

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
ADRIANA COUTO RAMOS

**The use of certainty adverbs in Brazilian Portuguese and American English:  
a Semantic/Pragmatic approach**

BELO HORIZONTE  
2015

Adriana Couto Ramos

**The use of certainty adverbs in Brazilian Portuguese and American English:  
a Semantic/Pragmatic approach**

Dissertação apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Mestre em Estudos Linguísticos.

Área de Concentração: Linguística Teórica e Descritiva  
Linha de Pesquisa: Estudos Linguísticos baseados em Corpora  
Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Heliana Ribeiro de Mello

Belo Horizonte  
Faculdade de Letras da UFMG  
2015

*à memória da minha mãe, Judit Couto,  
que me ensinou a ter fé, força, e coragem, nessa ordem.*

## ACKNOWLEDGEMENTS

À minha orientadora Prof<sup>a</sup> Heliana Mello, por saber muito sobre tudo. Obrigada pela orientação que vai além da acadêmica.

Ao Prof. Tommaso Raso, pela confiança apesar das minhas raízes fluminenses.

Aos membros da Banca Examinadora, por aceitarem o convite para leitura e discussão deste trabalho.

À CAPES pelo suporte financeiro durante o mestrado e ao Programa de Pós- Graduação em Linguística da UFMG, especialmente à Prof<sup>a</sup>. Emília Mendes, coordenadora do PosLin, tão prestativa e solícita.

Aos colegas e amigos do LEEL, pelas trocas de sempre.

Ao Fred, pelo pessimismo e pelo domingo perdido me ajudando.

To Sarita, Max and Billiam, who bored themselves proofreading my English.

À Gabriela, prima e fisioterapeuta em formação, que desfez muitos nós nos últimos dois anos.

À minha irmã Barbara, que me faz rir todos os dias, mesmo na pior das situações. Obrigada por segurar minha barra no momento mais difícil das nossas vidas.

Agradeço, sobretudo, à mamãe e papai, de quem eu tive a sorte de nascer filha. Obrigada por me ensinarem a explorar e descobrir, sempre com “juízo”. À mamãe, por ter sido simplesmente a melhor de todas. E ao papai, obrigada por trazer suco enquanto eu trabalhava, por esse sorriso que eu herdei e por ser o melhor amigo que eu pudesse ter. Espero ainda poder te dar muitas alegrias.

*“Tout ce qui vaut la peine est quasi impossible.”*  
Leo Apostel

## ABSTRACT

‘Adverbs of certainty are multifunctional and have social meanings on different dimensions.’ (Simon-Vandenberghe; Aijmer, 2007). This suggests that adverbs assume a polipragmatic behaviour and can hold, amongst others, the semantic function of modalizers as well as the pragmatic function of politeness markers. In this dissertation, based on data extracted from two comparable spontaneous speech corpora, the main objective is to map and describe the semantic and pragmatic uses of adverbs and adverbial expressions of certainty in both Brazilian Portuguese (BP) and American English (AE). Our main research questions are: 1. how are adverbs of certainty employed in both languages, 2. to what extent does sociocultural variation determine both type and frequency of the indexes, and 3. is there a clear boundary between the semantic and pragmatic content of a certain index? In order to meet the preceding questions, this paper contemplates two main parts: the first one dedicated to the adaptation of a subcorpus of AE to make it comparable to the already existing BP one, and a second part that contemplates a qualitative study of certainty adverbs in such languages. The BP data was extracted from the C-ORAL-BRASIL corpus and the AE data from the Santa Barbara Corpus of Spoken American English. It became clear that an accurate picture of adverbs of certainty can only be achieved in a heteroglossic perspective, associated with type of context, social roles and power relations. The analysis shows important differences in the use of adverbs of certainty across both languages, especially when it comes to the use of equivalent adverbs, which are used for different purposes in each language. BP speakers use 1.6x more adverbs of certainty as modal and/or politeness markers than AE speakers, and that may indicate that especially in BP, adverbs of certainty are not merely epistemic markers but also indexes of the speaker’s stance, power and status. Moreover, the analysis demonstrates an upward curve representing an increased use of modal adverbs in lower diastaty in BP if compared to higher ones, which may indicate socioculturally based differences in the expression of politeness in the two groups. We hope the contrastive analysis of the data will help establish a network of relations and meanings between the semantic and pragmatic use of certainty adverbs, and hopefully contribute to the investigation of modal choices with a concern with their rhetorical effects.

*Keywords: modality, politeness, adverbs of certainty, spoken corpora, Semantics, Pragmatics*

## RESUMO

“Advérbios de certeza são multifuncionais e possuem razões sociais em diferentes dimensões.” (SIMON-VANDENBERGEN; AIJMER, 2007). Isso sugere que os advérbios de certeza possuem um comportamento polipragmático e podem adquirir diversas funções no discurso, dentre elas a função semântica de marcar modalidade e a função pragmática de marcar a polidez linguística. O principal objetivo dessa dissertação, baseada em corpora de fala espontânea, é mapear e descrever o uso semântico (como modalizadores) e pragmático (como estratégias de polidez) dos advérbios e locuções adverbiais de certeza em português brasileiro (PB) e inglês americano (IA). As principais questões exploradas nesta pesquisa serão 1. de que forma os advérbios de certeza são empregados nas línguas estudadas, 2. até que ponto a variação diastrática determina o tipo e frequência destes índices modais entre falantes, e 3. existe uma fronteira clara e delimitada entre o uso semântico e pragmático de um determinado índice? Para cumprir o objetivo de descrever o comportamento dos advérbios de certeza na expressão da modalidade e polidez em português e inglês, o presente trabalho contempla duas etapas principais. A primeira se dedica à construção do subcorpus de inglês comparável àquele já disponível para o português brasileiro, e a segunda parte se ocupa com a descrição semântico-pragmática dos advérbios de certeza. Os dados analisados foram extraídos dos corpora C-ORAL-BRASIL e Santa Barbara Corpus of Spoken American English, que contêm amostras comparáveis de fala espontânea do português brasileiro e inglês americano, respectivamente. Durante a análise ficou claro que uma descrição cuidadosa dos advérbios de certeza precisa considerar o tipo de contexto, as regras sociais e as relações de poder da interação, numa perspectiva heteroglóssica. A análise demonstrou importantes diferenças no uso de advérbios de certeza entre as línguas estudadas. Os falantes de BP usam 1,6 vezes mais advérbios e locuções adverbiais de certeza como marcadores modais/de polidez que os falantes de IA, o que sugere que, especialmente em BP, esses advérbios vão além de meros marcadores epistêmicos, e se tornam índices da postura social do falante. Ademais, a análise demonstrou que falantes de menor escolaridade em PB usam mais advérbios modalizadores de certeza do que falantes mais escolarizados, o que pode indicar diferenças socioculturais na expressão da modalidade e polidez nos dois grupos. Espera-se que a análise contrastiva dos dados permita estabelecer uma rede de relações e significados entre o uso semântico e pragmático dos advérbios de certeza, contribuindo para a investigação do comportamento das escolhas modais dos falantes em conjunto com seus possíveis efeitos retóricos.

*Palavras-chave: modalidade, polidez, advérbios de certeza, corpus oral, Semântica, Pragmática*

## LIST OF IMAGES

Image 4.1: Words per gender .....	42
Image 4.2: Uttered words per age group .....	43
Image 4.3: Uttered words per schooling level .....	43
Image 7.1.: Chart of strategies: positive politeness .....	69
Image 7.2: Chart of strategies: negative politeness .....	70
Image 9.1: Frequency of mesmo per text genre .....	84
Image 9.2: Frequency of claro per text genre .....	89
Image 9.3: Frequency of com certeza per text genre .....	93
Image 9.4: Frequency of exatamente per text genre.....	95



## LIST OF TABLES

Table 2.1: Prosodic X Informational Patterns .....	25
Table 4.1: Description of acoustic quality tags .....	33
Table 4.2: Informational Tagging .....	40
Table 4.3: Speaker's age and schooling classification .....	41
Table 4.4: Header .....	42
Table 5.1: The distribution of the adverbs of certainty in the corpora .....	49
Table 6.1: Propositional modes .....	56
Table 7.1: Corpus based reference table of speech acts classes and types .....	63
Table 7.2: PP Maxims .....	67
Table 8.1: Total Frequency of the adverbs .....	76
Table 8.2: Frequency of the adverbs as modalizers/politeness markers .....	80
Table 9.1: Position of mesmo within the utterance .....	82
Table 9.2: Position of claro within the utterance .....	87
Table 9.3: Position of com certeza within the utterance .....	91
Table 9.4: Position of of course within the utterance .....	96
Table 9.5: Position of sure within the utterance .....	98
Table 9.6: Feature analysis of the central adverbs of certainty .....	104
Table 9.7: Communicative Context BP .....	109
Table 9.8: Communicative Context AE .....	109
Table 9.9: Usage per schooling level.....	111

## TABLE OF CONTENTS

<b>1. INTRODUCTION .....</b>	<b>13</b>
1.1 Part I: The creation of the Santa Barbara Corpus of Spoken American English minicorpus .....	15
1.2 Part II: A Semantic/Pragmatic approach to the use of adverbs of certainty .....	16
1.3 Objectives .....	16
1.4 Organization .....	17
<b>PART I:</b>	
<b>The creation of the Santa Barbara Corpus of Spoken American English minicorpus ....</b>	<b>19</b>
<b>2. THE LANGUAGE INTO ACT THEORY .....</b>	<b>22</b>
<b>3. THE C-ORAL-BRASIL .....</b>	<b>28</b>
<b>4. SANTA BARBARA CORPUS OF SPOKEN AMERICAN ENGLISH.....</b>	<b>31</b>
4.1 The process of adaptation .....	32
4.1.1 Choosing the files for the minicorpus .....	33
4.1.2 The transcription .....	34
4.1.2.1 Prosodic Units .....	35
4.1.2.2 Paralinguistic Noises .....	35
4.1.2.3 Hesitation and interrupted words.....	36
4.1.2.4 Unintelligible words or segments .....	36
4.1.2.5 Alphabet letters .....	36
4.1.2.6 Acronyms .....	36
4.1.2.7 Interjections .....	37
4.1.2.8 Apheretic Forms .....	37
4.1.2.9 Pronouns .....	37
4.1.2.10 Anonymity.....	38
4.1.3 Alignment .....	38
4.1.4 Peer revision .....	38
4.1.5 Informational Tagging .....	39
4.2 The final design .....	40

## **PART II:**

<b>A Semantic/Pragmatic approach to the use of adverbs of certainty.....</b>	<b>45</b>
<b>5. ADVERBS OF CERTAINTY .....</b>	<b>46</b>
5.1 The list of adverbs studied in this dissertation .....	49
<b>6. THE CATEGORIES OF MODALITY AND EVIDENTIALITY .....</b>	<b>55</b>
6.1 Modality .....	55
6.2 Evidentiality .....	59
<b>7. POLITENESS.....</b>	<b>61</b>
7.1 Traditional theories of linguistic politeness .....	61
7.1.1 Grice (1975) .....	62
7.1.2 Illocutions .....	62
7.1.3 Leech (1983) .....	64
7.1.4 Brown and Levinson (1987) .....	68
<b>8. METHODOLOGY .....</b>	<b>72</b>
8.1 Corpus Linguistics .....	72
8.2 Methods.....	74
<b>9. DATA ANALYSIS AND RESULTS .....</b>	<b>81</b>
9.1. Mesmo .....	81
9.1.1. General remarks .....	81
9.1.2. Syntagmatic and paradigmatic description .....	81
9.2. Claro .....	86
9.2.1. General remarks .....	86
9.2.2. Syntagmatic and paradigmatic description .....	87
9.3. Com certeza .....	90
9.3.1. General remarks .....	90
9.3.2. Syntagmatic and paradigmatic description .....	90
9.4. Exatamente .....	93
9.4.1. General remarks .....	93
9.4.2. Syntagmatic and paradigmatic description .....	94
9.5. Of course .....	95
9.5.1. General remarks .....	95
9.5.2. Syntagmatic and paradigmatic description .....	95
9.6. Sure(ly) .....	98

9.6.1. General remarks .....	98
9.6.2. Syntagmatic and paradigmatic description .....	98
9.7. Exactly .....	100
9.7.1. General remarks .....	100
9.7.2. Syntagmatic and paradigmatic description .....	101
9.8. Definitely .....	102
9.8.1. General remarks .....	102
9.8.2. Syntagmatic and paradigmatic description .....	102
9.9. Summing up .....	104
9.10. A note on communicative context .....	108
9.11. A note on schooling level .....	110
9.12. A quick note on infrequent adverbs .....	113
<b>10. CONCLUSION .....</b>	<b>114</b>
<b>REFERENCES .....</b>	<b>118</b>
<b>APPENDIX .....</b>	<b>123</b>
Appendix 1 .....	123
Appendix 2 .....	125
Appendix 3 .....	126

## 1. INTRODUCTION

Modality is an interesting yet somewhat cryptic topic, with a long-established record of study. Indeed, the first account to modality dates back to 350 B.C and was primarily located in Logic<sup>1</sup>, when Aristotle devotes a chapter of his book *De Interpretatione* to his interpretation of modality. Today, the most cited works on the topic are grounded in the linguistic area<sup>2</sup>. Although the conceptual heart of modality is located in the linguistic area of Semantics, the interpretation of a certain index as modal or not cannot depend on semantics only in speech:

(1.1) É mesmo // (bfamcv06)

(1.2) É mesmo // (bfamcv07)

The examples above illustrate the fact that communicative contexts, illocution and prosodic features play an important role in disambiguating the uses of a certain index<sup>3</sup>. We need to broaden semantics in order to understand why, for example, the same lexical items can assume two different meanings and functions depending on the prosodic cues and semantic-pragmatic context in which this lexical item is inserted. Semantics can therefore integrate with other linguistic areas gainfully, even in the study of the semantic phenomenon of modality.

In this dissertation, based on data extracted from two comparable spontaneous speech corpora, one of the main objectives is to map and describe the semantic and pragmatic uses of the most frequent adverbs and adverbial expressions of certainty in both Brazilian Portuguese (BP) and American English (AE). Special attention will be given to the pragmatic function of marking politeness and the semantic function of signaling epistemic certainty. The BP data were extracted from the C-ORAL-BRASIL<sup>4</sup> and the AE data from the Santa Barbara Corpus of Spoken American English (SBC)<sup>5</sup>. Both corpora were compiled in accordance to methodological principles that distinguish them from other corpora available since the

---

<sup>1</sup> Aristote 2007.

<sup>2</sup> Bally 1932, Lyons 1977, Quirk et al. 1985, Biber & Finnegan 1988, amongst others

<sup>3</sup> As research in spoken language has demonstrated. See Simon-Vandenberg & Aijmer 2007, Culpeper 2011, Moneglia 2011 and Raso & Mello 2012.

<sup>4</sup> The project C-ORAL-BRASIL is coordinated by professors Tommaso Raso and Heliana Mello from the Federal University of Minas Gerais and funded by FAPEMIG, CNPQ, and UFMG and integrates the C-ORAL-ROM project, a set of corpora of the four main European Romance Languages, namely Portuguese, Italian, Spanish and French.

<sup>5</sup> The Santa Barbara Corpus was compiled by researchers in the Linguistics Department of the University of California, Santa Barbara. The Director of the Santa Barbara Corpus is Prof. John W. Du Bois, working with Associate Editors Wallace L. Chafe and Sandra A. Thompson (all of UC Santa Barbara), and Charles Meyer (UMass, Boston).

transcriptions show specific concern with the representation of illocution through prosody. This characteristic will be further explored and evidenced throughout the dissertation. Our main research questions are: 1. how are adverbs of certainty employed in each language, 2. to what extent does sociocultural variation determine both the type and the frequency of the lexical indexes, and 3. is there a clear boundary between the semantic and pragmatic content of a certain index?

A lot of uncertainty surrounds the definition of modal certainty. Before all else, there is no agreed definition of the topic up to this day. Yet despite the importance of the type of context, social roles and power relations in modality<sup>6</sup>, the vast majority of research on the topic pays little attention to the role of rhetorical effects of modal choices. Most studies are concerned with the modality in itself, and the lexical items that express the different types of modality. Simon- Vandenberg & Aijmer's 2007 book goes against the flow of traditional studies on the topic by analysing epistemic markers in a heteroglossic perspective, from a semantic-pragmatic approach. The authors propose a different methodology to the analysis of modal certainty that involves studying how different situations and rhetorical aims of the speaker can throw more light on the uses of a certain index. Their book, entitled *The Semantic Field of Modal Certainty*, was the inspiration for the analysis conducted in the second part of this dissertation. In the book, the authors point out that 'adverbs of certainty are multifunctional and have social meanings on different dimensions' (Simon-Vandenberg; Aijmer 2007:48). This suggests that those adverbs can hold concurrent functions, and amongst those, the semantic function of modalizers as well as the pragmatic function of politeness markers were the ones that appealed to me the most. That sentence aroused my interest in studying the polipragmatic behaviour of those adverbs contrastively in speech, using Brazilian Portuguese and American English to maybe elicit some socioculturally based differences in the expression of certainty and politeness in those two languages.

My aims with this dissertation are both descriptive and methodological. The methodological goal comes to corroborate with the Language into Act Language (LAcT - Cresti 2000) on how to adapt a corpus to its theoretical background to serve to the aims of a specific linguistic analysis -, and to corroborate with Simon-Vandenberg & Aijmer (2007) on how advantageous it can be to analyse modal certainty from a semantic-pragmatic perspective. The

---

<sup>6</sup> Simon-Vandenberg & Aijmer 2007:4.

descriptive goal encompasses describing in detail the uses of adverbs of certainty in BP and AE, contrastively and comparatively when possible and/or relevant, since examining correspondences in other languages can uncover subtle meaning distinctions which could go unnoticed in monolingual only studies.

