized frequencies of perhaps, maybe, possibly, and probably in MICASE and VOICE*

	VOICE	***MICASE***	***Log-likelihood***
possibly	*32*	*50*	*- 4.87; p < 0.05*
perhaps	*342*	*93*	*- 212.22; p < 0.0001*
probably	*515*	*606*	*- 461.69; p < 0.0001*
maybe	*1757*	*832*	*+ 9.74; p < 0.01*
total	*2646*	*1581*	*+ 366.53; p < 0.0001*

Table 8 presents the total number of the normalized occurrences of *maybe*, *probably*, *perhaps*, and *possibly* in the two corpora investigated VOICE and MICASE. It shows that the epistemic adverbs *maybe* and *perhaps* occur more often in the non-native speech, whereas *possibly* and *probably* occur more in the native speech. This finding suggests that native speakers use longer (more syllables) adverbs compared to non-native speakers. The difference in the use of *perhaps* and *maybe* in the two corpora is significant. The number of the epistemic adverb *perhaps* is three times higher in the non-native speech compared to the native speech. The epistemic adverb *maybe* occurs around twice as more in the non-native speech compared to English native speech.

Table 8 indicates that the log-likelihood values are at $p < 0.05$ for *possibly*, $p < 0.0001$ level for *perhaps* and *probably*, $p < 0.01$ for *maybe.* Therefore, the frequencies for the four epistemic adverbs range between significant and highly significant which indicate that there is a difference in the use of these epistemic adverbs in the native and non-native speech. The examples below demonstrate the use of *maybe* in the native and non-native speech.

(27) *maybe* it's the same with er (.) what we have with russia. (VOICE: EDcon521:1319)

(28) *maybe* something'll come out of it. i don't know. (MICASE:LEL565SU064)

The higher frequency of the epistemic adverb *maybe* in VOICE compared to MICASE could be explained by its property of being highly colloquial as suggested by Greenbaum (1969: 194). Unlike *maybe*, the epistemic adverb *perhaps* seems to be equally suitable in formal and informal varieties (ibid). One potential explanation of the higher use of *maybe* by non-native speakers could be related to the less academic properties of VOICE compared to MICASE. In fact, the latter includes academic speech events recorded at the University of Michigan. However, the interactions recorded in VOICE cover a range of different speech events in terms of domain, including professional, educational, and leisure.

I will consider this hypothesis: the higher use of *maybe* by non-native speakers could be explained by the less academic properties of VOICE compared to MICASE. To test the hypothesis, I restricted the search for *maybe* to the educational domain (ED) which is composed of 260,935 words in VOICE. The results of *maybe* in this restricted educational domain returns 568 instances after normalizing the frequency to 1,000,000 words. According to these findings presented in Table 9, the number of instances of *maybe* has decreased which shows a potential correlation between the domain of use and the number of occurrences of *maybe.* Thus, this result presented in Table 9 below supports the hypothesis which is: the higher use of *maybe* by non-native English speakers could be explained by the less academic properties of VOICE compared to MICASE.

Table 9. *Comparison of the use of maybe in different domains*

Domains	**Number of words**	**Instances of maybe**
MICASE	*1,848,364*	*832*
VOICE	*1,023,082*	*1757*
VOICE (ED)	*260,935*	*568*

6.2 Frequencies of *perhaps*, *maybe*, *probably*, and *possibly* preceding and following verbs in VOICE and MICASE

The previous section showed that the two corpora differ statistically in terms of the number of four epistemic adverbs. Following the calculation of the frequencies of *perhaps*, *maybe*, *possibly*, and *probably* in the native and non-native speech, I have investigated their occurrences in middle and initial-end positions.

To gain a more in-depth understanding of the usage of the epistemic adverbs under scrutiny in these positions, I had to investigate their presence preceding and following the verbs. The middle positions, which occur before the verb, include four different categories which are the following: initial-medial, medial, medial-medial and end-medial (Hoye 1997:148). To have a better conceptualization of the positions, the following examples by Hoye (1997:148) illustrate the positions:

(29) initial-medial They *possibly* may have been sent to London.
(30) medial They may *possibly* have been sent to London.
(31) medial-medial They may have *possibly* been sent to London.
(32) end- medial They may have been *possibly* sent to London.

The second position that I will focus on is the occurrences of the adverbs that follow the verb named initial-end position (ibid). The adverbs in the initial-end position occur directly after the verb. The example below shows the second position that I will focus on which is the initial- end position. In my thesis, the adverb that occurs in the initial-end position will also be referred to as the adverb following the verb as presented in the example below:

(33) initial-end They may have been sent *possibly* to London. (Hoye 1997:148)

The choice to focus on the middle and initial-end positions is grounded on two different reasons. The first motive lies behind Biber et al's (1999: 772) study that supports the overuse of epistemic stance adverbials in the middle positions. In fact, Biber et al. (1999:874) believe that one reason of the high frequency of stance adverbials in the medial position might be explained by the speaker's tendency to soften suggestions in conversations.

The second reason behind the choice to study only the middle positions of the four epistemic adverbs is made after several failed attempts to locate the initial and end positions of the adverbs in the spoken corpora. The challenge manifests in defining the beginning and the end of the sentences in conversations. For instance, to determine if an adverb occur in initial position, the adverb should precede the subject (Hoye 1997: 148). However, in the spoken corpora, the sentences do not necessarily begin with a subject and in some cases, the subject is even missing from the sentence. These examples better illustrate the difficulty I faced to study the initial and end positions:

(34) yeah *maybe* look some</9>thing up on the internet. (VOICE EDwsd242:150)
(35) *maybe* NEXT week tuesday (.) we can do it for. (VOICE: EDwgd497:162)

These two examples above show that the speakers might use the epistemic adverbs in different ways in the spoken corpora. Examples (34) and (35) lack a subject which explains the challenge I faced to study the presence of the adverbs before the subject. Example (35) shows that the boundaries of a sentence in the spoken corpora is not that simple to analyze.

It is difficult to refer the unit of speech as the spoken language cannot rely on the notion of sentence as a reference unit. This is a problem in the analysis of specific speech properties. In fact, several linguists have suggested that in the spoken corpora, it is difficult to consider a sentence as a fully

adequate notion (Raso et al. 2014: 332). In the default to perceive the boundaries of the sentences, I have chosen to focus on the occurrence of *maybe*, *perhaps*, *probably*, and *possibly* preceding and following the verbs.