In order to meet the preceding purposes, this dissertation contemplates two main parts: the first one dedicated to the creation of an adapted AE minicorpus to make it comparable to the already existing BP one, and a second part that offers a quantitative study of the adverbs and adverbial expressions of certainty in such languages.

### 1.1. Part I: The creation of the Santa Barbara Corpus of Spoken American English minicorpus

Dealing with spoken language has its theoretical premises and methodological consequences. This dissertation has its basis on the study of informal spoken language based on data extracted from corpora concerned with the theoretical premises of spontaneous speech analysis. The corpora used in this study follow the theoretical and methodological parameters of the Language into Act Theory – LAcT - (Cresti 2000), which provides a model to process and analyse speech taking into consideration its intrinsic prosodic cues.

The first part is a description of my contribution to a major project. This contribution consists in the adaptation of a subcorpus of the SBC to the methodological parameters of the already existing Italian and Brazilian minicorpora of the C-ORAL-ROM<sup>7</sup> project. The product is a minicorpus of 30,105 words suitable for contrastive studies amongst the four main European Romance languages. This adaptation expands the possibilities of contrastive studies within the C-ORAL-ROM project to include the global lingua franca to the comparable set of corpora. The AE adapted minicorpus will soon be available online to the academic community, at the DB-IPIC platform<sup>8</sup>, an XML research platform designed to simplify the study of the information structure of spoken language. Having an adapted AE minicorpus magnifies the dissemination of the C-ORAL-ROM project in its theoretical/methodological aspects.

---

<sup>7</sup> The aim of the C-ORAL ROM project is to provide the linguistic community and speech industry with a comparable set of corpora of spontaneous spoken language for the main romance languages, namely French, Italian, Portuguese and Spanish.

<sup>8</sup> Panunzi;Gregori 2011, at <http://lablita.dit.unifi.it/app/dbipic/>

Moreover, by extending the possibilities of comparative linguistic investigations within the project, we are able to establish a broader network of relations and meanings amongst the four main Romance languages and a non-Romance one of great academic reach – English.

Originally, the second part of this dissertation would be conducted on the basis of the analysis of the minicorpora of BP and AE only. I noticed, however, that expanding the data analysed would be advantageous to the description and I decided to use not only the data of the adapted minicorpus, but to adapt other texts upon demand.

## 1.2. Part II: A Semantic/Pragmatic approach to the use of adverbs of certainty

It would be too optimistic to think that we can track down all the variables of modal certitude only with the analysis of adverbs of certainty. However, studying them in conjunction with other functions they may acquire in speech is a good way to start solving the puzzle.

The second part of this dissertation is based on three main theoretical premises that will be tested in the data studied and further discussed throughout the dissertation:

### I. “*Contexts warrant more or less certainty.*” (Simon-Vandenberg & Aijmer 2007)

This notion states that the communicative context can condition the frequency and the types of adverbs of certainty. This suggests that these lexical indexes should be examined in context in order to describe their use more accurately.

### II. “*A recurring finding seems to have been that one factor which distinguishes the discourse styles of social classes is the use of certain types of adverbs.*” (Simon-Vandenberg & Aijmer 2007)

The statement suggests that diastratic variation might hold a powerful influence in the use of specific types of adverbs of certainty.

### III. “*We only say we’re certain when we are not.*” (Halliday 2004)

This means that speakers only mark their certainty in speech if what they are saying is subjected to the hearer’s judgement.

## 1.3. Objectives



This dissertation has two main objectives:

- I. Contribute to the creation of the SBC minicorpus, adapting the Santa Barbara Corpus of Spoken American English to the methodological parameters adopted in the compilation of the C-ORAL-ROM minicorpora.
- II. Map and describe the semantic and pragmatic uses of adverbs and adverbial expressions of certainty in Brazilian Portuguese and American English.

Aligned to the first general objective above are the following specific goals:

- a) Establish parameters for the selection of texts to be adapted.
- b) Contribute to the prosodic segmentation and alignment of the transcribed texts.
- c) Carry a final revision of the adapted texts, with special attention to the revision of segmentation;
- d) Map and measure the following information in the minicorpus: number of male and female speakers, number of words divided by type of interaction, number of utterances (simple and complex), and number of tone units.
- e) Contribute to the creation of the minicorpus specifications to be made available to the academic community.

The second general objective suggests the following specific ambitions:

- f) Examine the features that are important in the description of the use of adverbs and adverbial expressions of certainty in each language.
- g) Evaluate the boundaries between the use of certainty adverbs as modal or politeness markers and determine if those boundaries can be clearly described or not.
- h) Analyse the sociolinguistic variation in the use of modal adverbs of certainty between speakers using the metadata available.

#### 1.4. Organization

As previously stated, this dissertation mainly holds two separate but co-dependant parts. The second half of the dissertation would not be possible without the work of adaptation carried in part one. This dissertation is thus organized into two parts: the first one being dedicated to the careful description of the work performed in the adaptation of the SBC with the purpose to make it comparable to the already existent set of corpora of Brazilian Portuguese and Italian.

The second part will comprise the study of adverbs and adverbial expressions of certainty. Here I will initially present the core characteristics of the object of study, namely certainty adverbs and adverbial expressions, together with a brief panorama of the study that served as the basis for the analysis conducted in this dissertation (Simon-Vandenberghe; Aijmer 2007). I will then proceed to the chronological panorama of modality and politeness in the literature. The following section will deal with the description of the methodological procedures adopted in this research, followed by a discussion of the data and results found, comparatively across the languages studied when possible. In the final section I will reiterate the key points and outline the conclusions made, detailing the main outcomes and the questions for further discussion in a near future.

**PART I:**

**The creation of the Santa Barbara Corpus of Spoken American English minicorpus**

This research, as stated before, has its grounds on the study of spontaneous speech. The study of spoken language has had significant growth in the past couple of decades, following the development of high quality technology for the recording and treatment of speech (Moneglia 2011). Linguistic analysis based on empirical observation of spontaneous speech offers an insight into the linguistic phenomena observed in authentic data. Speech corpora offers the opportunity to study language the way it is operated by speakers in their everyday interactions. This alone constitutes the greatest advantage of corpora-based linguistic studies.

Vizcaíno (2007) emphasises this idea by pointing out that speech corpora provide the researcher with the possibility of studying our main communication channel - oral speech - in different communicative contexts, emulating what happens in real life. In addition to that, the author makes important remarks on the crucial role that prosody plays in linguistic studies dealing with spoken language. She states that prosody is essential when describing the meanings and functions of a certain object of study in speech, as it guarantees the appropriate interpretation of a variety of meanings and functions. Spoken corpora contain (or *should* contain) prosodic data that can be very helpful depending on the aims of the research, such as intonation and hesitation. Vizcaíno also points out that a good oral corpus should offer extra information which sometimes cannot be accessed by the recording itself, such as information about the situation and its context. Furthermore, the author emphasises the importance of having access to speaker information, which is essential to sociolinguistic analysis.

Once compiled, corpora must be made available in a way that guarantees easy access and interpretability. Moneglia (2011) lists some important aspects to be considered in spoken corpora compilation and annotation so that the final result can reach as many researchers as possible, which will assure great impact within the linguistic community. Over and above, the steps proposed by Moneglia guarantees that the corpora will be satisfactorily interpreted and explored in many different linguistic areas. According to the author, three steps must be observed when compiling spoken corpora. Firstly, the compilation process must assure that the corpus represents a variety of contexts as well as of speakers. Secondly, speech transcription should be segmented and analysed according to prosodic cues, as to mark the flow of the speech in the text. In addition to that, oral corpora should provide the researcher

with the alignment of text and sound in order to guarantee that oral corpora are adequately interpreted and easily explored by researchers in many linguistic areas<sup>9</sup>.

It is also imperative to highlight the importance of studying spoken language in the light of Pragmatics. In everyday oral interaction, speakers exchange information with one or more interlocutors, in a specific and dynamic context, performing acts with every utterance. These are the main characteristics of spontaneous speech, which can be systematically defined as the realization of unplanned linguistic acts in a free interaction amongst interlocutors that could be easily modified (Cresti & Scarano 1999). A corpus that portrays the oral diamesy, in which speakers interact and use language in everyday activities, is also a corpus that offers the opportunity to deeply understand how language is organized in spontaneous speech interactions. This purpose, however, can only be achieved if the data are extracted from corpora concerned with the representativeness of spontaneous speech. A corpus designed to represent the oral diamesy of a given language must do so by offering an ideal balance of various communicative situations which portray the reality of how speakers use language ordinarily. This is the main concern of both the C-ORAL-BRASIL and the adapted SBC minicorpora, and their theoretical premises and methodological consequences will be explained in detail in the following sections in this first part of the dissertation.

---

<sup>9</sup> Examples 1.1 and 1.2 presented in the introduction of this dissertation also serve to illustrate that notion.

## 2. THE LANGUAGE INTO ACT THEORY

The Language into Act Theory (Cresti 2000) – LAcT, offers a model for studying how spontaneous speech is prosodically and informationally organized. It was developed following intense empirical observation of oral texts with the core and common characteristic of being spontaneous. The texts analysed come from the LABLITA *corpora*<sup>10</sup> with approximately 40 hours of oral speech produced in natural contexts whose linguistic arrangement happens simultaneously to the oral production. Those spontaneous interactions of two or more speakers were transcribed, prosodically segmented and extensively analysed. The theoretical/methodological assumptions of LAcT were grounded on examples of speech as it naturally happens in the majority of oral interactions of a typical speaker on their everyday life. The premises of LAcT evolved from the observation of data and the generalisations made from the regularities found, which seem to share the same theoretical gist.

A logical but somehow commonly ignored fact about spoken language is that it is realised through sound. This obvious characteristic is extremely relevant to the model once it assigns methodological consequences to the study of the speech. By consistently observing spontaneous language, Cresti (2000) realised that the existing models for documenting and describing spoken language were not enough to represent it. She noticed that speech should not be segmented following primarily syntactic, semantic or orthographic criteria. In reality, those criteria can forge the misinterpretation of the phenomena that are inherent to the oral diamesy of language. Cresti understands that speech can only be truly represented if prosody is given a central role in that process.

LAcT offers a pragmatic model for both prosodic and information analysis of spontaneous speech and has its roots in the Speech Act Theory (Austin 1962). For the LAcT, each utterance conveys an action, notion that was first introduced by Austin in 1962. According to him, when a speaker utters a sentence, s/he is acting in three different, simultaneous and inseparable levels: the locutive, illocutive and perlocutive levels. The first one is purely linguistic, or the action of saying something. The second one – the illocutive level -, carries the communicative intention of the locutive material. The third and last level raises different

---

<sup>10</sup> The LABLITA *corpora* mainly counts with interactions between adults, but also with media broadcast and language acquisition recordings.

opinions on different authors. According to Austin (1962), the first author to propose the notion of speech acts, the perlocutionary act is both the motivation for the illocutionary act and the effect caused by it on the hearer. A conversation would therefore be considered a chain of perlocutions that are both the effect of a previous act and the motivation for the upcoming act. To Searle (1969), the perlocutive level of a speech act is only the direct effect caused by a certain illocution. That notion is widely disseminated in the literature on the topic but has been questioned over the years. Cresti (2000) has a different take on the matter. According to the author, the perlocution is the emotive motivation of an illocutionary act. This emotional drive is particular to each person and causes each speaker to react differently and accordingly to any given illocution.

Correspondingly to that notion, Cresti (2000) states that the utterance corresponds to both locutive and illocutive levels, once it is the linguistic part of a speech act that necessarily conveys an action. Her model – LAcT - offers an alternative to studying language taking into consideration the relationship between the linguistic domain (locution according to Austin), as well as the action domain (illocution), in the lights of a prosodic interface. In this model, the utterance is described as the smallest linguistic unit that can be pragmatically and prosodically interpreted by itself. What defines an utterance here is not its semantic or syntactic independence, but its pragmatic autonomy as a linguistic unit. This definition allows for sentences such as the ones highlighted below<sup>11</sup> to be considered and analysed as utterances<sup>12</sup>:

(2.1)        \**DAN: no / I don't know how to play it //*  
               \**JEN: oh // okay / I'll teach you //*

(2.2)        \**FRE: see / the day before yesterday / I did ice cream / right // Balian //*  
               \**RIC: hum hum //*

For LAcT, there are no restrictions regarding the lexical or syntactic nature of an utterance. That notion classifies one of the main differences between this and previous definitions of

---

<sup>11</sup> For a better understanding of the examples given here, consider the following transcription criteria (RASO;MELLO 2012): a) the asterisk followed by three capital letters (such as ‘\*DAN’) indicates the speaker and his/her turn; b) the ‘/’ portrays a tone unit perceived as non-terminal; c) the ‘//’ indicates a tone unit perceived as terminal and the end of an utterance, d) the ‘&’ followed by a segment indicates interrupted words and finally e) the ‘[.]’ indicates a non-terminal prosodic break followed by the number of words that should be cancelled.

<sup>12</sup> Audio file in the CD enclosed

what an utterance may or may not be. In this model, the definition of utterance is not based on semantic or syntactic parameters, which means that expressions such as ‘oh’ and ‘hum hum’ can be categorised as utterances since they are prosodically and pragmatically autonomous.

Prosody plays a central and important role in the model. It not only provides the cues for the segmentation of speech in utterances, but also mediates the relationship between the pragmatic and linguistic units (locutive material and illocution). Let us observe the following example:

- (2.3)      \*DAR: *do what you want / with the time you have //*  
              \*DAR: *learn //*  
              \*DAR: *give //*  
              \*DAR: *whatever //*  
              \*PAM: *love //*  
              \*DAR: *love //*  
              \*PAM: *love //*

In speech it is common to have the same locutive material conveying different illocutions. Similarly, it is also possible to have different locutive material conveying the same illocution. As demonstrated by the example above, even though the locutive content of the utterances is the same, the actions are completely different. The first ‘love’ is a question, the second a confirmation, and the third an irony. The matter of illocutions will be dealt in further detail in part II of this dissertation, as it also plays an important role in the analysis carried out in the next part of this paper.

These examples highlight the importance of prosody in mediating the relationship between the pragmatic and linguistic dimensions of an utterance (Cresti 2000). The adequate interpretation of the utterance’s communicative value can only be determined if the acoustic signal is accessed together with the transcription.

Prosody is not only intrinsically connected to the illocutionary act. It also plays an equally important role in segmenting each utterance in tone units and in assigning each tone unit a specific informational function. According to LAcT, the utterance can be simple, if realised in



a single tone unit, or complex, if executed in two or more tone units. The identification of an utterance is realised through perceptual criteria that can be identified by any speaker of a given language even with little training. In a brief summary, the perceptual identification of an utterance is done through the sense of an intonational break perceived as terminal and represented in the transcriptions by the double slash ('//'). In other words, the hearer feels like the speaker has finished speaking, has completed an action. The utterance can also be segmented into other tone units this time perceived as non-terminal, represented by the single slash ('/'), which can also be observed in example (2.3) above. Note that in between 'do what you want' and 'with the time you have' there is not any pause. It is important to highlight that tone units or prosodic breaks do not necessarily correspond to pauses (Moneglia 2011). According to the author, only 40% of the non-terminal breaks are followed by pauses. When it comes to prosodic breaks perceived as terminal, that number rises to 60%. This empirical analysis shows that pauses are not sufficient in signaling the end of tone units, terminal or non-terminal.

For LAcT, each tone unit corresponds to an informational one (Cresti & Moneglia 2010), and each informational unit has its specific intonational pattern. This systematic correspondence between informational and tone units was developed in accordance to the perceptual phonetics work named *A Perceptual Study on Intonation. An Experimental Approach to Speech Melody* carried by t'Hart (t'Hart; Collier; Cohen 1990). This close relationship between both theories was summarized for better visualisation in the chart<sup>13</sup> below:

Table 2.1: Prosodic X Informational Patterns

Prosodic Pattern		Information Pattern	
	root		Comment
(prefix)	(suffix)	(Topic)	(Appendix)
	(parenthesis)		(Parenthetical)
(incipt)	(phatic)	(Incipt)	(Phatic)

On the prosodic pattern of t'Hart and his research group (1990), root patterns are a mandatory and necessary configurations in an utterance, which is also true on the information pattern

<sup>13</sup> Adapted from Cresti & Moneglia 2010.

proposed by Cresti (2000), with the Comment unit (COM) on that role. The prefix and suffix patterns are optional: the prefix can precede a root configuration or another prefix, while the suffix pattern follows the root. These prosodic patterns were incorporated to LAcT as both theories share the premise of being perceptual. LAcT states that each information unit has a specific intonation unit, with its own core characteristics, and their correspondence is standardized as shown above. Only one type of information unit is mandatory, which carries the illocutionary force of the utterance. This is called the Comment unit, which has the prosodic pattern of nuclear unit or *root*.

Information units are identified via three criteria: functional, intonational and distributional. Those three criteria in combination allows for the identification of information units in speech. The function the unit has in the utterance, its intonational profile and the position it occupies in the utterance in relation to the COM unit are equally important in assigning the tone unit its correspondent informational function. Example 2.4 below illustrates what was said so far:

(2.4) \*ALI: *so they go barging in on yyy // =COM= so mom /=TOP= felt obligated to ask those two idiots to lunch // =COM=*

In the example above, there are two utterances, a simple (one tone unit) and a complex one (two or more tone units). On that note, simple utterances bear only the COM unit, whilst complex utterances have not only that necessary unit but alternative ones that can either help the realisation of the illocutionary force or are related to the interpersonal and expressive channels of the illocution. The second utterance, a complex one, carries not only the Comment unit, but also another central unit in the informational articulation of utterances: the Topic (TOP), which has the prosodic pattern of a prefix and precedes de COM unit. The TOP has the function of establishing the scope of application of the illocutionary force, transmitted by the COM unit.