In this regard, Raso et al. (2014: 332) do not recommend analyzing any form of text that do not anticipate a form of text-sound alignment. In fact, it is considered inadequate for a realistic examination of the structuring of speech. In the same way, they question the boundaries of a sentence in the spoken corpora. Today, different ways are used to better analyze the spoken corpora such as ToBI, tones and break indices, which is a system that help transcribe the intonation patterns and other aspects of the prosody of English utterances (Raso et al. 2014: 334). In other words, ToBI is a set of conventions that transcribes and annotates the prosody of speech. There are symbols used, such as H an L to denote whether the speaker is using a high (H) or low (L) tone. When combined to a star *, H* and L* are associated to a stressed syllable. Other symbols mark the alignment of the tone and detect prosodic structure boundaries (ibid: 196). These different symbols make it easier to identify the boundaries of a sentence and thus the position of a word in the clause becomes detectable.

The focus in this study is on the middle positions, in particular, on the four epistemic adverbs that occur before the verb and on the initial-end position that includes the adverbs that are located following the verb. I will first present normalized frequencies and the log-likelihood values for the occurrence of *maybe*, *perhaps*, *possibly*, and *probably* both before and following the verb as seen in Table 10. The results presented in Table 10 give an overview on the occurrence of *perhaps*, *maybe*, *possibly,* and *probably* next to the verbs in the two corpora. Next, Table 11 presents the occurrences of the four adverbs before the verbs in the native and non-native speech. Lastly, Table 12 presents the results of the occurrences of the four adverbs after the verbs. To have a fair comparison of the counts, the results shown in the three tables below were normalized to 1,000 words.

Table 10. *Normalized frequencies of maybe, perhaps, possibly, and probably next to the verbs in the two corpora*

	MICASE	***VOICE***	***Log-likelihood***
perhaps	*243*	*137*	*+9.26; p < 0.01*
probably	*641*	*447*	*+ 553.35; p< 0.0001*
possibly	*489*	*312*	*+ 2.07; p > 0.05*
maybe	*199*	*200*	*- 0.00; p > 0.05*
Total	*1572*	*1096*	*+ 4.43; p < 0.05*

Table 10 indicates some differences in the occurrence of the four epistemic adverbs next to verbs in MICASE and VOICE. Overall, native speakers use the four epistemic adverbs in higher frequencies next to verbs. The most remarkable differences concern the epistemic adverbs *probably* and *perhaps* that occur in high frequencies in the middle positions in the native speech compared to the non-native speech. The results show that the epistemic adverbs *maybe* and *possibly* do not show a large difference in the native and non-native speech. Examples (36) and (37) illustrate the use of *maybe* in the two corpora

(36) we can *maybe* do that tomorrow (.) but right now it's important to: shortly tell (.) other people what you think. (VOICE: EDwsd303:111)

(37) m, you wanna *maybe* think about that argument and and, *maybe* even, um, and talk about it a little bit. (MICASE: OFC115SU060)

Based on the log-likelihood results presented in Table 10, *perhaps* and *probably* are statistically significant at $p < 0.0001$ for *probably* and at $p < 0.01$ for *perhaps*. These results indicate that the differences in the use of *probably* and *perhaps* in the middle position in the native and non-native speech is significant. Based on the log-likelihood test, *possibly* and *maybe*, are not statistically significant in the two corpora at $p > 0.05$. Example (38) demonstrates the occurrence of *perhaps* next to the verb

in the native speech. Example (39), however, shows the lack of occurrence of *perhaps* in the middle position in the non-native speech.

(38) Without *perhaps* realizing it there's the electron, which are called E-minus it has an electric charge. (MICASE : COL485MX069)

(39) here in malta *perhaps* i think it's even (.) in other countries (.) that our second language was English. (VOICE: EDint328:572)

After observing the occurrences of *maybe*, *perhaps*, *possibly*, and *probably* next to the verbs (before and after the verb). Table 11 presents the differences in the use of the four adverbs before the verbs in the two corpora. The results show that English native speakers use the four epistemic adverbs more frequently before the verbs compared to non-native speakers.

Table 11. *Normalized frequencies of maybe, perhaps, possibly, and probably preceding the verbs in the two corpora*

	MICASE	***VOICE***	***Log-likelihood values***
perhaps	*110*	*100*	*+0.12; p > 0.05*
maybe	*97*	*106*	*- 0.54; p > 0.05*
possibly	*424*	*212*	*+9.57; p < 0.01*
probably	*515*	*288*	*+ 44.83; p< 0.0001*
Total	*1146*	*706*	*+ 3.99; p < 0.05*

As seen in Table 11, the sum of the four epistemic adverbs occurring before the verbs is at $p < 0.05$. Thus, the two corpora are statistically significant when it comes the occurrences of these epistemic adverbs before the verbs. Comparing every epistemic adverb independently in the two corpora shows that the epistemic adverbs *probably* and *possibly* occur before the verbs in higher frequencies in the

English native speech compared to non-native speech. In fact, based on log-likelihood values, *probably* is at $p < 0.0001$ and *possibly* is at $p < 0.01$. These results indicate that the difference between the two corpora is statistically significant concerning the position of *possibly* and *probably* preceding verbs. The difference between the use of *maybe* and *perhaps* in the two corpora, however, is not significant as reflected in the values presented in Table 11. Examples (40)– (44) illustrate the use of *perhaps*, *maybe*, *possibly,* and *probably* before the verbs by native and non-native English speakers:

(40) you *possibly* have a mistake in the day in anyone of these three situations. (MICASE: COL999MX059)

(41) they *probably* go through all the planets yeah. (MICASE: DIS495JU119)

(42) You have Puerto Rican English and you *probably* have Colombian English. although, uh depending on the on the contact situation you could have, you could have dialect leveling in that sense too. (MICASE: STP355MG011)

(43) There *perhaps* is also I think there is some sort of er our declining fertility rates in the sense that we are not spending our entire lives in care. (VOICE: PRpan13:28)

(44) I believe you *maybe* know with e: r with munich? (VOICE: PBmtg300:1582)

Next, Table 12 shows the frequencies of the use of *perhaps*, *maybe*, *possibly*, and *probably* following the verbs in the native and non-native English speech. The sum of the four epistemic adverbs following the verbs show that, overall, non-native speakers use the four epistemic adverbs more often after the verbs compared to native speakers with 450 and 427 instances respectively. On the one hand, the epistemic adverbs *perhaps* and *maybe* occur more frequently following the verbs in the native speech compared to the non-native speech. On the other hand, the epistemic adverbs *possibly* and *probably* occur more frequently in the non-native speech compared to the native speech in initial-end position (following the verb).