According to LAcT, there are two categories of information units: textual and dialogic. The textual units are the ones that build the text itself. The dialogic units act upon the hearer or the communicative situation. The textual units are named Comment (COM) – which can have the form of Multiple Comment (CMM) or Bound Comment (COB) in more complex patterns;

Topic (TOP), Appendix Comment (APC), Appendix Topic (ACT), Parenthesis (PAR), and Locutive Introducer (INT). The dialogic units are dedicated to the pragmatic fulfilment of the utterance, serving in the regulation of the interaction. They are the Incipit (INP), Conative (CNT), Phatic (PHA), Allocutive (ALL), Expressive (EXP), and Discourse Connector (DCT). There are also the units that have no informational value, namely Time Taking (TMT) and Scanning unit (SCA). We will not deal with every information unit in detail as that is not the focus of this particular study.

Both the compilation of the *corpora* in the C-ORAL-ROM project and the adaptation of the SBC follow the premises established by LAcT. The methodological criteria adopted in the compilation of the *corpora* and minicorpora mentioned will be explored in detail in the next two sections.

### 3. THE C-ORAL-BRASIL

The C-ORAL-BRASIL is a spontaneous speech corpus of Brazilian Portuguese (BP). It is prosodically segmented in utterances and tone units, and designed to study the informational structure and illocutions of the BP, based on the Language into Act Theory (Cresti 2000). The C-ORAL-BRASIL represents the *Mineiro* diatopy of Brazilian Portuguese (BP), as speakers are chiefly from the state of Minas Gerais (mainly from the surrounding cities of the capital Belo Horizonte). The corpus was compiled following the same methodological criteria adopted in the compilation of the C-ORAL-ROM, a set of spontaneous speech corpora of Italian, European Portuguese, Spanish and French. The corpora of the C-ORAL family are, therefore, comparable amongst themselves, and the BP corpus comes to add a non-European variety to the four main European Romance languages represented in the set of corpora mentioned above.

The C-ORAL-BRASIL lays plans for approximately 300,000 words in its final form, distributed in informal and formal portions. The main differences between those two portions are the planning of speech (completely unplanned versus slightly planned or rehearsed), the communicative channel (the informal part does not contain phone or media recordings, for instance), and the register itself, which is affected by the social domain (more or less formal)<sup>14</sup>. The informal part of the C-ORAL-BRASIL was completed in 2012 and contains 208,130 words. The formal part of the corpus is being currently compiled and transcribed.

The recordings are conducted in natural context while the speakers involved do the activities they would normally do in their daily lives. It is extremely important that the speaker feels comfortable with the recording equipment so it will not affect their performances in any sense. This can only be achieved with the help of high quality recording equipment, including wireless microphones. To ensure a high acoustic quality, the recordings of the C-ORAL-BRASIL were made with Marantz PMD660 Professional Solid State Recorder, and high resolution, non-invasive wireless equipment, mostly mono-directional clip-on microphones (Sennheiser EK/SK 100 G3). Whenever there were more than two interactants, an analog mixer (Behringer XEXYX 1222 FX) was also used.

---

<sup>14</sup> RASO & MELLO, 2012

As already discussed in the previous section, spoken language cannot be analysed without its sound asset. Due to that fact, the transcribed texts of the C-ORAL-BRASIL are aligned utterance by utterance with its correspondent acoustic signal. This is done with the software Winpitch (Martin 2004), which provides the tools for the text-to-speech alignment and puts together sound, text and spectrogram, consequently granting the analysis of a series of acoustic parameters.

The transcribed texts are equitably divided into monologues, dialogues, and conversations. They are also divided in accordance to the context they represent, which can be public (20% of the situations) or private (80% of the interactions). The group of public interactions is characterized by the speaker's behaviour as a public figure with a social function. For example, a talk amongst three participants in a union meeting would characterise a public conversation. Differently from a conversation amongst friends in which the speaker acts without such restrictions, and would therefore portray a private type of interaction. The texts from the corpus C-ORAL-BRASIL also present information about the context (situation, topic of conversation, date, place, and number of participants) and its participants (gender, educational level, profession, and birthdate). The corpus' compilation seeks first and foremost for great diversity of communicative situations, notwithstanding the diastatic variation.

Collecting minicorpora is a way of having a portion that is representative of the whole. The C-ORAL-BRASIL has a 30.000 words subcorpus of its informal portion, which has been used as the basis for studies in the project since early stages of compilation. Through statistical analysis (Raso, Mittmann 2012) carried in the subcorpus for the BP spontaneous speech corpus, it can be said that the C-ORAL-BRASIL as a whole operates proportionally to its subcorpus when it comes to the main units for speech measurement, namely utterances and dialogic turns<sup>15</sup>.

Both minicorpora from the C-ORAL-BRASIL and Italian portion of the C-ORAL-ROM compile 20 informal texts each in a total of approximately 30.000 words. The samples were extracted from their respective main corpora following the criteria below:

- a) representativeness of each type of interaction, namely monologues, dialogues and conversations, in number of words;

---

<sup>15</sup> See RASO; MELLO, 2012, chapter 5 for further information.

- b) great diaphasic variation without repetition of communicative situations;
- c) high acoustic quality;
- d) great diastratic variation with special attention to the avoidance of duplication of same informant taking part in another communicative situation;

These parameters aim to guarantee that the minicorpora will offer a balanced and representative sample of the corpus it attempts to picture. In total, the C-ORAL-BRASIL subcorpus represents 15% of the main corpus. It comprises 15 texts representing the private domain and 5 speaking for the public domain, divided in 7 monologues, 7 dialogues and 6 conversations, with approximately 1500 words each. Those texts were informationally tagged following the theoretical premises of LACT (Cresti 2000). The C-ORAL-ROM minicorpus of Italian has 14 texts representing the private domain and 6 picturing the public one, in 8 monologues, 7 dialogues and 5 conversations. Those texts also offer informational tagging and, together with the BP minicorpus, are available to the community at the online platform DB-IPIC (Panuzi; Gregori 2011).

#### 4. SANTA BARBARA CORPUS OF SPOKEN AMERICAN ENGLISH

The Santa Barbara Corpus of Spoken American English<sup>16</sup> - SBC - (Du Bois et al. 2000-2005) is a body of 60 recordings of naturally occurring spoken interaction of American English (AE) from all over the United States. The SBC represents a variety of people from different regional origins, ages, occupations, genders, and ethnic and social backgrounds. The corpus documents various ways in which people use language in their everyday lives. Telephone conversations, card games, food preparation, on-the-job talk, classroom lectures, sermons, storytelling and town hall meetings are among the interactions recorded.

The SBC is part of the International Corpus of English (ICE) and is the main source of data for the spontaneous spoken portions of the American component of the International Corpus of English. The SBC offers explicit information about the speakers and the communicative situation and also contains prosodic information that can be retrieved from the audio files. The original corpus comprises 60 texts for a total of approximately 249,000 words.

The SBC was not originally comparable to the C-ORAL family minicorpora in terms of design specification. Originally, the texts of the SBC were only readable with the CLAN software. The software, similarly to the WinPitch (Martin 2004), allows the alignment of sound and text. However, the acoustic information promoted by spectrogram analysis is not offered within the software. Similarly to the C-ORAL-BRASIL corpus, the SBC showed evidence of following prosodic segmentation criteria in their texts. For example, the end of very clear illocutions, such as complete questions, was originally marked. In the texts of the SBC, prosodic breaks perceived as terminal are called intonation units and marked with a bullet point. Other prosodic breaks, perceived as non-terminal, are called intonation subunits. Tone variations, pauses and some phenomena such as vowel lengthening are also marked in the original transcriptions, in contrast to the BP corpus.

Choosing the SBC as a potential corpus for the adaptation to C-ORAL-BRASIL's methodological premises took into consideration two core characteristics of the SBC. Firstly, the SBC compiles recording of spontaneous speech, which is the same portrayed by the corpora in the C-ORAL family. Secondly, the diversity of communicative situations seemed

---

<sup>16</sup> Available at <http://www.linguistics.ucsb.edu/research/santa-barbara-corpus>

to be comparable to the assortment found in the C-ORAL-ROM and C-ORAL-BRASIL minicorpora. Lastly, the acoustic quality of the majority of the recordings is good enough to enable prosodic investigations. That variation of context, together with the notion of natural interaction and good acoustic quality are the features that made the SBC potentially comparable to the C-ORAL-BRASIL.

As a result, therefore, we designed the SBC minicorpus to be prosodically segmented and aligned using the same methodological/transcription criteria and software adopted in the compilation of the C-ORAL-ROM spontaneous speech corpora. The final product is a subcorpus of 20 texts and approximately 30,000 words, using the same architecture of the Italian and Brazilian Portuguese minicorpora. One of the most important advantages of this adaptation is the comparability of the corpora, as they now share the same basic architecture and equivalent transcription and prosodic segmentation patterns.

#### 4.1 The process of adaptation

The adaptation of the SBC follows some steps previously adopted in the compilation of the C-ORAL-BRASIL corpus. In summary, the process of adaptation of the SBC can be divided into 5 stages:

1. Text selection: search for informal recordings with good acoustic quality and communicative situations comparable to the BP and Italian minicorpora;
2. Transcription: cleaning the original transcription of its original criteria and segmenting it according to the LACT (Cresti 2000);
3. Alignment and first revision: text-sound alignment with the software WinPitch and correction of possible segmentation errors;
4. Peer revision: the aligned texts were submitted to a second revision, this time done by two experienced transcribers, in order to discuss problematic cases regarding the prosodic segmentation;
5. Informational tagging: texts were manually tagged in information units according to the LACT.

The subsections below explain in detail how the files from the SBC were chosen and adapted to build a comparable minicorpus, stage by stage. My personal contribution in the process of



adaptation can be observed in most of the stages of compilation, chiefly in the choice of texts, their prosodic segmentation and revision.

#### 4.1.1. Choosing the files for the minicorpus

The files used to compose the adapted SBC minicorpus were chosen primarily on their acoustic quality and secondarily upon the situation they represented. The acoustic quality was judged in the software WinPitch by checking the quality of the formants and F0. The table below (Raso; Mello 2012) shows how the acoustic quality was classified:

*Table 4.1: Description of acoustic quality tags.*

<b>Tag</b>	<b>Description</b>
A	Extremely high quality. Almost no voice overlapping and/or background noise. Trustable F0 computation for (practically) the entire file.
AB	High quality. Low voice overlapping and/or background noise. Trustable F0 computation for (practically) the entire file.
B	Medium quality. Some voice overlapping and/or background noise. Trustable F0 computation for most part of the file.
BC	Mid low quality. Some voice overlapping and/or background noise. Trustable F0 computation for at least 60% of the file. Audio is clear for listening throughout the entire file.
C	Low quality. Some voice overlapping and/or background noise. Trustable F0 computation for at least 60% of the file. Some portions of the audio may not be clear for listening.

If we compare the acoustic quality of the SBC and C-ORAL-BRASIL minicorpora, the latter is clearly of superior quality - due to the high quality recording equipment used. It should be stressed that the SBC has recordings dating back from the 80's, when the best recording equipment available was still very basic in comparison to what is readily available nowadays. The recordings of the SBC were taped on Digital Audio Tape (DAT) and recorded in stereo at 32 kHz or 48 kHz, on Sony TCD-D6 or TCD-D7 portable DAT recorders, using small, stereo microphones. A few earlier (late 90's) recordings were made on high quality analogue cassette recorders. The use of such equipment limited the F0 calculation in the WinPitch, especially when the recordings took place in loud contexts with background noise.

The files were also chosen on the basis of the diaphasic variation they represent, in a great variety of conversational situations. The diversity of communicative situations found in the

SBC seemed to be compatible with the assortment found in the two pre-existing minicorpora which integrate the C-ORAL-ROM set of corpora. Those situations were sorted into monologues, dialogues and conversations. Once chosen, the sound file selected for the transcription adaptation was reanalysed in order to select the segment that best represented its type of interaction. The duration of the recordings in the original SBC varies between 25 and 30 minutes each file. The selected segment (between 5 and 10 minutes, in accordance with the C-ORAL-BRASIL's recordings approximate duration) was edited with the software Audacity, an open source, freeware audio editor<sup>17</sup>. The process of editing the audio involves cutting the selected part to be transcribed, which is later aligned with the assistance of the WinPitch.

#### 4.1.2. The transcription

The transcription criteria used in the adaptation of SBC follows the criteria adopted in the C-ORAL-BRASIL corpus. It should be faithful to the recording, with regard to the prosodic units and what was actually said. The two transcribers who participated in the SBC adaptation process began the process during their final year in English, at UFMG. They were experienced transcribers of the C-ORAL-BRASIL project, having spent at least 2 year working on different stages of the corpus' compilation - recording, transcription, revision and informational tagging. They were allowed to change the original transcription when deemed necessary, since the SBC is available online under a *Creative Commons* license<sup>18</sup>.

The C-ORAL-BRASIL transcription criteria were fashioned to represent the processes of lexicalization and grammaticalization in language. Therefore some words were not written in accordance to its traditional orthography in order to preserve its morphological and syntactic changes. Because of that, some phenomena were not represented in the transcription, such as phonological processes - including sandhi and vowel lengthening, as they do not characterise any of the processes mentioned above. These phenomena are, however, not lost, as the researcher can immediately identify them by hearing the recording in alignment with the text. The following subsections deal with the general prosodic and orthographical conventions<sup>19</sup>

---

<sup>17</sup> Available at <http://audacity.sourceforge.net/download>

<sup>18</sup> A Creative Commons license gives the public the right to share, use and build upon a work that was created.

<sup>19</sup> These conventions are discussed in further detail in chapter 2 of the book *CORAL-BRASIL-1: Corpus de referência do português brasileiro falado informal* (RASO;MELLO, 2012)

established in the compilation of the C-ORAL-BRASIL corpus and implemented in the adaptation of SBC.

#### 4.1.2.1. Prosodic Units

The symbols used in the segmentation of the utterances in prosodic units are as follows<sup>20</sup>:

- the double slash marks a terminal prosodic break, which signals the end of an utterance and the realization of an illocution.

(4.1) \*LYN: *well you can trim 'em too short //* (afammn01)

- the single slash represents a non-terminal prosodic break, which indicates the end of a tonal unit inside an utterance.

(4.2) \*LYN: *no / I don't have my own equipment at all //* (afammn01)

- the sign of addition indicates an interrupted terminal break, or, in other words, that the utterance was not completed.

(4.3) \*LYN: *they're <really> +* (afammn01)

- [n°] signals the phenomena of retracting (when the speaker restates what was said). The number indicates the number of words to be cancelled in the process.

(4.4) \*LYN: *we [/1] that was kind of a double thing that &he / we had in [/1] in another class so / it was kinda review for us //* (afammn01)

The asterisk signals the beginning of a turn, and the identification of the speaker is made with three capital letters. The angled brackets ('<' and '>') signal overlapping.

#### 4.1.2.2 Paralinguistic Noises

These are the sounds that do not convey an illocution, or an informational unit, such as laughs, coughs, throat clearing. They are transcribed as *hhh* and are not prosodically segmented. The same sounds can, however, have an illocutionary value, in which case they are segmented. If the laugh is, for example, an ironic answer to a question, or used as an obvious agreement to something that was said, it will be segmented, as can be seen in the example below:

---

<sup>20</sup> Audio files in the CD enclosed.

- (4.5) \**ALI: well I grew up in Philadelphia / and that's too close to New York //*  
\**SEA: <hhh> // (afamcv05)*

Some paralinguistic noises will be added to the comments' section of the header by the transcriber if deemed necessary.

#### 4.1.2.3 Hesitation and interrupted words

Hesitation and filler sounds are transcribed as *&he* no matter the quality of the sound produced. Words that are incomplete for any reason are signaled with a commercial 'e' (&).

- (4.6) \**LYN: but / anyway / &he + (afammn01)*

- (4.7) \**LYN: and they're just / a little bit for &f [/1] the first couple of days / you know / I mean / they're just / sore // (afammn01)*

#### 4.1.2.4 Unintelligible words or segments

Words or segments that could not be decoded by the transcriber in the original transcriptions were transcribed as *xxx*. If the transcribers could clearly decode the word or segment during the process of adaptation, it was then transcribed as heard. If not, the code used in for unintelligible words, in accordance with the C-ORAL family corpora's transcriptions, is *xxx* and *yyyy* for segments.

#### 4.1.2.5 Alphabet letters

Alphabet letters are transcribed as pronounced, regardless of the convention:

- (4.8) \**BER: I think between first and second or //*  
\**FRA: hum hum //*  
\**BER: I wasn't down in ey's / bee's / and cee's // (afamcv05)*

#### 4.1.2.6 Acronyms

Two different rules apply to acronyms - depending on the way it is pronounced. If the acronym is spelled, letter by letter, the rule described in the previous subsection is applied. In

other words the acronym is transcribed the same way as it is pronounced, sound by sound, forming a word:

- (4.9)        \*LAN: *don't they // we are //*  
               \*RAN: *any veeeffar //* (apubdl01)

If the acronym is pronounced as a word instead of letter by letter, the rule is to transcribe it as a single word in capital letters:

- (4.10)       \*BER: *the New York Opera //*  
               \*ALI: *the MET //*  
               \*BER: *hum hum //* (afamcv05)

#### 4.1.2.7 Interjections

Some interjections are very commonly used to express accordance, to negate or to question what was said. These were transcribed as *hum hum* and *hum*, as showed previously. Interjections such as *oh* and *wow*, in expressing an emotional state, were also found in the transcriptions:

- (4.11)       \*FRE: *oh / they could tell right away //* (afamd101)

#### 4.1.2.8 Apheretic Forms

The phenomenon of losing one or more sounds in the beginning of a word was registered in the adaptation. However, apheresis was only registered when considered a possible marker for lexicalization and not a problem generated by fast speech. For example, the form *course* (of course), *wanna* (want to), *kinda* (kind of), amongst others, were transcribed as such when pronounced that way.

#### 4.1.2.9 Pronouns

The cliticization of pronouns such as *you* and *them* were marked in the transcription as they possibly represent a grammaticalization process in English. When such pronouns were pronounced as 'u or 'em, they were transcribed that way:

(4.12) \**FRA: see what you do is / you sort of sidle up to 'em and you give 'em a bump* // (afamcv05)

#### 4.1.2.10 Anonymity

Personal names of speakers on the recordings, as well as other identifying information such as telephone numbers, have been replaced by yyy in the transcriptions, and have been altered to preserve the anonymity of the speakers by filtering the audio files to make these portions of the recordings unrecognizable. Pitch information is still recoverable from these filtered portions of the recordings, but the amplitude levels in these regions have been reduced relative to the original signal. These alterations were not done for this adaptation though; they had already been made by SBC's transcribers. In the original transcriptions, the anonymised names were replaced by pseudonyms, which were later adapted to yyy in accordance with the transcription criteria of the C-ORAL-BRASIL.

(4.13) \**SEA: and it is the geography / yyy* //  
\**ALI: it's whatever it is* // (afamcv05)

#### 4.1.3 Alignment

As previously discussed, the importance of text-sound alignment is unquestionable. This step provides researchers with the opportunity to listen to the segment while reading the transcription. This alignment is realized in the software WinPitch (Martin 2004), which can also be used for a series of acoustic studies.

The text-to-speech alignment is done illocution by illocution. In other words, a segment of the alignment corresponds to a prosodic break perceived as terminal or interrupted. This also means that the alignment process calls for a revision of the adapted transcriptions, as at this stage the transcriber is again able to select segment by segment and test its acceptability as a terminal or non-terminal prosodic break.

#### 4.1.4 Peer revision

During the alignment process, some cases were judged as problematic, mainly because of their non-correlation to what the transcribers were used to finding in the BP corpus. When dealing with those cases, the transcribers were in doubt about whether a segment should be

marked as a terminal or non-terminal break. Part of the aligned texts was then submitted to peer revision, to seek solutions for the most dubious occurrences, in order to improve consistency in the *minicorpus*. Let us observe the example below which illustrates one of the most frequent problematic cases:

(4.14) \*LYN: *that's another thing too / is I kinda had a general idea of [/1] of kinda how to do it just watching him / or [/1] and watching / people come to our place and doing it and stuff / you know // and / I don't know then / down there / &he / it's mandatory / you have to [/3] &he / to graduate / you know / or well / to / get the degree / you know / you have to take this class // (afammn01)*

At times, the transcribers had difficulty determining whether the prosodic break after the segment *you know* was to be considered terminal or non-terminal. Issues like these were solved by listening carefully to the segment and discussing the theory behind it until a solution was found, keeping in mind that the transcription should remain consistent.

#### 4.1.5 Informational Tagging

Once segmented and aligned, the texts of the SBC adapted minicorpus were submitted to informational tagging in accordance with the theoretical/methodological premises of the LACT - already explored in this paper. The process of informationally tagging a transcription cannot be done automatically and therefore is a time-consuming task that should be executed carefully and reviewed with close attention. The annotator who carried out this work was trained to undertake such task, and the annotation of each and every transcription was later submitted to peer revision with one of the coordinators of the project C-ORAL-BRASIL to ensure maximum quality.

Table 4.2 below presents the full list of information tags used in the annotation. Since I have not actively participated in this particular stage of the adaptation, I will not describe each information unit in detail in this paper<sup>21</sup>. The information units that are particularly relevant for the study described in part II of this dissertation will be further discussed then.

---

<sup>21</sup> For more information on the matter, see MITTMANN, 2012.

Table 4.2: Informational Tagging

Textual Units		Dialogic Units		Empty Units	
COM	Comment	INP	Incipt	SCA	Scannning
CMM	Multiple Comments	PHA	Phatic	EMP	Empty (incomplete units)
COB	Bound Comment	CNT	Conative	TMT	Time Taking
TOP	Topic	ALL	Allocutive	UNC	Unidentified
APC	Appendix Comment	EXP	Expressive		
APT	Appendix Topic	DCT	Discourse Connector		
PAR	Parenthesis				
INT	Locutive Introducer				

In addition to these tags, the sign *\_r* can be incorporated to any informational tag, meaning that the unit was found within reported discourse. The sign *\_s* can be added to the COM or COB tags to signalise subordination.

#### 4.2 The final design

The SBC minicorpus comprises 30,105 words in 20 texts. This number corresponds to the transcription of 2 hours and 25 minutes of spontaneous speech recordings. Each transcribed text has approximately 1,500 words, apart from three texts that count less than 1000 words and two that count more than 2000. The corpus is divided into two sections: familiar context (22,594 words) and public context (7,511 words). All of this together comprises 5 conversations, 8 dialogues and 7 monologues. Speech transcriptions were done in CHAT format (Macwhinney 2000) with implementation of a prosodic annotation system (Cresti; Moneglia 1997) already explored in this paper.

Each one of the 20 minicorpus files comprises the following components:

- a) Audio recordings in wav format;
- b) Transcriptions in txt and rtf formats;
- c) Header with metadata of the recording in txt format;
- d) Text-sound alignment in xml and wp2 formats to be opened with the software WinPitch;

Each recording comes with a header that contains all relevant metadata. The metadata of the recording and speakers was originally available in separate files that could be downloaded



from the official website. This data was adapted to the parameters adopted in the C-ORAL-BRASIL corpus. After the adaptation, the header offers information about the given recording's situation and speakers in one single file, as well as comments that the transcriber judged relevant for the accurate interpretation of the transcription. The data regarding the context of the recording include the title of the recording; the name of the file; the name of the participants and their social information; the date and place of the recording; the situation and topic of conversation; the classification of the interaction; the length of the transcribed segment; the number of words of the transcription; the acoustic quality of the recording; the names of the transcriber and reviser; and finally, the transcriber's comments. The speaker's metadata include the fictional name and initials of the speakers, their sex, age, schooling, role within the interaction and place of birth. Their age and schooling adhere to the following classification:

*Table 4.3: Speaker's age and schooling classification*

<b>Age</b>		<b>Schooling</b>	
<b>A</b>	18 to 25 years old	<b>1</b>	Incomplete basic level or up to 7 years of schooling
<b>B</b>	26 to 40 years old	<b>2</b>	Up to undergraduate degree as long as not having a profession related to university degree
<b>C</b>	40 to 60 years old	<b>3</b>	Professions dependent on a university degree
<b>D</b>	over 60 years old	<b>X</b>	Unknown
<b>X</b>	Unknown		

The filenames adopt the following logic:

- a) The first letter represents the language of the transcribed text ('a' for American English);
- b) The following three letters represent the context ('fam' for familiar and 'pub' for public);
- c) The subsequent two letters represent the text typology ('cv' for conversations, 'dl' for dialogues, and 'mn' for monologues);
- d) The last code is a number that identifies the transcription in its group.

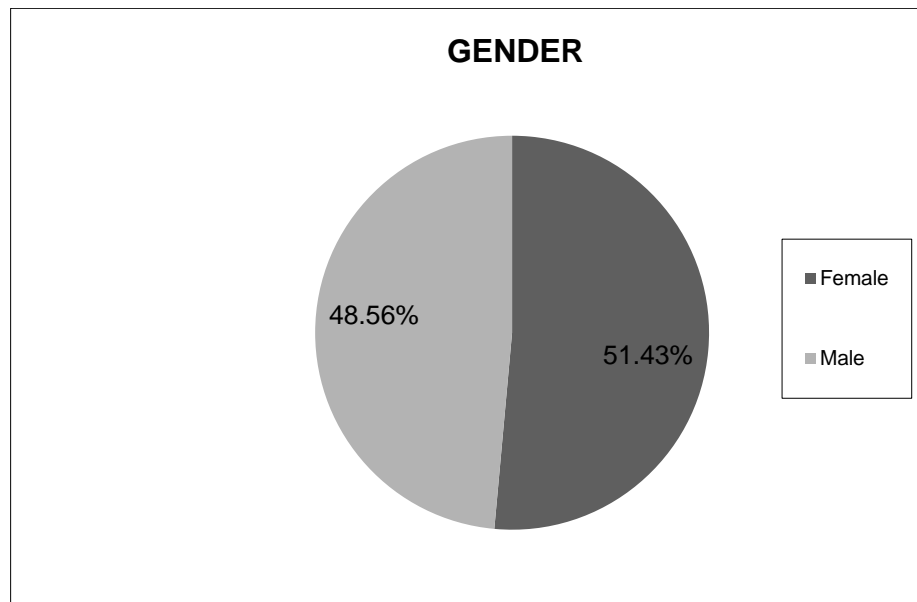
We end up with filenames such as 'afammn01' (American English, Familiar, monologues, 01) or 'apubcv03' (American English, public, conversations, 03). Table 4.4 below shows the header for the file afammn01 as an example:

Table 4.4: header

@Title: Actual Blacksmithing
@File: sbc0001 / afamnn01-actualblacksmithing
@Participants: LEN, Lenore, (woman, B, 2, student, participant, Los Angeles/CA) LYN, Lynne, (woman, A, 2, student, participant, Montana/MT)
@Date: 01/07/1988
@Place: living room of house trailer, outskirts of small rural town, Hardin, Montana
@Situation: Lynne is a student of equine science, and is the main speaker. She is telling Lenore (a visitor and near stranger) about her studies.
@Topic: vacation in Minnesota, horses
@Source: Santa Barbara Corpus
@Class: informal : particular : monologue
@Length:7'33''
@Words: 1794
@Acoustic_quality: B
@Transcriber: Adriana Couto Ramos
@Revisor: Adriana Couto Ramos and Frederico Cavalcante
@Comments: at minute 3'13"Lenore clears her throat. Apheretic forms: course (of course), kinda (kind of)

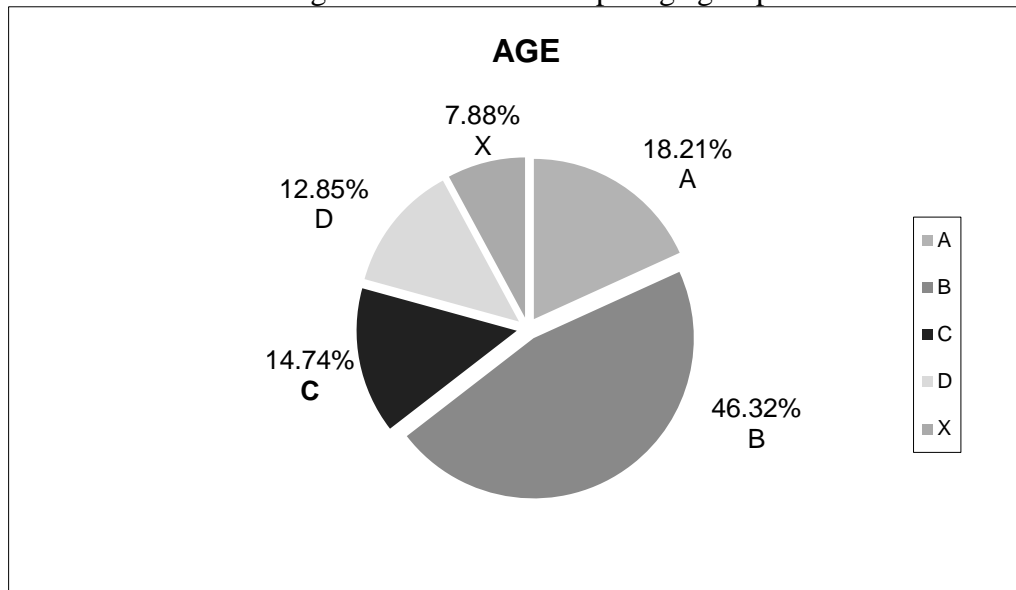
The main objective of the corpus adaptation was to achieve a good diaphasic variation which would be comparable to the variation found in the C-ORAL-BRASIL. Even so, let us look at the diastratic dimension of the subcorpus as it may be interesting for certain studies. The SBC *subcorpus* in its final design has the following distribution of uttered words per gender:

Image 4.1: Words per gender



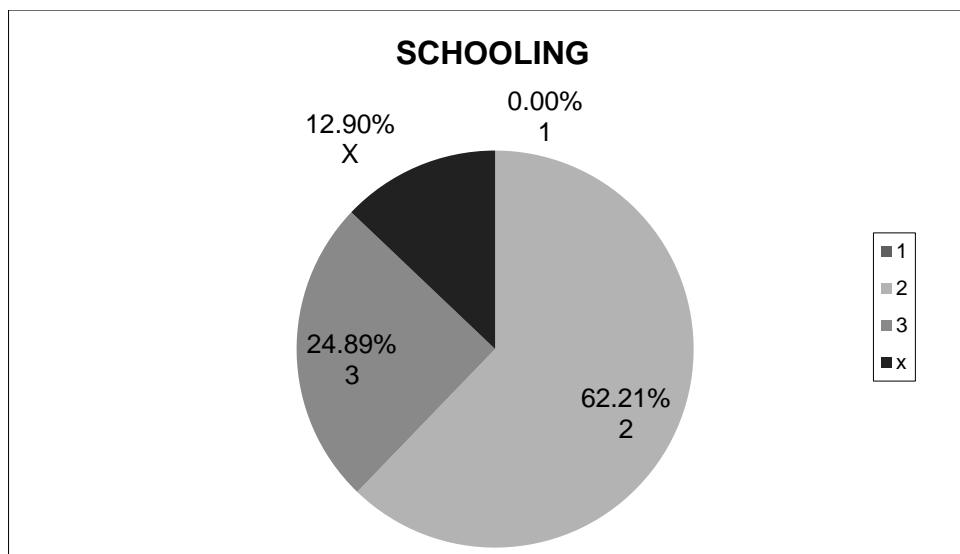
It can be said that when it comes to gender, the distribution is balanced and very close to ideal. The same cannot be said for the distribution of uttered words per age group as presented on the graphic below:

Image 4.2: Uttered words per age group



The pie chart above shows an uneven distribution. Group A, representing speakers between 18 and 25, produces 18.21% of the words in the corpus. Group C and D, symbolizing the age groups 41-60 and 60+ respectively, have similar percentages of uttered words. Group B (26-40 years old), however, is over represented with almost 50% of the total of words in the subcorpus. The other 7.88% of the words were uttered by speakers whose age group is unknown. Now let us analyse the distribution of uttered words per schooling level - in the graphic below:

Image 4.3: Uttered words per schooling level



This uneven distribution does not allow a comparative analysis between the C-ORAL-BRASIL and the SBC when it comes to educational level and its possible influence on the use of adverbs of certainty. The subcorpus is a reflection of the corpus as a whole, thus the same unbalanced distribution is observed on the totality of the SBC. Moreover, the incidence of speakers whose schooling level is not known is high, especially if compared to the C-ORAL-BRASIL<sup>22</sup>. This does not mean that the corpus is not representative of the society it aims to reproduce, but an uneven distribution of the data rules out certain forms of analysis.

Hitherto this dissertation has demonstrated how adapting a corpus to the theoretical and methodological grounds of another corpus can be advantageous to linguistic studies, whether the right conditions exist. The SBC adaptation process represents a big part of my research in Corpus Linguistics, and constitutes a great contribution to the project C-ORAL-BRASIL which was, therefore, considered important enough to be explored in greater detail in part I of this dissertation. During the making of this research, however, I have decided to analyse the whole SBC corpus instead of the adapted minicorpus only, as I noticed that it would be more gainful to the purposes of this dissertation to have a more numerous and detailed account of the behaviour of adverbs of certainty. The second part of this dissertation starts next and deals with the analysis carried out using the corpora presented in this first part of the dissertation, in accordance with the theoretical background of LAcT (Cresti 2000).

---

<sup>22</sup> The numbers for the diastatic variation in the C-ORAL-BRASIL can be found in Appendix 1.

**PART II:**  
**A Semantic/Pragmatic approach to the use of adverbs of certainty**

The second part of this dissertation deals with the analysis of the use of adverbs and adverbial expressions of certainty, with special attention to the functions of expressing epistemic certainty and linguistic politeness. After the definition of adverbs of certainty, I will present a chronological literature review on the phenomena of modality – section 6 -, and politeness – section 7. Section 8 immediately following is devoted to the methodological procedures adopted in the analysis. Section 9 of this part of the dissertation focuses more closely on the data analysis and results. And finally, the concluding section of this dissertation returns to the theoretical premises and objectives presented in the introductory section and highlights the main outcomes of the analysis.

## 5. ADVERBS OF CERTAINTY

In this section I will define what adverbs of certainty are and explain how I arrived at the list of adverbs and adverbial expressions selected for this study. I will begin with the definition of adverbs of certainty followed by a brief review of relevant research on the topic, with special attention to the study that influenced the making of this particular dissertation the most.

Adverbs of certainty, as suggested by the name itself, express a particular degree of certitude. In a scale of probability, one of the ends marking ‘extreme certainty’ and the other ‘doubt/uncertainty’, the adverbs analysed in this paper will express a connotation towards the higher certainty end. To illustrate that definition, let us analyse some examples<sup>23</sup> of adverbs and adverbial expressions that demonstrate the level of certitude of the speaker towards the content of the utterance, extracted from the C-ORAL-BRASIL and SBC corpora.