Based on the log-likelihood values presented in Table 12, the result of the total occurrences of *perhaps*, *maybe*, *possibly*, and *probably* in the two corpora is statistically highly significant at $p < 0.0001$.

Thus, the conclusion that the non-native English speakers use the four epistemic adverbs more often following the verbs compared to native speakers is correct. However, when comparing every epistemic adverb independently, the log-likelihood values are at $p > 0.05$ which means that the results concerning the comparison of every adverb independently is not statistically significant.

Table 12. *Normalized frequencies of maybe, perhaps, possibly, and probably after the verb in the two corpora*

	MICASE	**VOICE**	***Log-likelihood values***
perhaps	*134*	*106*	*+ 0.77; p > 0.05*
maybe	*102*	*94*	*+ 0.55; p > 0.05*
possibly	*65*	*91*	*- 0.21; p > 0.05*
probably	*126*	*159*	*- 2.89; p > 0.05*
total	*427*	*450*	*- 35.21; p < 0.0001*

Table 12 summarizes the results of the four epistemic adverbs located following the verbs, simply described as the initial-end position by Huddleston and Pullum (2002). The examples (45)– (51) found in MICASE and VOICE below better illustrate the occurrences of the four epistemic adverbs under scrutiny in the initial-end position (following the verb):

(45) So, if we look at the shape of our data that will help us determine *perhaps* what kind of uh aquifer we have. (MICASE: LES205JG124)

(46) We deliver *perhaps* twenty percent of all the teaching we deliver (VOICE: POmtg316:367)

(47) There' s *probably* whole branches of philosophy, that do nothing more than distinguish, soul from mind from self from personality (MICASE: OFC355SU094)

(48) It is *possibly* true and if you guys still have questions in your head, just raise your hand. (MICASE**:** LEL175MU014**)**

(49) To counter- attack *possibly* before, before our general assembly. (VOICE: POmtg403:508)

(50) With the drug you kill *maybe* ninety-nine-point-nine-nine of the cancer cells. (MICASE: LEL175SU106)

(51) I was thinking *maybe* the diversity unity was really useful (VOICE: EDwsd304:1084)

To conclude, the total occurrences of *perhaps*, *maybe*, *possibly*, and *probably* next to verbs is higher in MICASE compared to VOICE. On the one hand, native speakers tend to use the four epistemic adverbs more frequently before the verbs compared to non-native speakers. On the other hand, the non-native speakers tend to use the four epistemic adverbs more frequently in the initial-end position (following the verbs) compared to the native speakers. The two examples below illustrate this conclusion. Example (52) illustrates the occurrence of *probably* before the verb in the native speech. Example (53) illustrates the occurrence of *probably* following the verb in the non-native speech.

(52) so then you *probably* move to the Fleming Building. (MICASE: MTG999ST015)

(53) it all started *probably* when i (.) when i went to the: to the united states for: an exchange. (VOICE: EDsed31:227)

6.3. Verbs that occur next to *perhaps*, *maybe*, *possibly*, and *probably* in VOICE and MICASE

In this section, the aim is to examine the most common verbs that occur next to *perhaps*, *maybe*, *possibly,* and *probably* in MICASE and VOICE. The results show the differences and similarities in the occurrence of the four epistemic adverbs next to verbs in native and non-native English speech. This research question is inspired by Tucker (2001) who carries out a thorough analysis on the epistemic adverb *possibly.* Tucker (2001) focuses solely on the occurrence of the epistemic adverb *possibly* next to the verbs and modal verbs, in my study, however, I will observe the most frequent verbs that occur next to *perhaps*, *maybe*, *possibly* and *probably* in MICASE and VOICE.

Tucker's study is based on 2000 random selected citations from the Bank of English COBUILD corpus (Sinclair 1987). The corpora compromise around 0.4 billion words of running text. According to Tucker's findings the most common verb that co-occur with the epistemic adverb *possibly* is the

verb **to think** (2001:192). The second most common verb co-occurring with *possibly* is the verb **to be**. Examples (54) – (56) illustrate the co-occurrence of the verb **to be** with the epistemic adverb *possibly*.

(54) Dresser. Their collection of Dresser pots **is** *possibly* the world's finest. The walls are lined with Tucker (2001: 193)

(55) Antonio in Twelfth Night because he **is** *possibly* a gay outcast. The logical conclusion is a (Tucker 2001:194)

(56) why you think that is # <M02> I **think** *possibly* low unemployment er er is obviously (Tucker 2001:191)

In addition, Biber et al. (1999: 373) examine the most common lexical verbs that occur in conversations. The results show that among most common lexical verbs that occur in conversation are the following: *get*, *say*, *go*, **know**, **think**. Furthermore, Biber et al. (1999: 428) state that the verbs **to be**, **to have** are commonly used as they have two functions main and auxiliary. Based on Biber et al's. (1999) and Tucker's (2001) conclusions, I will examine the most common verbs that occur next to the epistemic adverbs and which are the following: **to be**, **think**, **know**, **have**. I will examine how the four epistemic adverbs *perhaps*, *maybe*, *possibly,* and *probably* co-occur with these verbs and I will also include all tenses of these chosen verbs. Figures 5 and 6 show the differences in the use of these four verbs next to the epistemic adverbs under scrutiny in VOICE and MICASE respectively. To have a fair comparison of the counts, the results shown in Figure 5 and 6 have been normalized to 1,000 words.

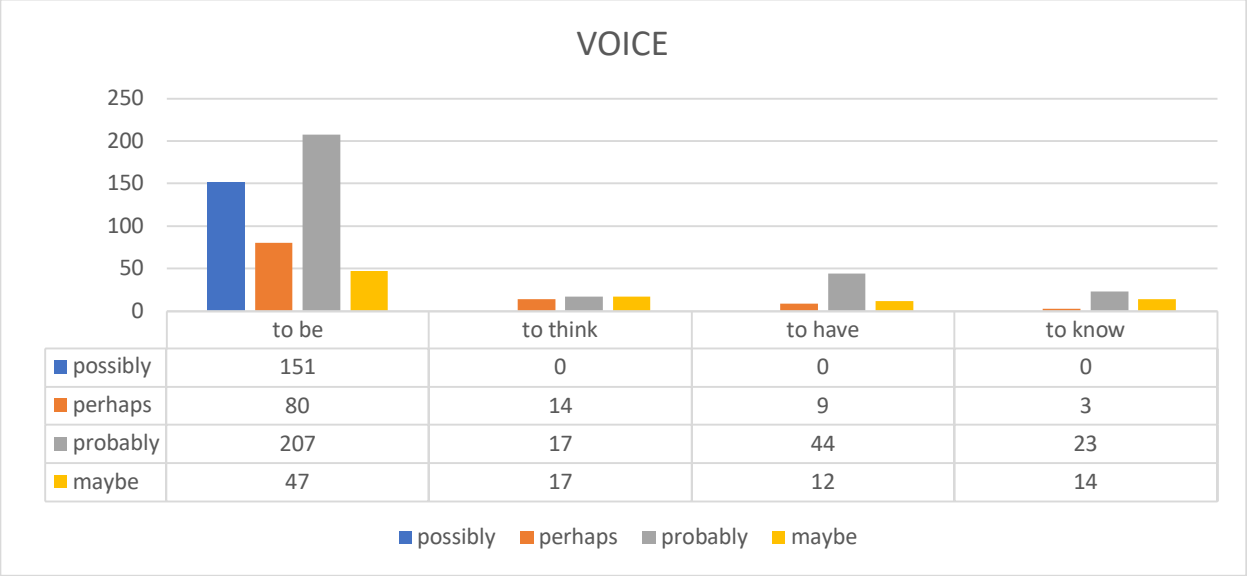

Figure 5. *The occurrences of possibly, perhaps, maybe, and probably next to the verbs in VOICE*

Based on Figure 5, a number of conclusions can be drawn. First, the epistemic adverb *possibly* does not occur with the verbs **to think**, **to have** and **to know** in the non-native English speech. Secondly, *perhaps, maybe, possibly*, and *probably* occur the most with the verb **to be**. In VOICE, *probably* occurs the most next to the four verbs. In many instances in the corpus, the verb **to know** occurs in expressions such as "I don't know" and "you know" especially with the epistemic adverbs *maybe* and *probably*. Examples (57)– (60) illustrate this idea:

(57) we just need to provide a space you **know** *maybe* have some juice in the break. (VOICE: EDwsd464:244)

(58) I **don't know** *maybe* one or two weeks or a month (.) to ask them again. (VOICE: ED-wsd464:244)

(59) I **don't know** *maybe* that (.) such a conference is also a very important step in that direction. (VOICE: PRpan294:197)

(60) hh yeah (.) yeah there is also one other one er er er CHAIN one other (.) er chain which you **know** *probably* from Croatia. (VOICE: PBmtg463:675)

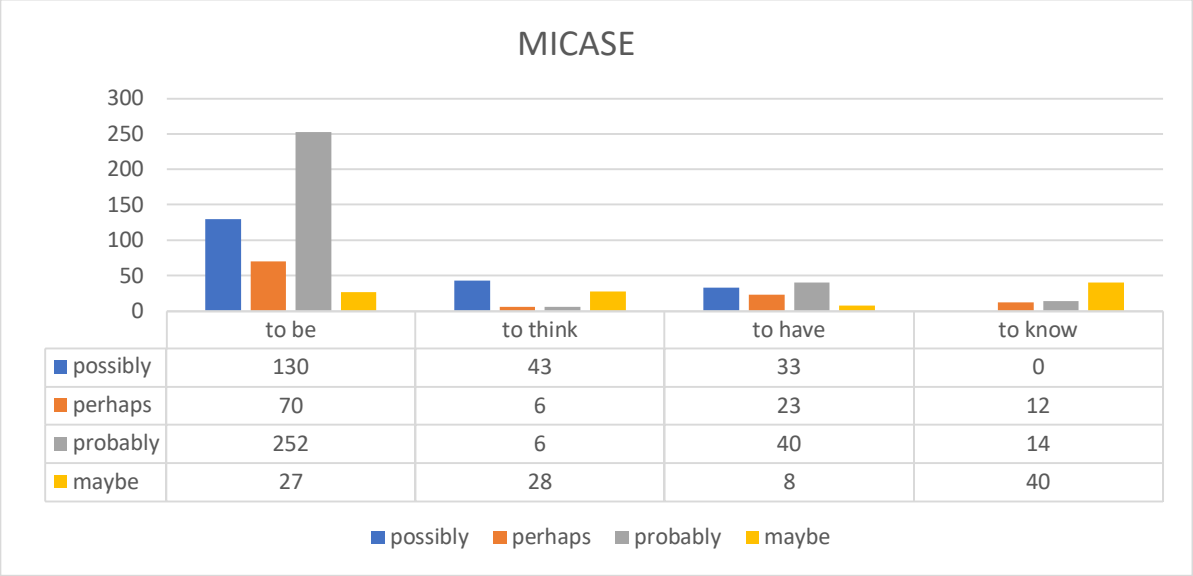

Figure 6. *The occurrences of possibly, perhaps, maybe, and probably following the verbs in MICASE*

Based on Figure 6, a number of conclusions can be drawn. First, *possibly* does not occur with verb **to know** in the native English speech. Secondly, all the epistemic adverbs except *maybe* occur most with **to be**. *Maybe* is the only epistemic adverb that occurs mostly with verb **to know**. In many instances in MICASE, **to know** occurs in expressions such as "I don't know" and "you know" especially with the epistemic *maybe*. The examples (61) and (62) illustrate this idea:

(61) you **know** *maybe*, the only thing that, can really be done at this point is like she needs to really take a look at her internal structure mhm because no matter what, if she can't retain people whether they're interns or volunteer. (MICASE: STP560JG118)

(62) **don't know** *maybe,* like from here to here. and he was always seasick he was just one of those people that never got over being seasick. so you know he was always throwing up and he never could really sleep. (MICASE: LEL175JU154)

Figure 5 and Figure 6 show that both native and non-native speaker use *perhaps*, *maybe*, *possibly*, and *probably* next to verb **to be**. The only exception is with *maybe* that occurs mostly with **to know** in MICASE. In addition, the epistemic adverb *possibly* does not occur with verbs **to think** and **to**

have in VOICE. One similarity between the native and non-native speech is with the epistemic adverb *possibly* that does not occur with **to know**.