### (5.1) Mesmo (bpubcv01)

*\*BRU: <deve ser um sacco> mesmo / né //*

### (5.2) Of course (sbc060)

*\*JON: poster //*

*\*ALN: a benefit // it was a poster // that's just [/2] that's not a [/3] that's not what they call a seriograph // the Joe & n [/3] the Joe & he / Namath /<who was of course> an <Alabama> football player //*

---

<sup>23</sup> Audio files in the CD enclosed.

\*JON: <yeah> // <yeah> //

The examples show that the adverbs here studied express a higher level of certainty with different levels of certitude amongst themselves. Each and every adverb or adverbial expression listed in this study will be positioned towards the extreme level of certainty within a probability scale, but they will not necessarily occupy the same position on that scale thereby expressing more or less assurance depending on their use (Simon-Vandenberg; Aijmer 2007:69).

Two semiotic concepts are important when studying adverbs of certainty, namely indexicality and reflexivity. Simon-Vandenberg and Aijmer (2007) discuss those concepts in the book entitled *The Semantic Field of Modal Certainty*:

‘(...) because adverbs of certainty are indexical they can convey information which is only implicit in the context, such as information about social identity, authority, solidarity and politeness. Other indexical parameters are heteroglossic features referring to the way in which speakers take up interpersonal positionings such as countering or challenging. The parameters which are indexed by the adverbs also involve discourse functions and stance (epistemic, affective). Reflexivity explains that adverbs of certainty function as contextualization cues showing how the discourse is organized.’ (Simon-Vandenberg; Aijmer 2007:44)

The passage above exposes the necessity of studying adverbs of certainty in context, once that context sets out the possible roles to be assumed by these naturally multifunctional lexical items. This close relationship between adverbs of certainty and the context in which they occur indicates that the use of those adverbs is not only tied in with the semantic side of the linguistic analysis (modality) but also with pragmatic indexes such as social identity, power relations and nature of the interaction.

Another concept that helps explaining the multiple posts an adverb of certainty can assume in speech is the notion of reflexivity. Simon-Vandenberg and Aijmer (2007) explain that adverbs of certainty have the important role of structuring the discourse while it is being produced by referring to aspects of the context and sometimes modifying it. In Simon-Vandenberg and Aijmer’s words, ‘reflexivity explains that adverbs of certainty function as contextualization cues showing how the discourse is organized’ (Simon-Vandenberg; Aijmer 2007:44). The use of that type of adverbs demonstrates the speaker’s stance in regard

to the preceding or following utterance, or their awareness of ‘how their utterances fit in with preceding or following discourse’

Earlier studies of modality (Greenbaun 1969; Lyons 1977; Perkins 1983, amongst others) tended to focus almost exclusively on the semantic features of the indexes, with little or zero concern for the rhetorical effects of modal choices. Since the late eighties, however, that early view was expanded to incorporate the pragmatic features of those modal markers. In 1987, for example, Coates highlights the polypragmatic function of epistemic modals in informal speech, saying that those markers can convey 5 different communicative goals:

‘(a) saying something on the topic under discussion; (b) being sensitive to the face-needs of the various addressees; (c) qualifying assertions to avoid total commitment to a point of view which they may want to withdraw from; (d) qualifying assertions to encourage the flow of discussion; (e) creating cohesive text’ (Coates 1987:130)

It seems to me that the goals described in letters *a* and *e* deal with the semantic side of modal indexes, whilst letters *b* and *d* outline a pragmatic feature of them, or the social-oriented conduct of being well-mannered. The goal presented on letter *c* seems to set ground in both semantic and pragmatic functions, not only marking the speaker’s commitment towards the uttered content but also the speaker’s concern with the impact of his/her stance on the hearer. Östman (1985), when talking about the study of pragmatic particles, suggests that in order to study those indexes, we should consider three main features, named involvement (stance/point of view), culture-coherence (cultural norms) and politeness. He says that those three parameters ‘provide the necessary ‘context’ with reference to which research into implicit pragmatics is doable’ (Östman 1985:104).

In defiance of what is expected of those kinds of adverbs by definition – to mark the speaker’s commitment to the truth of the proposition – adverbs of certainty have gone through a pragmatic development, becoming not only merely epistemic markers, but also indexes of different rhetorical aims (Simon-Vandenberg; Aijmer 2007:81). Certainty adverbs have a high frequency of occurrence in speech because of their polysemous nature and thereby capability of assuming different functions. With that in mind, I believe that an accurate picture of adverbs of certainty can only be achieved in a heteroglossic perspective, taking into consideration their multifunctional nature associated with type of context, social roles and power relations. That is my concern in the description of the modal indexes that constitute the object of study of this dissertation, and the reason why I chose to study these adverbs in the



lights of a semantic-pragmatic approach, describing who uses those indexes with whom, when and why.

### 5.1. The list of adverbs studied in this dissertation

Assuming adverbs of certainty as those expressing a high degree of probability, or a strong commitment to the truth of what is being said, I arrived at a list that roughly corresponds to the list displayed in the book that inspired the formulation of this study, *The Semantic Field of Modal Certainty*, by Simon-Vandenberg and Aijmer (2007), for British English. Table 5.1 shows the list of the adverbs studied with their frequencies in the corpora analysed, namely C-ORAL-BRASIL and Santa Barbara Corpus of Spoken American English.

Table 5.1: The distribution of the adverbs of certainty in the corpora

C-ORAL-BRASIL		SBC	
Type	Freq.	Type	Freq.
mesmo	114	of course	41
claro	58	sure(ly)	39
com certeza	27	exactly	19
exatamente	24	definitely	18
justamente	7	certainly	10
sem dúvida	5	obviously	10
necessariamente	4	necessarily	6
obviamente	3	evidently	4
logicamente	2	no doubt	4
evidentemente	2	for sure	3
deveras	1	indeed	1
certamente	1		
claramente	1		
de verdade	1		
TOTAL	250	TOTAL	155

The types found are exemplified below, starting with the Brazilian Portuguese ones<sup>24</sup>:

(5.3) Mesmo (bpubdl02)

\*EUG: *porque / o' que borracha boa //*

\*JAN: *é // ela é boa mesmo // melhor a preta do que a branca / né //*

(5.4) Claro (bpubcv07)

\*JOS: *que mineiro que cê conhece que conversa assim / "tu só fala em ir pra roça" //*

<sup>24</sup> Audio files in the CD enclosed.

\*SON: não // **claro** que não //

(5.5) Exatamente (bfammn36)

\*ADE: <morreu> [/1] morreu inclusive uma [/1] uma atriz / té famosa na época / nũ  
[/1] nũ lembro o nome dela //

\*ADR: mas era o [/1] o que [/2] um barco que tinha muito mais gente do que podia /  
era <isso> //

\*ADE: <exatamente> // esse mesmo //

(5.6) Com certeza (bpubdl10)

\*FAB: né / nós nũ &s [/1] <temos / &he> / como se diz / são [/1] são problemas que  
[/1] de outra alçada //

\*ADA: <com certeza> //

\*FAB: <nũ é do [/1] de> professor //

(5.7) Justamente (bpubcv01)

\*EMM: <e // o congelamento e pa preservar> //

\*MAR: pra nao &per + <justamente // pra nao perder> / &he / os fatores de  
coagulacao //

\*BRU: <ah / so> //

(5.8) De verdade (bpubcv06)

\*DEB: <não / mas por enquanto não> hhh / é / por enquanto é //

\*REN: <por enquanto é segredo> //

\*DEB: <mas é [/1] é uma> coisa bem pessoal porque [/1] por causa dos números de  
[/1] de gravidezes que teve / eu me sinto / um pouco responsável // **de verdade** //

(5.9) Sem Dúvida (bfamcv01)

\*GIL: <n' é não> / <o Galáticos> / &e [/1] tá fora // <principalmente depois que  
eles tão> fazendo um campeonato e não nos chamaram / eles <tão fora> //

\*EVN: não / <Galáticos / sem dúvida> //

\*LUI: <e ainda> falaram que o campeonato deles é / “realmente bom” //

\*LEO: <é> //

\*GIL: <é> //

(5.10) Necessariamente (bpubmn07)

\*REC: por exemplo / o meu nome / é Renato // eu posso fazer este movimento // eu  
posso fazer a datilologia do meu nome // mas eu também tenho um sinal // que é esse  
// que é Renato // que não **necessariamente** outro Renato será esse também // poderá  
ser outro // o próprio Gilberto tem um sinal // que é este aqui // e você / como você se  
chama //

(5.11) Evidentemente (afammn36)

*\*ADE: fusquinha bateu de frente com / uma carreta / e o [/1] o Fusca ficou [/1] passou pela / primeira [/1] primeira roda da frente / e ficou / entre a [/1] a [/1] a / segundo / conjunto de roda / e a roda / da frente // ficou no [/1] bem no meio da carreta // achatado / **evidentemente** //*

(5.12) Obviamente (bpubmn13)

*\*JAN: <eu> penso assim // eu / sempre acho que o dinheiro não é um [/2] é um [/1] um fim / é um meio / <não é> //*

*\*MMM: <hum hum> //*

*\*JAN: eu nã coloco / o lucro / como objetivo / e sim / a qualidade / nã é // e o dinheiro vem por consequência // **obviamente** / não é //*

(5.13) Logicamente (bfammn06)

*\*JOR: com as amizades adquirida / que nós chamamos de "network" / &he / me apareceu uma outra / hhh oportunidade dentro de uma outra multinacional / aonde eu fui desenvolver / um trabalho de vendas / &he / junto / ao mercado / concorrente dessa empresa onde eu estava / e lá eu fiquei um período / desenvolvendo o mesmo tipo de trabalho / **logicamente** com um salário melhor / hhh e por amizade eu fui cair / em uma multinacional / que eu dei uma virada no produto //*

(5.14) Deveras (bfamdl29)

*\*ELI: um serviço importantíssimo / esse / tá //*

*\*ALV: sei //*

*\*ELI: se você não achar / eu acho // **deveras** <importante> //*

*\*ALV: <pegar música> para o casamento //*

(5.15) Certamente (bfammn01)

*\*MAI: a cobra tava &en [/2] continuou enrolada nele // **certamente** ea tava querendo fazer o seguinte / eu / eu matei esse / eu vou matar o resto tudo / da [/1] &he / dentro da casa // nã sei / né / a imaginação hhh dum [/1] dum animal / o que que pode ser / né //*

(5.16) Claramente (bfamdl09)

*\*LUC: mas / por exemplo / o Van Gogh / ele [/1] &c [/1] &e [/1] **claramente** / a [/1] as pinceladas / são muito importantes pra ele // qual que é o tipo da pincelada / e tal //*

Now the types found in the SBC corpus for American English:

(5.17) Of course (sbc015)

*\*JOA: he doesn't even eat cake with rum in it //*

*\*LEN: no // **of course** not // no alcohol //*

(5.18) Sure (sbc001)

\*DOR: *it won't last long // but / it **sure** does make a mess in the house //*

\*LYN: *yeah / it <makes a mess> //*

\*DOR: *<I need new filters> //*

(5.19) Exactly (sbc047)

\*FRE: */ and making <your life impossible> //*

\*RICH: *&he / <didn't enjoy doing anything> / she always> was / you know> //*

\*FRED: *<yeah / **exactly**> //*

(5.20) Definitely (sbc002)

\*JAM: *Or what they teach / and what happens socially / could be two different things / you know //*

\*MIL: *I mean / Brazilians **definitely** don't &he /do it like that hhh //*

(5.21) Certainly (sbc016)

\*BRA: *<just let me know> //*

\*TAM: *<and I'll> **certainly** let you know about this thing / &he / I'll be calling you back //*

(5.22) Obviously (sbc060)

\*ALN: *but that's my story on that machine // that damn thing / and of course / I didn't know [/3] **obviously** I didn't know what I was missing / if I'd done [/2] I'd done that hhh //*

(5.23) Evidently (sbc023)

\*LIN: *did you see some charm there //*

\*JAN: *hum hum //*

\*SUE: *uhn //*

\*LIN: *okay //*

\*JAN: *hum hum //*

\*LIN: *hum hum // <this is for the movie> //*

\*EVE: *<she **evidently**> did //*

(5.24) Necessarily (sbc023)

\*LOR: *<no but that's> tragedy // <that's> not **necessarily** passion //*

(5.25) No doubt (sbc004)

\*KAT: *and <that was the only> <way to do it> //*

\*CAR: *<**no doubt**> //*

(5.26) For sure (sbc047)

\*RIC: *at least <it's insured> //*  
\*FRE: *<yeah> //*  
\*RIC: *if anything happens to anybody else //*  
\*FRE: *for sure //*

(5.27) Indeed (sbc024)

\*DAN: *oh / he's giving you some problems over there //*  
\*JEN: *he is **indeed** //*  
\*DAN: *go for that one //*

As already discussed, these adverbs are polypragmatic and can acquire different functions within an utterance, but I chose to concentrate this study on their semantic function of modalizers and pragmatic function of politeness markers. The adverbs and their functions will be explored in further detail on the section dedicated to the data analysis and results. The adverbs which are less frequent (10 examples or less) will be treated more in passing.

Before moving any further, I would like to make some remarks on the list presented. There is, amongst linguists who committed themselves to the study of certainty adverbs, an undeniable difficulty in defining which adverbs may and may not figure within the scope of certitude, resulting in different lists for the same topic. For methodological reasons, I decided to follow the list proposed by Simon-Vandenberg and Aijmer (2007:69)<sup>25</sup> since it is coherent and attends the needs of this research. From the empirical observation of the data collected for this study, however, I made some amendments to their list, which I would like to elucidate.

As stated before, the list of adverbs of certainty analysed in this dissertation roughly corresponds to the list presented by Simon-Vandenberg and Aijmer (2007:69), but I have included some adverbs that are not mentioned in that particular list, such as adverbs of certainty in positions of response (as exemplified above for adverbs such as *exactly* and *for sure*). Since I am adopting a broad perspective of the adverbs of certainty as to especially include their pragmatic function of expressing politeness, including adverbs in the position of response became relevant for this particular study. These adverbs are *exactly* and *sure* in AE, and in BP, *exatamente* and *justamente*. Moreover, due to the multifunctionality and polissemey of some of the adverbs in the list proposed by Simon-Vandenberg and Aijmer (2007), the correspondents in BP at times numbered more than one. In other words, a single adverb in

---

<sup>25</sup> The table can be seen in Appendix 2.

English can have more than one corresponding adverb in Portuguese, possessing the exact same meaning depending on the context. I find this to be the case of *justamente*, for instance, which is one of the possible translations for *exactly* in its epistemic form. Similarly, *of course* can assume both the forms of *logicamente* and *mesmo* in BP, amongst other forms. This explains why there are more types in BP than in AE: at times, there is more than one translation for a particular adverb, one that corresponds to a prototypical equivalent and another one that corresponds to an equivalent which is more appropriate in some contextual uses.

## **6. THE CATEGORIES OF MODALITY AND EVIDENTIALITY**

In the previous section we discussed the classification and properties of the adverbs and adverbial expressions that fashion the object of study in this dissertation. It was pointed out that drawing strict boundaries when it comes to the classification of adverbs is a challenging task, considering the several pragmatic functions that these adverbs can assume.

The adverbs and adverbial expressions studied in this dissertation are inherently epistemic as they are naturally connected to the notions of certainty, and this is the reason why they are often regarded as evidentials. In this section I will place the adverbs presented within the general discussion of modality and briefly discuss their relationship with evidentiality, by presenting a broad overview of the most relevant frameworks available and positioning the adverbs studied within one of them.

### **6.1. Modality**

Given the vast list of publications on the topic, a concrete definition of what modality may or may not be is still pending. In this section I will explain the reasoning behind the most popular and disseminated interpretations for the phenomena and call attention to some issues encountered with the use of certain terms.

Throughout the time I have studied modality, I have noticed that the definitions proposed so far tend to refer back to one common concept. For this reason, I believe that the best way to comprehend modality as a whole may be by understanding its intrinsic roots and the possibilities that emerge from them. Accordingly, I will present a chronological description of this semantic phenomenon – modality - as to situate this dissertation within the framework of the general discussion on the topic.

The notion of modality first appeared in a non-linguistic environment, within the domain of Philosophy and Logic. The notion of proposition is an essential aspect for the description of modal logic in the philosophical tradition. A proposition is a concept used to refer to the content or meaning of a purposeful declarative sentence, and has the quality of being either true or false (McGrath 2014). This definition follows the Aristotelian interpretation that a

proposition is any given sentence that affirms or denies the predicate of a subject (Aristote 2007). With those definitions in mind, it is easy to identify the deep connection that modality has with the notion of truth values in the philosophical sense. According to von Wright (1951), the propositional trait of bearing the truth has four different modes: modes of truth, modes of knowing, modes of obligation and modes of existence. The characteristics of these modes are summarized in the table below for better visualization<sup>26</sup>:

*Table 6.1: Propositional modes*

<b>Modes of Truth Alethic</b>	<b>Modes of Knowing Epistemic</b>	<b>Modes of Obligation Deontic</b>	<b>Modes of Existence Existential</b>
Necessary	Verified	Obligatory	Universal
Possible		Permitted	Existing
Contingent	Undecided	Indifferent	
Impossible	Falsified	Forbidden	Empty

The existential mode is not considered as a branch of modal logic, but as a trait of the proposition. Only the first three modes apply to modal logic, whereas the last one deals with the logic of truth concepts, or truth logic. Modal and truth concepts, according to von Wright (1951:6), are different but not logically disconnected. In modal logic, a proposition cannot assume two different characteristics of the same mode at the same time, being, for example, verified and falsified simultaneously. A property can be, however, universal and empty at the same time, which means that two different traits of truth concepts can coexist concomitantly.