Table 13. *Log-likelihood values of the occurrence of the epistemic adverbs and verbs in MICASE and VOICE*

	Log-likelihood values
possibly	+ *42.09; p < 0.0001*
perhaps	+ *84.74; p < 0.0001*
probably	+ *663.55; p < 0.0001*
maybe	+ *111.26; p < 0.0001*

The log-likelihood results show that the total co-occurrence of *perhaps*, *maybe*, *probably,* and *possibly* with the four verbs are statistically significant in the two corpora at $p < 0.0001$. This statistical result suggests that native and non-native English speakers use the four adverbs next to the verbs **to be**, **think**, **know**, **have** differently before and after the verbs in the two corpora.

6.4. Gender-based findings on the use of *probably, perhaps, maybe*, and *possibly* in VOICE and MICASE

In this section, I will examine one social aspect of language which is gender differences using corpus. The aim of this investigation is to find out whether there is a gender-based difference in the use of the four epistemic adverbs, first, in MICASE and then in VOICE. The results of this comparison are illustrated in Tables 14 and 15. Next, I compare the difference in the use of *perhaps*, *maybe*, *possibly*, and *probably* among native and non-native female, and later among native and non-native male. The

results of these comparisons are presented in Tables 16 and 17. The unknown gender category included in Table 14 and 15 is not taken into account due to the low frequencies in this category.

The reason behind shifting the focus to sociolinguistics, is based on the potential answers the findings can provide concerning pragmatics which will be tackled in the discussion section. To have a fair comparison of the values, all the results seen in the tables below are normalized to 1,000,000 words. The normalized frequencies are based on the real number of words in the two corpora. For instance, in MICASE, *maybe* occur in the female speech 922 times. I used the equation below to calculate the normalized frequency of *maybe in* MICASE. With (/) stands for division sign, (X) stands for multiplication sign and (=) stands for equals sign.

Frequency of *maybe* per million words = 922/ 1,848,364 X 1,000,000

$$=498.8 \approx 499$$

Table 14 presents the normalized frequencies of the four epistemic adverbs *perhaps, maybe, probably,* and *possibly* in MICASE and VOICE. A number of comments can be mentioned based on these two tables. First, Table 14 reveals that females use the four epistemic adverbs under scrutiny more frequently than males in MICASE, particularly, *possibly, maybe,* and *probably.* In MICASE, 920 instances of these adverbs are reported to occur in the female speech compared to 661 in male speech. The epistemic adverbs *possibly*, *probably,* and *maybe* occur in higher numbers in the female native speech compared to the male native speech. The only epistemic adverb that appears to be used more frequently in MICASE among male is *perhaps*.

Table 14. *Normalized frequencies of perhaps, maybe, possibly, and probably found in MICASE*

	Male	***Female***	***Unknown***
possibly	*22*	*28*	*0*

perhaps	*52*	*41*	*0*
probably	*254*	*352*	*1*
maybe	*333*	*499*	*0*
total	*661*	*920*	*1*

Next, I will compare the female and male use of the four epistemic adverbs in VOICE. The results presented in Table 15 reveal that the use of these four epistemic adverbs is more common among non-native women. In addition, the epistemic adverbs *perhaps* and *probably* occur in higher numbers in the male speech compared to female speech. *Maybe* and *possibly*, however, are used more commonly by females in VOICE.

Table 15. *Normalized frequencies of perhaps, maybe, possibly, and probably found in VOICE*

	Male	***Female***	***Unknown***
possibly	*13*	*20*	*0*
perhaps	*228*	*130*	*0*
probably	*282*	*233*	*0*
maybe	*761*	*991*	*5*
total	*1,284*	*1,374*	*5*

Based on Tables 14 and 15, there is a difference in the use of the four epistemic adverbs under scrutiny between males and females. Both in MICASE and VOICE, the total values of the four epistemic adverbs are higher among female speakers. The epistemic adverbs *maybe*, *probably*, and *possibly* are used more by females compared to males in MICASE. In VOICE, the epistemic adverbs *possibly*, and *maybe* occur more frequently in the female speech compared to male.

After comparing the occurrences of the four epistemic adverbs in the male and female speech within the same corpora, I have next tried to find out whether there are any differences in the native and non-native female use of these four epistemic adverbs. In other words, I compared the use of *perhaps*, *maybe*, *possibly*, and *probably* between the native and non-native female speakers in MICASE and VOICE. Table 16 presents the frequencies of the four epistemic adverbs used by females. The finding shows that, overall, non-native female speakers use the four adverbs under scrutiny more often than female native speakers.

Table 16. *Normalized frequencies of the female use of the four epistemic adverbs in MICASE and VOICE*

	MICASE	***VOICE***	***Log-likelihood***
possibly	*41*	*13*	*+1.79; p > 0.05*
perhaps	*97*	*223*	*- 61.71; p < 0.0001*
probably	*469*	*289*	*+31.55; p < 0.0001*
maybe	*616*	*779*	*- 225.71; p < 0.0001*
total	*662*	*1275*	*-132.67; p < 0.0001*

After comparing the native and non-native female use of *perhaps*, *maybe*, *probably*, and *possibly*. The results show that, overall, female non-native speakers use more often the four epistemic adverbs than female native speakers. In particular, the epistemic adverbs *possibly* and *probably* occur mostly in the native female speech, while *perhaps* and *maybe* occur mostly in the female non-native speech.

Next, I tried to find out whether there are any differences in the use of the four epistemic adverbs under scrutiny between native and non-native male speech. Table 17 shows that, overall, male non-native speakers use more often the four epistemic adverbs compared to male native speakers. *Possibly*

and *probably* occur more in the native speech among male, whereas *perhaps* and *maybe* occur more among males in the non-native speech. These results and the log-likelihood values are presented in Table 17.

Table 17. *Normalized frequencies of the male use of the four epistemic adverbs in MICASE and VOICE*

	MICASE	***VOICE***	***Log-likelihood***
possibly	*51*	*20*	*+ 3.35; p > 0.05*
perhaps	*75*	*127*	*-153.10; p < 0.0001*
probably	*650*	*238*	*- 2.04; p > 0.05*
maybe	*922*	*1014*	*- 235.77; p < 0.0001*
total	*1223*	*1277*	*- 248.65 p < 0.0001*

Based on Tables 16 and 17 a number of comments can be mentioned. First, both native male and female speakers use *possibly* and *probably* more frequently. In addition, both non-native male and female speakers use *perhaps* and *maybe* more frequently. These findings support the results seen in section 6.1 that indicate that native speaker tend to use the epistemic adverbs *possibly* and *probably* more often than non-native speakers. The non-native speakers, however, use more often the epistemic adverbs *maybe* and *perhaps.*

Based on the log-likelihood values presented in Tables 16 and 17, there is a significant difference between (native and non-native) females, and between (native and non-native) males in the use of the four adverbs. In fact, the total log-likelihood values in the two tables is at $p < 0.0001$. Only the epistemic adverb *possibly* in Table 16, and *possibly* and *probably* in table 17 are at $p > 0.05$. On this

account, it is possible to suggest that, overall, non-native males and females use *perhaps*, *maybe*, *possibly*, and *probably* more often than native females and males**.**

In this section, first, I examined the gender-based difference in the use of the four adverbs in MICASE. The results show that native women use the four epistemic adverbs more frequently than native men. Next, I examined the gender-based difference in the use of the four adverbs in VOICE. The results show that non-native women use the four epistemic adverbs more often than non-native men. Accordingly, both native and non-native women use the four epistemic adverbs more frequently than native and non-native men. It is worth mentioning that this gender-based difference is more significant in MICASE compared to VOICE. After observing the gender-based difference in the two corpora. I proceeded by comparing (native and non-native) women in the two corpora and (native and non-native) men in the two corpora. The results indicate that, overall, native speakers tend to use *possibly* and *probably* more often than non-native speakers. The non-native English speakers, however, use more often *maybe* and *perhaps.*

7. Discussion and conclusion

This chapter summarizes the thesis and adds a few reflections to this study. The aim of this comparative study is to investigate the frequencies and the positions of the epistemic adverbs, *perhaps*, *maybe*, *probably,* and *possibly* between native and non-native speakers. In addition, I examined the most common verbs that occur next to the four epistemic adverbs under scrutiny in MICASE and VOICE. Lastly, I observed the gender-based differences concerning the use of the four epistemic adverbs in the two corpora. The general aim of this thesis is to examine the similarities and differences in the usage of these adverbs in VOICE, which represents the non-native speech data, and MICASE, which represents the native speech data.

To begin the discussion, it is worth mentioning a well-known remark concerning the adverbs stated by Chomsky's (1965) which is the following: "adverbials are a rich and as yet relatively unexplored system, and therefore anything we say about them must be regarded as quite tentative." (quoted here in Austin et al. 2004:1). The adverbs of stance have been marginally addressed in the literature and my study is an attempt to provide an overview of the way *perhaps*, *maybe*, *possibly*, and *probably* occur in the native and non-native English speech.

First, I have studied the frequencies of the four adverbs in MICASE and VOICE. The quantitative findings indicate that overall, *maybe*, *perhaps*, *probably*, and *possibly* occur more often in the non-native speech compared to the native speech. The non-native speakers use English as their second or foreign language, thus the more frequent use of the four epistemic adverbs under scrutiny in their speech could mirror their uncertainty when speaking compared to the native speakers. The results

show that both native and non-native speakers favor the use of the adverbs under scrutiny in the same order. The order of this preference is the following: *maybe* is the most common used adverb followed by *probably*, *perhaps*, and finally *possibly*. One potential explanation of the highest use of *maybe* could be a matter of a slight differences in the strength of the epistemic evaluation. In fact, *maybe* is considered stronger than *possibly* in the epistemic evaluation (Nuyts 2001: 56). Despite the same order of preference of these four epistemic adverbs in the native and non-native speech, their number of occurrences in the two corpora differs. The results show that, on the one hand, non-native speakers use *maybe* and *perhaps* more frequently compared to the native speakers. On the other hand, English native speakers use *probably* and *possibly* more frequently than the non-native speakers. It is noteworthy to remark that English native speakers use longer (more syllables) adverbs compared to the non-native speakers.

Once the whole picture on the frequencies of *perhaps*, *maybe*, *possibly*, and *probably* is examined between native and non-native speakers. The study broadens to investigate the position of the four epistemic adverbs in a clause. In fact, the position of the adverbs varies within a single language. The same adverb may occur in different positions within the same language. The aim of this investigation is to account for this variation. The positions chosen in this thesis are the middle and the initial-end which refer to the epistemic adverbs that occur preceding and following the verbs. By doing so, it is possible to detect the similarities and differences in the use of the four epistemic adverbs between the native and non-native speakers of English.

The examination of the occurrence of the four adverbs preceding and following the verbs in the two corpora, MICASE and VOICE, has revealed some differences. Overall, English native speakers tend to use the epistemic adverbs under scrutiny more frequently before the verbs and verb auxiliaries.

The non-native speakers, on the other hand, tend to use the epistemic adverbs in higher frequencies following the verbs in conversations. These conclusions are based on the comparison of the total number of occurrences of the four adverbs in the two corpora before and after the verbs. The results of the use of *perhaps*, *maybe*, *possibly* and *probably* independently show the following: First, the epistemic adverbs, *probably*, *possibly*, and *perhaps* occur in the native speech more often in the middle position (before the verbs) compared to the non-native speakers. Secondly, the native speakers favor the use of *perhaps* and *maybe* in the initial-end position (following the verbs) more often than the non-native speakers.

When examining the distribution and positions of *perhaps*, *maybe*, possibly, and *probably*, it is important to consider the pragmatics, i.e., to consider their usage and the contexts in which they are used. Lakoff (1990:6) emphasizes the significance of pragmatics in language studies and says the following:

> People use language to achieve certain aims, the broad topic of the politics of language provides a way of investigating how the forms of language facilitate its function, and how they are created to serve functions. Until we have ways to ask and investigate those questions, I don't think we will have much understanding of language, no matter how many formal rules we produce, how many recorded conversations we transcribe

Pragmatics sees language as communication, i.e., in terms of functions (Lakoff 1990:28). Through pragmatics, it is possible to explain the speaker's intention and the hearer's understanding (ibid). In fact, the occurrence and placement of adverbs cannot be explained by syntactic and semantic reasons

alone. Engel (2012: 325) states that the placement of the adverbs in English could be affected by a variety of factors such as syntactic, semantic as well as information-structural considerations. According to Hoye (1997:212), when adverbs are used in initial position, it renders the speaker's assertion. In this regard, it is argued that the epistemic adverbs placed in initial position, prior to the subject, can even be used in a manipulative sense "to seduce the addressee into believing the content of the proposition" (Hoye 1997: 213). The example below better illustrates this idea:

(63) *maybe* you should footnote that. (MICASE: SEM300MU100)

When adverbs are used in the end of the clause, they signal the speaker's presence as seen in the examples below:

(64) okay i'm going to (.) to read (a bit) *probably* @@ (.) (VOICE: EDcon4:276)
(65) and initial bitterness units *maybe*? (MICASE: SGR195SU127)

The occurrence of the adverbs in the middle positions (before and after the verbs), puts emphasis on the verb. It could also add emotional overtones to make the propositional content sound more acceptable (ibid: 214).