Following this first consideration of modality in the lights of the Philosophy, some authors started to study it from a linguistic perspective. Jespersen (1924) was one of the first authors to take modality to the linguistic domain and stated that modality is concerned with ‘expressing a certain attitude of the speaker towards the content of sentences’<sup>27</sup>. About a decade later, Bally (1932) introduced the concepts of *modus* and *dictum* to his definition of modality, defining it as ‘the linguistic form of an intellectual, emotive or volitive judgment [*Modus*] that a rational being utters [*Dictum*] on a perception or a representation of their conscience’<sup>28</sup>. Aligned with that notion, Kiefer (1994) claims that modality is ‘the speaker’s cognitive, emotive or volitive attitude towards a state of affairs’ (Kiefer 1994:2516). In 1977,

<sup>26</sup> VONWRIGHT 1951:6

<sup>27</sup> JESPERSEN 1924:313

<sup>28</sup> ‘Modalité est la forme linguistique d’un jugement intellectuel, d’un jugement affectif ou d’une volonté qu’un sujet pensant énonce à propos d’une perception ou d’une représentation de son esprit’. (BALLY 1932). My translation.



Lyons suggested that modality is ‘the speaker’s opinion or attitude towards the proposition that the sentence expresses or the situation that the proposition describes’ (Lyons 1977:452). Later on, Palmer (1986) studied modality within the realm of real and unreal (realis and irrealis<sup>29</sup>), and approached his analysis based on that binary distinction. He defined modality as ‘the grammaticalization of speaker’s subjective attitudes and opinions’ (Palmer 1986:16). Bybee and Fleischmann (1995) say that ‘when the proposition of an utterance in the most neutral semantic status, i.e. factual or declarative, is subject to further addition or overlay of meaning, this extension represents modality’ (Bybee; Fleischmann 1995:2). A more recent work conducted by Nuyts (2001) presents modality from a functional and cognitive perspective, stating that he is interested in how ‘humans conceptualize the world and how linguistic structure and processing relate to that’, from an empirical observation of ‘functional structure of modal expressions in all its facets’ (Nuyts 2004:139). Nuyts defines (epistemic) modality as ‘a speaker’s evaluation of the likelihood of a state of affairs, as expressed in language’ (Nuyts 2001:16). In a broader perspective of the phenomenon, Tucci (2009) describes modality as ‘all kinds of subjective/evaluative meaning’ (Tucci 2009:1).

At this point, bewildered by the many different definitions of modality in all its different guises through the years, it is clear to identify what Nuyts once called ‘the modal confusion’<sup>30</sup>. The problem with defining the semantic category of modality demonstrates the complexity surrounding the theme, and the puzzling use of ambiguous terms can only accentuate the issue. The term ‘proposition’, for example, or its counterpart ‘state of affairs’ is often used in the definitions of modality, as observed above. This characterizes a problem, as, according to Tucci (2009), propositions do not describe the scope of modality in spoken language, which is organized in utterances. In her own words, ‘the scope of modal index cannot be set on hypothetical *propositions*, as the written idealization of language leads us to do, but has to be considered within the lexical, morphological and semantic domain of the informational structure of a sentence.’ (Tucci 2009:2).

Similarly, it is very common to come across definitions of modality with the notion of attitude attached to it. This term also needs disambiguation. When we talk about modality and

---

<sup>29</sup> ‘The realis portrays situations as actualized, as having occurred or actually occurring, knowable through direct perception. The irrealis portrays situations as purely within the realm of thought, knowable only through imagination.’ (PALMER 2001:1)

<sup>30</sup> See NUYTS 2005. The modal confusion: On terminology and the concepts behind it.

attitude, we are dealing with two different sets of phenomena, even though those terms are broadly used as synonyms or complements. There is a need to disambiguate those terms and detach their meanings from one another in order to solve the modal puzzle. Based on a recent empirical and experimental work developed by Mello and Raso (2011), modality and attitude are to be considered instances of different phenomena. Attitude is the expression of social effects controlled by the speaker, and can affect both the locutive content of an utterance - irony and expression of surprise, for example - and the relationship between interactants in any given communicative event – i.e. attitudes of arrogance and courtesy (Moraes et al. 2010). Modality, on the other side, does not concern the relationship speaker-hearer. It concerns the judgement of the speaker towards the locutive material of the utterance (Mello; Raso 2011). Moreover, while attitude is prosodically marked, modality does not occur through prosodic cues. In summary, we can say that the same locutive material can be modalized in many different ways and can likewise express many different attitudes. Thus modal and attitudinal markers can happen simultaneously in any given utterance as they appertain to different levels and should therefore be disassociated.

Cresti and Firenzuoli (2001) claim that it is necessary to distinguish modality from illocution, as the concepts are also mixed at times. According to the authors, illocution is the attitude of the speaker towards the hearer, or the action that this speaker completes from the content of the enunciation. These linguistic actions can be understood as complaints, confirmations, and as advice, to quote a few. Illocutions are prosodically marked in part of the utterance, differently from attitudes, whose prosodic cues can be observed in the utterance as a whole. The authors also state that modality is the evaluation of the speaker towards their own verbalization and does not depend on the relationship between speaker and context. Modality, therefore, belongs to Semantics, as it characterizes a cognitive process of representation linked to a referee and not to a linguistic action, in opposition to attitude and illocution, phenomena that belong to Pragmatics. The notion of illocutions will be explored in further detail in the next section of this dissertation, once it is deeply linked to the study of politeness.

When dealing with modality, we come across an important paradox initially suggested by Lyons (1977). The author stated that the non-modalized assertions express a greater commitment from the speaker than the modalized ones. In other words, a speaker will only make use of epistemic markers if the truth of what is being said cannot be verified in the

discourse as obvious. Halliday (2004) resumes that concept when affirming that ‘we only say we are certain when we are not’ (Halliday 2004:147). The linguist argues that if a speaker unconsciously considers his assertion to be true, he will simply utter it without making use of any modal marker, i.e. *Maria has left*<sup>31</sup>. However, if the same speaker was to add a modal marker such as an adverb of certainty – *Maria has definitely left* -, that means that the speaker admits a dubious element and tries to correct that inaccuracy by emphasizing his/her expression of certainty. To further illustrate that paradox, let us observe the following example extracted from the C-ORAL-BRASIL corpus.

(6.1) \*MAI: *a cobra tava &en [/2] continuou enrolada nele // certamente ea tava querendo fazer o seguinte / eu [/1] eu matei esse / eu vou matar o resto tudo / da [/1] &he / dentro da casa //*

The passage is part of a recording in which the main speaker is telling the story of a snake from the north of Minas Gerais that is supposedly capable of eating human beings. The speaker here uses the adverb *certamente* to validate the truth of what he affirms to be what the snake was thinking. This is a clear example that shows how speakers can say they are certain about something that in reality constitutes a matter of opinion, or, as demonstrated above, something clearly impossible to be verified in the real world.

## 6.2. Evidentiality

Another issue encountered by linguists is coming to terms with the domains of modality and its relationship to evidentiality. Deciding if they are instances of the same phenomenon or two distinct phenomena is a theme for many linguistic researches nowadays. The source of the issue comes from the fact that both evidential and modal markers can indicate the speaker’s commitment to the truth of what is being said, with only one variation: the mode of knowing or the source of the information.

To Saeed (1997:133), evidentiality is ‘a term for the ways which a speaker qualifies a statement by referring to the source of information’. Chafe (1986) assumes an all-embracing definition and puts evidentiality in one category together with epistemic modality.

---

<sup>31</sup> Extracted from Halliday 2004:625.

Throughout the years, many different views on evidentiality and its precise relationship to modal categories were presented. I will briefly present a summarized list of those views:

- I. Evidentiality and epistemic modality in one category: Bybee (1985), Chafe (1986), Palmer (1986);
- II. Evidentiality closely related to epistemic modality, both adjoined under one modal subcategory in a privileged relationship: Hengeveld (1989), Palmer (2001), Simon-Vandenberg & Aijmer (2007), introducing the notion of epistemological modality, propositional modality, and evidential modality, respectively;
- III. Evidentiality separate from epistemic modality but also modal: Cornellie (2009) and Narrog (2005);
- IV. Evidentiality not modal: Anderson (1986), Bybee et al. (1994), Saeed (1997) and Haan (2006).

My intention with this brief list of the most influential views on the relationship between evidentiality and modality is to argue that evidentiality is not a consistent category. However, it is noticeable that both categories are clearly closely related. For the purposes of this dissertation, I will follow a broad view on the relationship of the phenomena and abstain myself from the discussion on the precise definition of both categories for now. When a modal marker has the characteristic of qualifying the source of information, that feature will be stated as so, in accordance to what is proposed by Simon-Vandenberg & Aijmer (2007). In this dissertation I shall use the notion of modality to denote the expression of the speaker's judgement (*Modus*) towards the referential or cognitive content (*Dictum*) of the utterance (Cresti 2000, 2001, 2002).

## 7. POLITENESS

Many different strategies are used by speakers in order to avoid threatening the social-oriented conduct of being polite. The linguistic phenomenon of politeness is susceptible to cultural elaboration: a particular society's values and ideas have a direct impact on the views of what constitutes 'good' behaviour in communication. In other words, an individual is judged as more or less polite depending on the standard of a certain community (Leech 1983). English, for example, might sound a more polite language to a Brazilian due to the use of certain politeness markers, whilst the use of the same politeness markers for a native speaker of English is regarded as conventional behaviour.

Politeness can be marked in speech through many different linguistic features. Speakers make use of verbs, adjectives, honorifics and other lexical items in order to avoid being rude. Modal adverbs of certainty have also developed this pragmatic function: speakers use them as devices in order to protect their public self-image.

The next section introduces the theoretical background to the analysis developed. A general overview of the most traditional perspectives developed throughout the years will be presented to describe the phenomenon of politeness, with special attention to the concepts and methodology considered in the making of this research.

### 7.1. Traditional theories of linguistic politeness

When analysing strategies used by speakers in the expression of politeness, we come across many different theoretical accounts that aim to describe the phenomenon. It is safe to say, however, that these theories have something in common: they are couched in the general Gricean framework. Some of them attempt to revise the theory, while others presuppose Grice's contribution and add a framework to it. In this section I will introduce two main theoretical models for the description of politeness, giving special attention to the one that will be adopted in the analysis proposed in this dissertation (Leech 1983). Before this however, I will present a brief literature review on the assumption that represents the common ground between both politeness models, the Cooperation Principle (Grice 1975), followed by the notion of illocutions, which is intrinsically related to the study of courtesy.

### 7.1.1. Grice (1975)

The author claims that there is a way language is used among speakers: they have to cooperate so they can communicate. He labelled this assumption the Cooperative Principle and proposed four maxims based on the following general formulation:

“Make your conversational contribution such as it is required at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged”<sup>32</sup>

The maxim of Quantity states that the speaker should make his/her contribution as informative as is required. The maxim of Quality claims that the speaker should not say what he believes to be false. According to the maxim of Relation, the speaker`s contributions must be relevant. The fourth and last maxim – Manner – states that the speaker should avoid obscurity of expression, ambiguity and unnecessary prolixity. The maxims represent what speakers should do in order to express themselves efficiently in an interaction. In other words, they do not classify rules one has to obey. Instead, they are logical principles for people to consider if they want to communicate a clear message.

As previously stated on part I of this dissertation, speech acts play a crucial role not only in LAcT (the Language into Act Theory) but also in the theoretical assumptions regarding the phenomenon of politeness. Since speech acts were already explored on the section dedicated to the presentation of LAcT, I will now focus on the illocutionary level of those acts, and point out its importance to the study of politeness

### 7.1.2. Illocutions

According to Austin (1962) illocution is what speakers do when they speak or the communicative intention of speakers when they say something. Searle (1969) systematized this notion and proposed five categories for illocutions, namely assertives, expressives, commissives, directives, and declaratives. The first category embodies the acts with which the speaker reveals his commitment to the truth of what is being said. In the second category, speakers reveal their emotions towards what is being said (i.e. congratulations and apologies). The commissives show the speaker`s commitment to future actions (i.e. promises). When it comes to the directives, the speaker asks the hearer for a particular action (i.e. asking and giving instructions). In the last category – the declaratives - the speaker changes reality by

---

<sup>32</sup> GRICE (1975) p.26.

saying something - this is what happens in baptisms and marriages, when the priest changes the previous reality by declaring a child as baptised or a couple as husband and wife.

Cresti (2000) hones the idea of both authors yet further by stating that, when studying illocutions, two aspects must be taken into consideration: the pragmatic and the linguistic entities. The former is the linguistic act, whilst the latter constitutes the utterance. The relationship between them is marked by intonation. According to Cresti (2000), illocution is the attitude of the speaker towards the listener. Cresti and Firenzuoli (2001:5) suggest a new taxonomy for the representation of speech acts classes and types. The authors claim that the Searlian framework cannot be applied to the concrete domain of speech, as it is not enough to capture real data. This happens because the Searlian taxonomy ignores speech's real characteristics of production since it has a logical lexical nature rather than an empirical prosodic/pragmatic one. Hence, from an empiric perspective, they suggest five illocutionary classes that roughly correspond to the Searlian taxonomy, but, due to having a prosodic-pragmatic nature, they present subcategories that better translate what can be observed in spontaneous speech (Cresti; Firenzuoli 2001:7):

*Table 7.1: Corpus based reference table of speech acts classes and types*

<b>REPRESENTATIVES</b>	<b>DIRECTIVES</b>	<b>EXPRESSIVES</b>	<b>rites</b>
Concluding	Distal recall-not visible object	Exclamation	Thanks
Weak assertion	Distal recall- visible object	Expression of contrast	Greetings
Answering	Proximal recall	Expression of obviousness	Apologies
Commentary	Distal deixis	Softening	Welcome
Strong assertions	Proximal deixis	Expression of surprise	Congratulation
Identification	Presenting (object/event)	Expression of fear	Wishes
Verification	Introducing (person)	Expression of relief	Compliments
Claim	Request of information	Expression of uncertainty	Declaration of legal value
Hypothesis/ supposition	Request of action	Expression of doubt	<i>condemnation</i>
Explanation	Order	Expression of certainty	<i>condolences</i>
Inference	Total question	Expression of wish	<i>baptism</i>
Definition	Partial question	Expression of disbelief	<i>promise</i>
Narration	Alternative question	Expression of pity	<i>bet</i>
Describing	Request of confirmation	Irony	
Quotation	Reported speech	Complaint	<b>REFUSALS</b>
Objection	Announcing	Imprecation	
Confirmation	Advising	Insinuation	
Approval	Warning	Derision	
Disapproval	Suggesting	Provocation	
Agreement	Proposal	Reproaching	

Disagreement	Recommend	Hint
	Invite	Encouragement
	Prompt	Assuring
	Permit	Threatening
	Authorize	Giving up
	Prohibition	
	Instruction	
	Regret	

The variety of types of conventional activities allowed by the linguistic system is to be found within the main classes of representatives, directives, and expressives, which outnumber the subsequent two categories therefore containing the relevant variation.

Leech (2005) points out that the politeness phenomenon is oriented by the communicative aims of the speakers, which can come from an illocutionary or social demand. The illocutionary aims are the primary ambitions we have when communicating. We may want to convince someone of something, for instance. In alignment to these objectives, speakers also aim to maintain a harmonic relationship between them and their addressees, which would characterize a social demand. Both illocutionary and social objectives are often in agreement. For example, when a speaker compliments the hearer, the illocutionary aim of giving a compliment suits both the social objective of being polite and maintaining a good interaction with the hearer. There are, however, situations in which these targets are misaligned. In a situation in which the speaker wants to criticise the hearer, for example, the illocutionary aim automatically harms the social aim of maintaining concord. In the next section we will see that the illocutionary and social aims constitute the basis for the concept of politeness in its positive and negative polarities.

### 7.1.3. Leech (1983)

The author suggests a principle with its own set of maxims to complement Grice's Cooperative Principle (CP). In the book entitled *Principles of Pragmatics*, Leech introduces the concept of Politeness Principle (PP).

Leech says the CP does not explain why people are often so indirect when conveying what they mean. According to him, politeness is a principle to be added to the CP since it is given a higher rating than cooperation in certain situations, in different societies. The author supports this argument by giving many examples. One of them demonstrates a very common strategy, when a speaker leaves behind the maxim of quality in order to avoid being impolite: the



speaker chooses to tell a 'white lie', declining an invitation by saying that s/he has another engagement. Leech claims that in many situations, politeness principles overrule the CP: if a person has to choose between being cooperative and being polite, s/he will most likely choose to be polite. According to the author, the PP explains why people seem to be less informative, less truthful, less relevant, less clear and less brief in many occasions: mainly because they want to be polite.

Later on, in the paper entitled *Politeness: Is there an East-West divide?* (Leech 2005), the author restates some of the notions discussed on his first attempt to describe the phenomenon, in response to the criticism his work had received. First of all, he emphasizes that the PP is a goal-oriented behaviour and has no intention of being prescriptive. According to the author, the PP postulates a set of social goals which co-exist with illocutionary goals such as informing, influencing, praising, criticising, and boasting, amongst many others. He also points out that, differently from other authors, he never committed himself to a universalist position, since he understands politeness as a social dependable phenomenon. Considering the criticism his work had received, Leech restates his Principle of Politeness based on studies he carried on after the publication of the PP in the 80's. This time, the author says the PP is analogous to Grice's CP, and is 'a constraint observed in human communicative behaviour, influencing us to avoid discord or offence, maintaining concord' (Leech 2005:6). The main change lies on the fact that the PP must not be taken as a prescription of the politeness phenomenon, but rather a constraint on a goal-oriented communicative behaviour that co-exist with other communicative goals.