(66) we'd *probably* pick estimator three even though it's a little bit biased. (MICASE: LAB575JU095)
(67) I think *probably* the people up here have a dog, and it comes over here and runs. (MICASE: LAB175SU032)

The findings in section 6.2 show that English native speakers use *perhaps*, *maybe*, *possibly*, and *probably* in the middle position (before the verb) more often than the non-native speakers. Considering Hoye's (1997:213) suggestions, this result has two possible interpretations. First, English native speakers tend to use the four epistemic adverbs before the verbs to put emphasis on the verbs more often than the non-native speakers. The second interpretation could be that English native speakers tend to use the four epistemic adverbs before the verbs with the intention to soften their propositions

more often than the non-native speakers. A definite conclusion concerning the occurrence of the epistemic adverbs in the middle position is difficult to state. Analyzing the initial and end positions, in this case, would have given a better explanation of the results.

Next, I have examined the occurrence of the four epistemic adverbs next to verb **to be**, **to have**, **to know**, and **to think** between native and non-native speakers of English. The choice of these verbs is based on the findings of Biber et al. (1999) and Tucker (2001) that suggest that these verbs are among the most common. The findings in my study indicate that both the native and the non-native speakers of English use *perhaps*, *maybe*, *possibly*, and *probably* the highest next to the verb **to be**. The only exception is with the epistemic adverb *maybe* that occur more frequently with the verb **to know** in the native speech. Particularly, expressions such as "you know" and "I don't know" occur with *maybe*. If I go back to the correlation between the occurrence of the adverbs in the middle position and the verb as suggested by Hoye (1997: 213). It possible to think that the high frequency of *maybe* next to verb **to know** put emphasis on the state of "knowledge" or "limited knowledge" of the speaker. One similarity between the native and non-native speech concerns the epistemic adverb *possibly* that does not occur with **to know**.

Lastly, after examining the frequencies, positions, and environments of *perhaps*, *maybe*, *possibly*, and *probably*, I have shifted the focus to another area of linguistic which is sociolinguistics. I compared the way the four epistemic adverbs are used by male and female speakers. The aim was to examine if there are any gender-based differences which can be useful to understand language in use and the contexts in which it is used.

It is important to understand how languages evolve, thus, to keep examining the same questions with a new perspective. The examination of the gender-based differences in the two corpora show some differences in the use of the four epistemic adverbs. The findings indicate that female speakers use the four epistemic adverbs relatively more often than male in MICASE and VOICE. Only the epistemic adverb *perhaps* in VOICE, and *probably* and *perhaps* in MICASE are favored more by males compared to females. In addition, I have examined the use of the four adverbs under scrutiny between female in the two corpora, and male in the two corpora. The results show both male and female non-native speakers use the four epistemic adverbs more often than the native speakers. The higher use of *perhaps*, *maybe*, *possibly*, and *probably* among women in both native and non-native speech could have two different interpretations. The first reading is proposed by Lakoff and Bucholtz (2004: 79) who suggests that women are unassertive compared to men in their speech. The second interpretation is proposed by Facchinetti (2012 :346) who suggests that women might be more aware of the different functions of epistemic modal forms in their speech because of their sensitivity to interpersonal aspects of talk. It is difficult to give a clear-cut to the question concerning the higher use of the epistemic adverbs among women. It is worth mentioning, however, that this gender-based difference is more significant in MICASE.

From my point of view, the quantitative methods were suitable for this thesis. Calculating the normalized frequencies adds more accuracy to the study than if the raw frequencies were used. Furthermore, calculating the log-likelihood values of the epistemic adverbs under scrutiny made it possible to assess the significant difference in frequency between the native and non-native speech, and thus to obtain reliable answers to the research questions.

As I move forward to discussing the methodology, I believe that the methods used were suitable for the first and fourth questions. However, concerning the second and third research questions, I have faced some challenges. When dealing with spoken corpora, some challenges arises concerning the position of the epistemic adverbs in sentence. Unlike the written language, where it is possible to detect the syntactic analysis of a sentence, analyzing the spoken language raises some issues. To spot the sentence in the written language capitalized first words, commas, and full stops help the researchers conduct their analyses. However, as the data analyzed in my study consists of spoken language, to detect the beginning and the end of a clause is challenging. My first plan was to study three positions of these adverbs, initial, middle and end. When I examined the spoken corpora to begin the search, I noticed it was not possible to detect the boundaries of the sentences which lead me to focus only on the middle and initial-end positions.

After summarizing the results and adding some reflections based on the theory and the findings, I will now give my opinions concerning the future research. I see possible ways to carry on the research about this topic that has plenty to offer for English Linguistics. In my study, I focused on *perhaps*, *maybe*, *possibly*, and *probably* and their occurrences in the middle position (before the verb) end initial-end position (after the verb). This research could be expanded by including the initial and end positions with the help of systems such as ToBi to make it easier to detect the boundaries of the clause in the spoken language. In addition, this study is a comparison of four epistemic adverbs in American native speech and non-native English speech. It would be interesting to include another English variety such as British English and examine their differences and similarities.

References

Aarts, Bas. 2011. *The Oxford Modern English Grammar.* New York: Oxford University Press.

Alvanoudi, Angeliki. 2014. *Grammatical Gender in Interaction: Cultural and Cognitive Aspects.* Boston: Brill

Austin, Jennifer R, Engelberg, Stefan, Rauh, Gisa. 2004. *Adverbials: the interplay between meaning, context, and syntactic structure.* Amsterdam; Philadelphia: John Benjamin publication.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.

Brown, David. (n.d.). *The grammar Lab*. http://www.thegrammarlab.com/?p=160

Chomsky, N. 1965. *Aspects of the Theory of Syntax.* Cambridge: MIT Press.

Collins, Peter. 2009. *Modals and Quasi-modals in English*. Amsterdam: Editions Rodopi.

Crystal, David. 2012. *English As a Global Language*. Cambridge: Cambridge University Press.

Davis, Alan. 2003. *The Native speaker: Myth and Reality*. Clevedon: Multilingual Matters.

Deterding, David. 2013. *Misunderstandings in English As a Lingua Franca: An Analysis of ELF Interactions in South-East Asia*. Boston. Berlin: De Gruyter.