Leech defines politeness as forms of behaviour that establish and maintain comity, which is the ability that the participants of a given social interaction have to engage in interaction in an atmosphere of relative harmony. The linguist claims that the notions of truth and politeness are very frequently in conflict, as they are not always compatible. Speakers tend, for example, to agree with their hearers even though they clearly do not, to avoid discord. He considers the phenomenon of politeness to be relative, varying in many different dimensions, depending on the social standardised rules of a certain community. To the author, politeness is a matter of degree, and determining the appropriate degree of politeness in a given society depends on other scales of value (Leech 2005:19), which he describes as:

- a. Vertical distance: the social distance between interactants (i.e. status, power, role and age, amongst others);
- b. Horizontal distance: the type of relationship shared by the speakers (i.e. intimate, familiar, acquaintances, stranger, etc);
- c. Weight or value, i.e. how large is the benefit, cost, favour and obligation, amongst others;
- d. Strength of socially defined rights and obligations;
- e. Self-territory and other-territory.

His system of politeness constraints are coherently articulated with the notions of cost (negative politeness) and benefit (positive politeness), which interact with the notion of illocution. On the matter, Leech says that some illocutions (i.e. orders) are inherently impolite, and others (i.e. offers) are inherently polite. These illocutions will ask for different politeness strategies, either a face enhancing act (a positive construction of increasing the estimation) or a face threatening act (intended to lessen the degree to which the speaker's goals are imposed on the hearer). This suggests that depending on the cost or benefit inherent to the illocution, the speaker will choose a politeness strategy to maintain the goal of comity. Negative politeness is employed to minimize the lack of courtesy of a naturally impolite illocution and the positive polarity of the phenomenon is used to maximise politeness in inherently polite illocutions. In other words, negative politeness deals with the avoidance of discord whilst positive politeness seeks concord between the parts, in accordance with the nature of a given illocution.

The Politeness Principle offers a series of maxims<sup>33</sup>, which Geoffrey Leech has proposed as a way of explaining how politeness operates in conversational exchanges. These constraints come in pairs, and are presented in the chart below:

---

<sup>33</sup> In 2005 Leech restates the notion of maxims and says he decided to avoid the term because it implies a moral imperative, which does not reflect the nature of his PP. He started using the term *constraints* instead. (LEECH 2005:16)

Table 7.2: PP Maxims

<p><b>Tact</b></p> <p><i>Minimise cost to other</i></p> <p><i>Maximise benefit to other</i></p>	<p><b>Generosity</b></p> <p><i>Minimise benefit to self</i></p> <p><i>Maximise cost to self</i></p>
<p><b>Approbation</b></p> <p><i>Minimise dispraise of other</i></p> <p><i>Maximise praise of other</i></p>	<p><b>Modesty</b></p> <p><i>Minimise praise of self</i></p> <p><i>Maximise dispraise of self</i></p>
<p><b>Sympathy</b></p> <p><i>Minimise antipathy between self and other</i></p> <p><i>Maximise sympathy between self and other</i></p>	<p><b>Agreement</b></p> <p><i>Minimise disagreement between self and other</i></p> <p><i>Maximise agreement between self and other</i></p>

As it can be observed, there are 6 maxims and they occur in pairs. The first column shows hearer-oriented maxims and the second column displays the speaker-oriented ones. We can also observe that the first submaxim illustrates the negative side of the courtesy phenomenon whilst the second submaxim represents the positive side of it. Moreover, it can be seen that the second submaxim corresponds to a natural implicature of the first. The tact maxim states that the speaker should not be inconvenient to the hearer, and, according to the author, it is more powerful than the generosity maxim which states that the speaker should be helpful when it comes to the hearer's needs. The approbation maxims says that the speaker should avoid disagreement and praise the hearer, and it is more important than the modesty maxim which affirms that the speaker should be modest and even dispraise him/herself. The last pair is the sympathy/agreement maxims. The first states that the speaker should be caring and supportive of the hearer's needs while the second maxim of the pair claims that interactants should avoid disagreement or be less direct when expressing it.

In summary, politeness in this theoretical frame is understood as the set of social norms established by each society that aims to adjust the adequate behaviour of its members, conventionalising manners and conduct. What is adequate to the standard norm is considered polite and what is not is labelled as impolite. He emphasizes that politeness is an asymmetrical phenomenon, meaning that what is polite to the speaker may be impolite for the hearer and vice-versa. He also suggests that cross-cultural variability will lie in the relative importance given to one of the maxims facing another (Leech 1983:107).

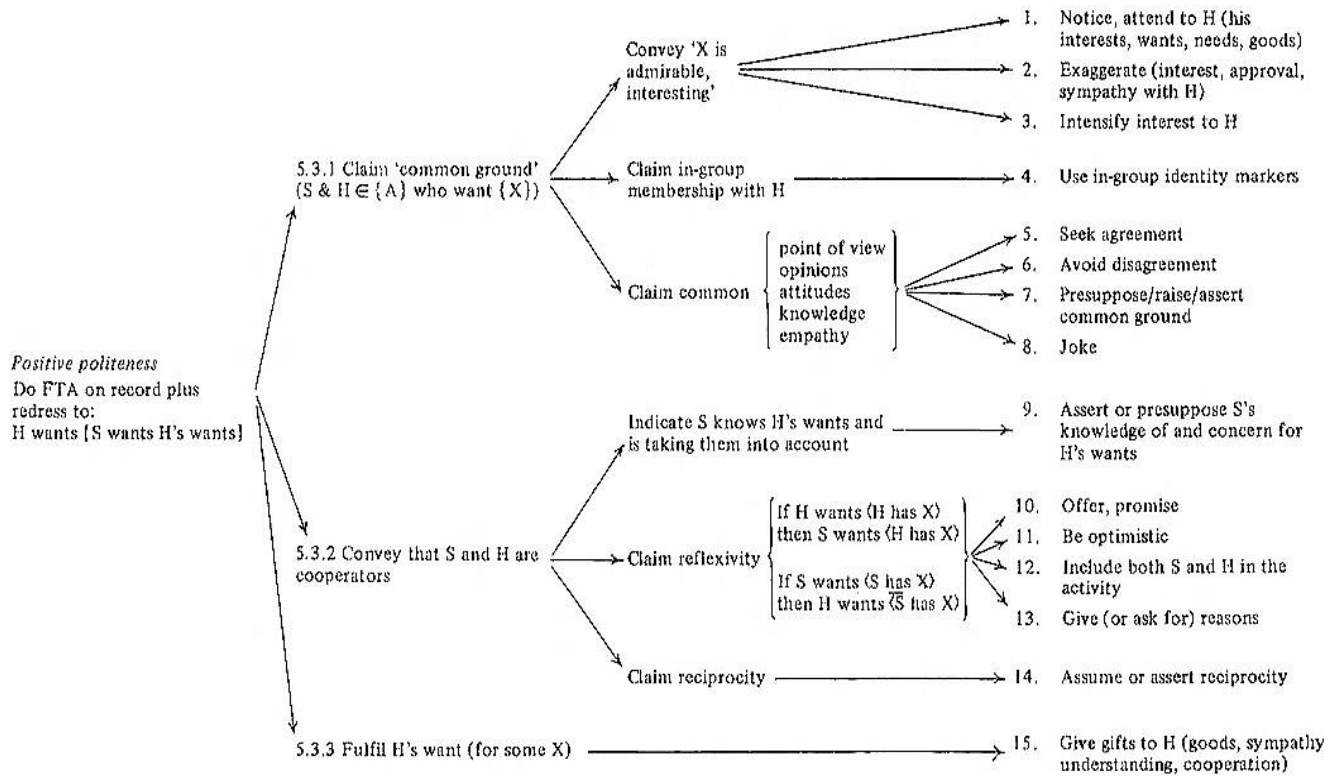
#### 7.1.4. Brown and Levinson (1987)

In the book entitled *Politeness: some universals in language usage*, the authors use the notion of face to explain the phenomenon of politeness in communication. Face is defined in their framework as the ‘individual’s self-esteem’, or, in other words, ‘the public self-image that every member of society wants to claim for himself’ (Brown & Levinson 1987:2). Brown and Levinson refine the notion of face by saying that it is universal but linked to cultural values, such as social honour, virtue and religious concepts. They say that, although the notion of face is considered to be highly abstract, it explains the main desires attributed by speakers to one another: the desire to be approved of and the desire to be free of imposition. The writers call those desires the interactant’s face wants. According to them, the strategies people use in order to protect their face wants add up to the notion of politeness: they argue that in human interaction, people tend to maintain one another’s face continuously, and politeness strategies are used when the notion of face is threatened. The face threatening acts (FTAs) are acts that infringe on the speaker’s face wants, or their need to maintain their self-esteem and be respected.

Levinson and Brown sum up politeness behaviour in four strategies: bald on record, positive politeness, negative politeness, and off-record-indirect strategy. The bald on record strategy does not use any politeness marker in the proposition to minimize the threat to the hearer’s face wants. The second strategy shows that the speaker recognizes he/she might threaten the hearer’s face, and therefore that this hearer has a face to be respected. Negative politeness recognizes the hearer’s face and also presupposes some level of imposition on him/her. Off-record indirect strategy is used when the speaker wants to avoid a direct face threatening act by being indirect. The authors present some techniques within these main strategies in order to illustrate how speakers lexically mark politeness in discourse. For succinctness, only the notions of *positive* and *negative politeness* will be discussed in more detail, as they also appear in other models.

According to the authors, positive politeness is the redress directed to the addressee’s desire that his actions and wants are considered desirable. With that aim in mind, the speaker tries to communicate his own wants as they were similar to the hearer’s wants. The linguistic realisations of positive politeness can be seen in the chart below (Brown & Levinson 1987:107).

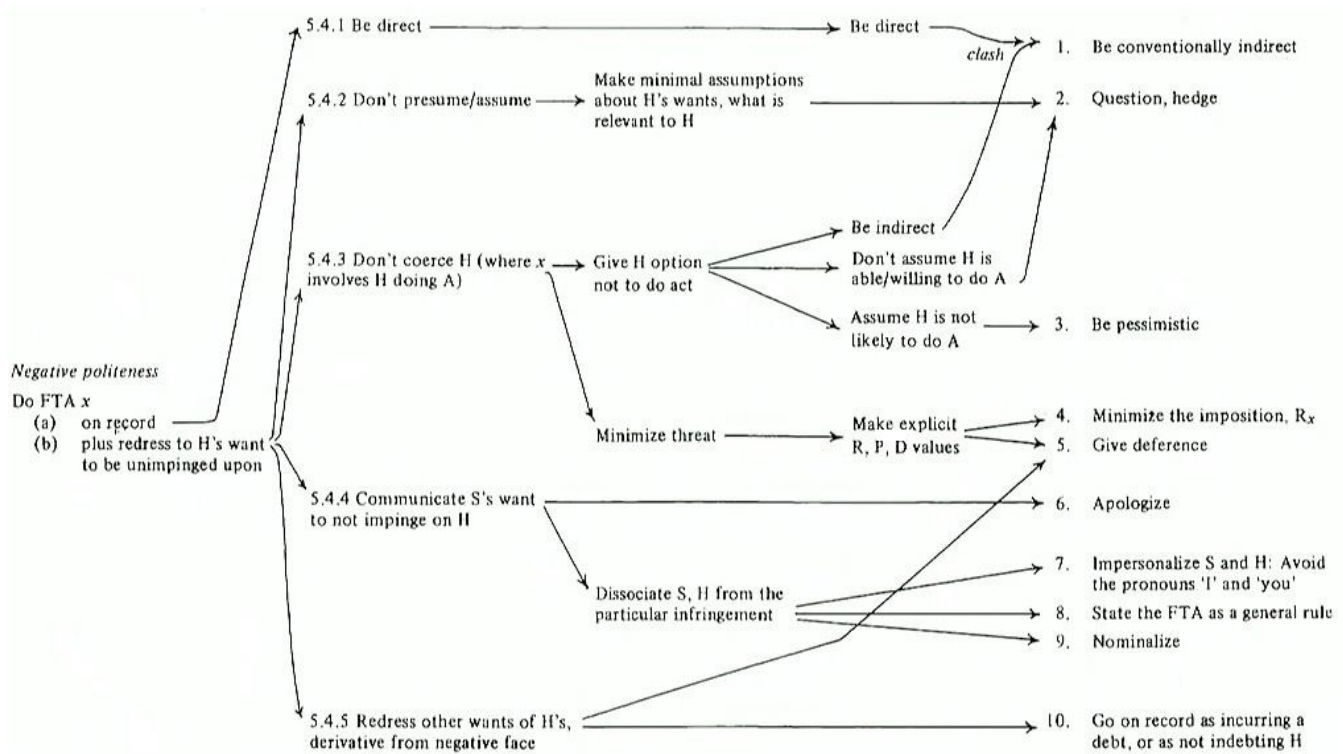
Image 7.1.: Chart of strategies: positive politeness



The image shows a scheme with the possible realisations of positive politeness, used by the speaker with redress directed to the hearer's positive face. There are 15 strategies listed in total, which appeals to the hearer's desire to be liked and to be approved of.

In negative politeness, the redress lies in an area restricted by imposition itself. The speaker identifies a potentially face threatening act and shapes language so as to avoid a particular imposition that the FTA invariably offers. In the chart below it is possible to observe the strategies used by speakers in the expression of negative politeness, identified by B&L (1987:108).

Image 7.2: Chart of strategies: negative politeness



The list for negative politeness strategies is not long when compared to the list presented before. The strategies here counteract towards the hearer's desire not to be impeded or to be left free to act as s/he likes. The speaker will therefore use strategies such as apologising, giving deference and avoiding direct personalisation.

According to Brown and Levinson (1987), three sociological factors are very important notions to be considered when studying politeness strategies. They affirm that the level of politeness which a speaker will use to an addressee depends on what they call 'P', 'D' and 'R', or, respectively, relative power of hearer over speaker, social distance between interactants and ranking of imposition involved in the doing of the face threatening act. The authors say there are other factors that can affect the choice of what politeness strategies to use in interaction, but, for cross-cultural studies, those three sociological factors are enough to explain politeness assessments. One of the objectives of B&L when studying the phenomenon is to explain the cross-cultural conflicts aroused from the discordance on what constitutes 'good behaviour' in communication.

The highly abstract notion of face has been strongly criticised by some linguists<sup>34</sup>. Leech (2005) objected that Brown and Levinson's model does not present 'universals' of politeness since it is based on a Western notion of individual *wants*, and therefore cannot claim to present a theory that covers all languages and cultures, regardless of the natural observation that cultural differences do play an important role in the perceptions of social distance between speakers, which is, on its own, key in choosing politeness strategies in interaction. For instance, the difference in the relationship between teacher and students is considered formal in North America, but the same is not true in Brazil. The social distance that students feel toward their teachers varies from culture to culture. Furthermore, B&L's model lacks in succinctness: their model is not as economical as others. This can be seen as a problem since there is a widespread philosophical presumption that simplicity is to be seen as a theoretical advantage. The logical principle known as *Occam's Razor*<sup>35</sup> states that when choosing competing hypotheses, the postulate with the fewest assumptions should be the one selected. For this dissertation I chose to follow the theoretical framework proposed by Leech (1983) with his Politeness Principle, firstly because it is succinctly presented and secondly because I believe Leech's model is more suitable to comparative studies as it does not presuppose social universalism. Leech himself (Leech 2005:5) reveals his hopes toward the application of his model to contrastive studies, as he believed crosslinguistic analysis could add important insights to the theoretical discussion of the phenomenon.

---

<sup>34</sup> See LEECH, 2005.

<sup>35</sup> 'Most philosophers believe that, other things being equal, simpler theories are better' (BAKER 2004). See BAKER 2004 for further information, available at <http://plato.stanford.edu/entries/simplicity/>

## 8. METHODOLOGY

### 8.1. Corpus Linguistics

Corpus Linguistics (CL) is a methodology for approaching the study of language (McEnery & Wilson 2001). This method has proved to be a valuable tool within the general discipline of Linguistics, as it provides the researcher with the power to search, process and analyse a vast and complex set of data. CL is not only concerned with what is “possible” within a language, but also and mainly with what will probably occur given the observations made in a corpus. Under these circumstances, it can be said that CL sees language as a probabilistic system, since large amounts of data tell the empiric researcher about tendencies and what is typically found in real-life language use. The linguist Manning (2003) provides a good insight into the nature of probabilistic linguistics. The author uses a quotation from Sapir - “Everyone knows that language is variable”- as a starting point for arguing in favour of probabilistic linguistics. In the first section of the article, the author makes some general criticism of the Chomskyan approach to Syntax, saying that this tends to ignore the fact that language usage varies in a number of ways, especially from speaker to speaker. Following this concept, he gives two main reasons why probabilistic models are important and helpful in language studies. These are:

- 1) Linguistics can be seen as a cognitive science, and human cognition is probabilistic by nature: humans have to make meaning out of incomplete and uncertain information about the world on a daily basis, often relying on probabilities for this.
- 2) Understanding the relationship between form and its meaning conditioned by context is more complex than just knowing whether a sentence is grammatical or not.

Corpus Linguistics looks at language as it is, and is therefore qualified as a combination of bottom-up with top-down strategies (McEnery & Hardie 2012): processing data leads to facts and new trends (bottom-up), that can confirm or not the researcher’s hypothesis and intuition (top-down).