Du Bois, John W., Wallace L. Chafe, Charles Meyer, Sandra A. Thompson, Robert Englebretson, and Nii Martey. 2000-2005. *Santa Barbara corpus of spoken American English*, Parts 1-4. Philadelphia: Linguistic Data Consortium.

Eckert, Penelope, and McConnell-Ginet, Sally. 2003. Language and Gender. New York: Cambridge

University Press.

Esimaje, Alexandra; Gut, Ulrike; Antia, Bassey Edem. 2019. *Corpus Linguistics and African Englishes*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Facchinetti, Roberta; Krug, Manfred; Palmer, Frank. 2012. Berlin; Boston: *Modality in Contemporary English*. De Gruyter Mouton.

Gonzalez-Marquez, Monica. 2007. *Methods in Cognitive Linguistics*. Amsterdam; Philadelphia: John Benjamins Publication.

Greenbaum, Sidney.1969. *Studies in English adverbial usage*. London: Longman.

Hoye, Leo. 1997. *Adverbs and modality in English*. London & New York: Longman.

Huddleston, Rodney, Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language.* Cambridge University Press.

Ippolit, Margherita. 2013. *Simplification, Explicitation and Normalization: Corpus-Based Research Into English to Italian Translations of Children's Classics*. UK: Cambridge Scholars Publishing.

Jannedy, Stefanie; Bod, Rens; Hay, Jennifer. 2003. *Probabilistic Linguistics*. Cambridge: MIT Press.

Jenkins, Jennifer. 2002. 'A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language'. *Applied Linguistics*, 83–103

Jenkins, Jennifer; Baker, will; Dewey, Martin. 2017. *The Routledge Handbook of English As a Lingua Franca.* London: Routledge.

Kennedy, Graeme. 1998. *An introduction to corpus linguistics.* London: Longman.

Kärkkäinen, Elise. 2003. *Epistemic stance in English conversation: a description of its interactional functions, with a focus on I think.* Amsterdam; Philadelphia: John Benjamins Publication.

Labov, William. 1966. *The social stratification of English in New York City.* Washington, DC

Lakoff, George. 1973. *Adverbs and Modal Operators; Adverbs and opacity: a reply to Stalnaker; Counterparts, or the problem of reference in transformational grammar.* Bloomington: IN

Lakoff, Robin Tolmach. 1990. *The Politics of Language.* the University of Michigan: Basic Books.

Lakoff, Robin Tolmach, and Bucholtz, Mary. 2004. *Language and Woman's Place: Text and Commentaries. Oxford:* Oxford University Press.

Leech, Geoffrey Neil. 1971. *Meaning and the English Verb.* London: Longman.

Lindquist, Hans, and Magnus, Levin. 2018. *Corpus Linguistics and the Description of English.* Edinburgh: Edinburgh University Press.

Mcelhinny, Bonnie; Hols, Marijke; Holtzkener, Jeff; Unger, Susanne; Hicks, Claire.2003. 'Gender, Publication and Citation in Sociolinguistics and Linguistic Anthropology: The Construction of a Scholarly Canon'*, Language in Society*, 299 –328.

McEnery, Tony, Xioa, Richard, Tono, Yukio. 2006. *Corpus-Based Language Studies.* London; New York: Routledge.

Mesthrie, Rajend. 2011. *The Cambridge Handbook of Sociolinguistics.* Cambridge: Cambridge University Press.

Newman, Matthew L. Carla J. Groom, Lori D. Handelman &James W. Pennebaker. 2008. 'Gender Differences in Language Use: An Analysis of 14,000 Text Samples'*, Discourse Processes,*

211 –36.

Nuyts, Jan. 2001. *Epistemic Modality, Language and Conceptualization: A Cognitive-Pragmatic Perspective.* Amsterdam; Philadelphia: John Benjamins Publishing Company.

Palmer. Frank R .1990. *Modality and the English Modals. London:* Longman.

Palmer, Frank R. 1979. *Modality and the English Modals.* London: Longman.

Palmer, Frank R; Roberta, Facchinetti; Manfred, Krug. 2003. *Modality in Contemporary English.* Berlin : Mouton de Gruyter.

Peterson, Elizabeth. 2019. *Making Sense of Bad English: An Introduction to Language Attitudes and Ideologies*. New York: Routledge.

Prodromou, Luke. 2008. *English As a Lingua Franca: A Corpus-based Analysis.* London; New York: Oxford University Press.

Ranta, Elina. 2019. 'The 'Attractive' Progressive – Why use the -ing Form in English as a Lingua Franca?' *Nordic Journal of English Studies*, 95.

Raso, Tommaso; Mello, Heliana; Berserik, Françoise. 2014. *Spoken corpora and linguistic studies.* Amsterdam, Netherlands; Philadelphia, Pennsylvania: John Benjamins Publishing Company.

Rayson, Paul. *(n.d.). Log-likelihood and Effect Size Calculator.* Lancaster University. *http://ucrel.lancs.ac.uk/llwizard.html*.

Seidlhofer, Barbara. 2011.*Understanding English as a Lingua Franca.* United Kingdom*:* Oxford University Press.

Seidlhofer, Barbara. 2004. *Research perspectives on teaching English as a lingua franca.* United

Kingdom: Oxford University Press.

Seidlhofer, Barbara. 2005. *English as a lingua franca.* United Kingdom: Oxford University Press.

Seidlhofer, B; Researchers: Angelika Breiteneder, Theresa Klimpfinger, Stefan Majewski, Ruth Osimk-Teasdale, Marie-Luise Pitzl, Michael Radeka. 2013. *The Vienna-Oxford International Corpus of English* [version 2.0 online]. Available at: http://voice.univie.ac.at./

Simpson, R. C., S. L. Briggs, J. Ovens, and J. M. Swales. 2002. *The Michigan Corpus of Academic Spoken English. Ann Arbor, MI:* The Regents of the University of Michigan. Available at: https://quod.lib.umich.edu/m/micase/.

Sinclair, John. 1987. Collins Birmingham University International Language Database. *The Bank of English COBUILD corpus*. Available at: https://collins.co.uk/pages/elt-cobuild-reference-the-collins-corpus

Sunderland, Jane. 2006. *Language and Gender.* Abingdon; New York: Routledge

Tagliamonte, Sali**.** 2006. *Analyzing sociolinguistic variation*. UK: Cambridge University Press.

Tucker, Gordon. 2001. 'Possibly alternative modality'. *Functions of Language*, 183 – 215.

Weatherall, Ann. 2002. *Gender, Language and Discourse.* England: Routledge.