CL cannot, however, claim to completely describe the language as a whole or have the ambition to answer every question in Linguistics. If we were to draw an analogy, we could



say that CL has to be seen as a tool, and just like any other tool it is suitable for certain purposes and not so suitable for others (McEnery & Wilson 2001). Manning (2003), in the same article, also presents some problems in working with an empirical approach in linguistics investigations. The author claims that the linguist must have considerable computer skills, otherwise dealing with large amounts of data becomes unfeasible. Secondly, Corpus Linguistics deals with the major problem which is the scarcity of data. They can search through a large corpus of millions of words and be short on evidence for the phenomenon studied to be able to validate a hypothesis, or to give an explanation or insight into the phenomenon. Sparsity of data is, in that sense, a major problem in probabilistic approaches to linguistics. Corpus Linguistics and its probabilistic approach still is, however, more accurate than relying on intuition and making up primary data, according to the author. At this point he proposes a way of dealing with these problems in linguistic analysis, by saying that the solution lies in the combination of probabilistic models with sophisticated linguistic theories. Corpus Linguistics can and must, on that account, integrate with other methods gainfully.

CL has been playing an important role in Semantics/ Pragmatics studies recently, especially when it comes to politeness/impoliteness (Culpeper 1996; Bousfield 2008; Leech 2005; amongst others) and modality (Wierzbicka 1994, Aijmer 2007; etc). Our attempt in the study carried in part II of this paper is to analyse the linguistic devices (narrowed down to adverbs of certainty) that mark the speaker's evaluation towards the locutive material as well as their coextensive function of mitigating what is being said, through an empiricist approach comparing Brazilian Portuguese and American English. In other words, in this paper we will present a study based on corpora, in which the main objective is to map, describe and compare the semantic and pragmatic uses of adverbs and adverbial expressions of certainty in both languages studied. We hope the contrastive analysis of the data will help establish a network of relations and meanings between the semantic and pragmatic use of certainty adverbs, which will hopefully contribute to the investigation of qualitative differences in the settings of sociolinguistic parameters of both modality and politeness in such languages. CL is, therefore, extremely relevant for this study since it establishes a certain criteria to be observed, or, put differently, CL provides the course in which the study will be conducted. Part II of this paper is characterized by an empirical and descriptive approach, and the methods here adopted will reflect the need to verify the occurrences, their distribution and characteristics within a balanced and representative sample of the languages analysed. We

will look for general patterns to try and understand the context that can possibly trigger the use of a certain structure in language, keeping in mind the methods suggested by CL.

## 8.2. Methods

The corpora analysed in this dissertation, as stated previously, were the Santa Barbara Corpus of Spoken American English (SBC) and the C-ORAL-BRASIL, in their informal parts. A minicorpus of the SBC was created by adopting the methodological criteria of the C-ORAL-BRASIL, but for this study the corpus was analysed in its entirety of approximately 200,000 words. The corpora are comparable in word count and present great diaphasic variation, but the distribution of the texts in each corpus do not follow the same logic. In contrast to the C-ORAL-BRASIL, the SBC is not sorted into monologues, dialogues and conversation. While most recordings of the C-ORAL-BRASIL last between 5 and 8 minutes, the SBC recordings have an average duration of 20-25 minutes. As a result, the SBC as a whole has fewer transcriptions than the C-ORAL-BRASIL. One single recording can, however, include two or more different communicative situations, on some occasions with different speakers. A single transcription of the SBC can start as a conversation and become monologic towards the end. Similarly, a recording can begin as a dialogue and become a conversation towards the end. As an example, transcription number 56 in the SBC begins as a dialogue between a horse buyer and the seller in an office and ends in a conversation between employees of the barn where the office is. The inclusion of multiple communicative situations and speakers within a recording allows for good diaphasic and diastratic variation, but is not ideal for corpus compilation. This variation meant comparability between the corpora is still possible, despite the SBC having fewer transcriptions than the C-ORAL-BRASIL. The latter, however, was built to have a great variety of communicative contexts, and is therefore more diverse than the SBC when it comes to that.

For this study, 42 texts of the informal portion of the SBC were analysed, each of which contains approximately 4,500 words, totalling approximately 200,000 words. The informal part of the C-ORAL-BRASIL also with approximately 200,000 words, is divided into 139 transcriptions of approximately 1,500 words each. As already stated, only a minicorpus of 20 texts of the SBC was adapted to the C-ORAL-BRASIL methodological criteria. For the purposes of this dissertation, I have analysed 42 texts of the original SBC and the excerpts

that contained samples of the adverbs analysed were prosodically segmented upon demand, according to the criteria established by LAcT (Cresti 2000).

Each text file was manually checked with the help of the software WinPitch (Martin 2004) and adverbs of certainty were coded. The annotation of the adverbs studied includes their examination at different levels, semantic and pragmatic, within and beyond the utterance. Because of this, the text to sound alignment offered by the software WinPitch was essential in determining the function of the adverb of certainty. The small size of the corpora analysed was advantageous as it made it easier to manually search for and annotate the data, but at the same time makes the study's picture of the phenomena studied less reliable. Despite these limitations, my aim with this study is to provide a description of the tendencies in the use of adverbs of certainty in BP and AE, from a semantic-pragmatic perspective, focusing on the functions of modality and politeness markers.

The methodological approach followed in this study was proposed by Simon-Vandenberg & Aijmer (2007), in the book *The Semantic Field of Modal Certainty*. The authors advocate that adverbs of certainty should be studied from a broader perspective, with attention not only to the meaning (Semantics) but also to the use (Pragmatics). This calls for a semantic-pragmatic analysis of these adverbs, which are more commonly studied in the lights of Modality – Semantics – only.

The methods suggested by Simon-Vandenberg & Aijmer (2007) were adapted to the research carried out in this dissertation and ended up reflecting the following steps:

- I. *Search for samples in the corpora of study*

The list of English adverbs of certainty used as basis for this study can be seen in Appendix 2. As stated previously in the section presenting adverbs of certainty, the list of adverbs studied here is roughly similar to the one suggested by Simon-Vandenberg & Aijmer (2007), with the addition of adverbs in the position of response. The corpora were manually searched for every adverb of certainty and the following numbers were found:

Table 8.1: Total Frequency of the adverbs

<b>C-ORAL-BRASIL</b>		<b>SBC</b>	
Type	Freq.	Type	Freq.
mesmo	341	sure(ly)	39
claro	58	of course	45
exatamente	49	exactly	37
com certeza	28	definitely	19
justamente	17	certainly	10
de verdade	6	obviously	10
sem dúvida	5	for sure	8
necessariamente	4	necessarily	6
evidentemente	4	evidently	4
obviamente	4	no doubt	4
logicamente	2	indeed	1
deveras	1		
certamente	1		
claramente	1		
<b>TOTAL</b>	<b>521</b>	<b>TOTAL</b>	<b>183</b>

Note that the table above shows the total frequency of the types in their adverbial function. Not all of them, however, figured within the category of modal or politeness marker, which leads us to the next step.

## II. *Feature analysis of the adverbs and classification*

In order to achieve a more accurate account of the use of adverbs of certainty, it is crucial to consider their usage in a social context as well as within the utterance. At this stage, I focused on who was using which adverb while speaking to whom, when and for what purpose, based on the following parameters:

### a) Position

The position of the adverb within the utterance was considered in this study in order to investigate if different positions correlate with different functions, as suggested by Simon-Vandenberg & Aijmer (2007:82). The authors distributed the adverbs of certainty found in their analysis between initial, medial, final and ‘other’ position. The same categories were followed for this study. Initial and final positions are very straight forward, and refer to the positions respectively introducing and closing the utterance. Medial position indicates that the adverb either immediately precedes or succeeds the verb. The ‘other’ position refers to the adverbs that could not be classified within one of the previous positions. For this study, I have

added the position ‘response’ to the group, since the authors do not deal with adverbs in this position in their study. In summary, the adverbs of certainty found in this study were classified in initial, medial, final, other and response.

#### b) Situation

The situation refers the context in which the adverb of certainty was found, with a description of the communicative situation, what was happening prior to the occurrence and the topic of conversation.

#### c) Text Genre

The texts of both the SBC and the C-ORAL-BRASIL were classified in argumentative, expository and narrative. In the argumentative genre, the speaker is trying to convince the hearer of something, by presenting his/her arguments. The expository genre aims at the explanation of a certain matter. In the narrative genre, the speaker tells a story, normally in a chronological order. It should be noted that any given transcription could have more than one genre attached to it depending on the part of the text analysed. A recording can have, for instance, a narrative and an argumentative portion.

#### d) Type of the communicative context

The transcriptions in both corpora were divided into public and private contexts. They are public when the speakers assume their public role with a social function (a meeting of teachers in a school, for instance). The texts were private when the speakers act as an individual in a more informal context – i.e. a conversation among family members.

#### e) Speaker’s metadata

The gender, age and schooling level of the speaker who used adverbs of certainty were also specified during the process of annotation of the data, for potential sociolinguistic analysis. The age and schooling classification follow the design proposed in the C-ORAL-BRASIL and already outlined in part I of this dissertation.

#### f) Relationship between interactants

The association of power and social distance in interaction is an important factor to be observed when defining whether the adverb was being employed in the expression of politeness<sup>36</sup>. For each occurrence of the adverbs, the relationship between speaker and hearer was stated. They could be friends, husband and wife, near strangers, work colleagues, etc.

#### g) Type of illocution

Illocution is a very important notion to be considered both in the study of politeness and modality, as previously discussed. An accurate account of the type of illocutions in which an adverb of certainty was found can only be achieved with a detailed prosodic analysis, which was not done for this particular study. For the analysis proposed here I have done a very superficial classification of illocutions based on the list provided by Cresti and Firenzuoli (2001) and presented on pp. 63 of this dissertation.

#### h) Function

As already discussed, adverbs of certainty have developed many other functions within the utterance besides marking epistemic modality. The most salient functions that adverbs of certainty performed in the corpora analysed were modalizer, politeness marker, emphazier and focalizer. Deciding the function of a specific occurrence was not an easy task, and to rule out misclassification due to the multi-functionality of adverbs of certainty, some tests and procedures had to be performed:

- i. *Substitution*: this method was suggested by Simon-Vandenberg & Aijmer (2007) and consists of replacing an adverb of certainty by another adverb of certainty of the same value and verifying if the utterance retained its meaning.
- ii. *Translation*: this test was also suggested in the book *The Semantic Field of Modal Certainty* (Simon-Vandenberg & Aijmer 2007). The authors suggest that, similarly to the method of substitution, if an adverb can be replaced by its equivalent in a different language and retain the original meaning of utterance, this would be evidence of a specific function of that adverb.

---

<sup>36</sup> A detailed account of the relationship between politeness and social values was given in the section dedicated to the discussion of linguistic politeness.

- iii. *“It is true that...”*: this is a test commonly used to verify epistemic modality in a given utterance. If the locutive material of the utterance can be put after “it is true that” and keep the meaning unaffected, it suggests that the adverb is an epistemic marker.
- iv. *“It was ... where/who/that...”*: it-cleft sentences were used as tests to check if an adverb of certainty was a focalizer. If the utterance with the adverb of certainty could be cleft without losing its meaning, that adverb could potentially be considered a focalizer instead of an epistemic marker, for instance.
- v. *Test of gradability*: to test if a given adverb was to be considered an intensifier, the element it refers to was submitted to a test of gradability. Many adjectives, for instance, are inherently gradable, and can take many degree modifiers. An adverb of certainty whose scope was a gradable adjective or verb, for example, was considered an intensifier that enhances the meaning of the word it modifies.

Examples of different adverbs in their various functions will be presented in the data analysis and results section of this dissertation.

### III. *Adaptation of the excerpt to the prosodic segmentation criteria established by LAcT*

Once classified as modal or politeness marker, the excerpt in which the adverb of certainty occurred was prosodically segmented following the methodology proposed by LAcT (Cresti 2000), which was explained in detail in part I of this dissertation. This process had only to be done for the SBC texts which were not part of the adapted minicorpus of English (whose process of adaptation was also explained in part I of this paper), to ensure comparability with the transcriptions of the C-ORAL-BRASIL corpus.

Based on these parameters, each adverb of the first list (shown above) was examined in context to see if they functioned as modalizers or politeness markers, in alignment with the focus of this study, which helped me to arrive at the list below:

Table 8.2: Frequency of the adverbs as modalizers/politeness markers

<b>C-ORAL-BRASIL</b>		<b>SBC</b>	
Type	Freq.	Type	Freq.
mesmo	114	of course	41
claro	58	sure(ly)	39
com certeza	27	exactly	19
exatamente	24	definitely	18
justamente	7	certainly	10
sem dúvida	5	obviously	10
necessariamente	4	necessarily	6
obviamente	3	evidently	4
logicamente	2	no doubt	4
evidentemente	2	for sure	3
deveras	1	indeed	1
certamente	1		
claramente	1		
de verdade	1		
<b>TOTAL</b>	<b>250</b>	<b>TOTAL</b>	<b>155</b>

The differences between the uses of these adverbs, as modalizers or politeness markers will become clearer in the next section of this dissertation, when the data will be further analysed, exemplified and detailed.



## 9. DATA ANALYSIS AND RESULTS

As shown on table 5.1, there are some adverbs that appear more frequently in the data analysed, and will therefore be given more attention in this section. Here I will discuss the uses of *mesmo*, *claro*, *com certeza* and *exatamente* for BP and *of course*, *sure*, *exactly* and *definitely* for AE. The adverbs with 10 or less occurrences will be treated in less detail or just mentioned. In the end of this section I will present a table with the summarised feature analysis of the most frequent adverbs for better visualisation. It is important to highlight that the views presented in this section reflect what was observed in small corpora, of approximately 200,000 words each. In order to confirm the tendencies discussed here, further analysis should be carried out in larger set of corpora.

### 9.1. *Mesmo*

#### 9.1.1. General remarks

*Mesmo* is by far the most frequent adverb in BP. As shown on table 5.1, it has an average frequency of 57 occurrences per 100,000 words. *Mesmo* also figures between the top 100 more frequent words in the corpus C-ORAL-BRASIL, as seen in Appendix 3. Even when compared to the most frequent adverb in AE (*of course* – 20 per 100.000 words), the difference is remarkable. In the Houaiss Dictionary of Portuguese (2001), *mesmo* is defined as “1. de modo exato; exatamente, justamente. 2. de fato, de verdade; realmente”. This definition suggests a feature shared by many other adverbs of certainty: that it originated as a manner adverb that developed itself into a sentence adverb (Simon-Vandenberghe; Aijmer 2007:85). Its high occurrence suggests that this adverb has developed various functions in discourse. In the following subsection, I will describe some of those functions based on the analysis of data, with special attention to the semantic and pragmatic functions of modal and politeness markers.

#### 9.1.2. Syntagmatic and paradigmatic description

It appears that *mesmo* is not very flexible when it comes to its position within the utterance. Table 9.1 shows that *mesmo* more frequently occurs in the final position of the utterance:

Table 9.1: Position of *mesmo* within the utterance

		<i>Positions</i>				
		<b>Initial</b>	<b>Medial</b>	<b>Final</b>	<b>Other</b>	<b>Response</b>
<i>Mesmo</i>		-	10 (8,77%)	101 (88,59%)	3 (2,64%)	-

*Mesmo* also occurs in medial and other positions, but when it occurs in the initial position *mesmo* does not classify as a modal or politeness marker within the utterance and therefore was not considered for this study. Consider the examples:

(9.1) \*LIA: *é // pa iluminar esses pontos // mesmo que você não tenha olheira //* (bfamd127)

(9.2) \*DIA: *<ah ô fessora> // que dia mesmo que cê / marcou a prova na nossa sala //* (bpubmn10)

(9.3) \*MUR: *vai // vai / celera // po' acelerar / senão ele nã sobe não // po' acelerar mesmo // fundar o pé // vai //* (bpubdl04)

The examples above show the types of *mesmo* that were not considered for this analysis, as they do not function as adverbs of certainty. In example (1), the adverb is in initial position followed by *que*. *Mesmo que* is an expression that connects the text with the wider context. It denotes the idea of accepting a situation that sits alongside a primary situation. The expression does not presuppose any commitment to the truth of the locutive material nor classifies a politeness marker. Example (9.2) shows *mesmo* followed by *que* again, but this time not as an expression in initial position. The adverb in this instance refers to the noun it precedes, bringing it to a focus, functioning as a focalizer. The last example shows *mesmo* in another of its very frequent functions, assuming the form of an intensifier. This non-modal meaning can sometimes be misclassified as a modalizer and therefore needs to be closely analysed. In example (9.3) we have *mesmo* following the verb *acelerar*, in final position. Now it is important to analyse the scope of the adverb in order to determine its function, doing this in conjunction with the semantic characteristics of the word it succeeds. The verb *acelerar* has the notion of gradability attached to it, especially in the context in which a driving instructor is giving commands to his student. In this particular case, *mesmo* comes to emphasize the verb *acelerar*, acquiring the meaning of 'a lot'. When trying to understand the meaning of *mesmo* here, instead of thinking of it in a scale of probability, we have to think of the adverb

in a scale of gradability, which classifies it as an intensifier instead of an adverb of certainty. Now let us analyse examples in which *mesmo* acts as a modalizer:

(9.4) \*LAU: *é verdade / Luzia // a outra vez tava mais tranqüila / <né> //*  
\*LUZ: *<tava> // tava úmida / &he / a coisa // agora não //*  
\*LAU: *<é> //*  
\*LUZ: *<ressecou> mesmo // ai / lá vem outro caminhão // espera aí / cara //*  
*segura a onda // (bfamdl03)*

(9.5) \*ALV: *eu sou // sou bicho do asfalto com força //*  
\*REN: *é //*  
\*ALV: *<totalmente do asfalto> //*  
\*REN: *<urbanóide mesmo> // (bfamdl28)*

The examples above show *mesmo* in final and post-verbal medial positions, where they are more likely to function as modal or politeness markers. Example (9.4) is a good illustration of how *mesmo* can carry an evidential modal status, which appeared in the data as a frequent characteristic of this adverb in its modal function. In (9.4), the speaker is concluding her train of thought by using *mesmo* to certify her assumption that the landscape around them had dried up since they last saw it. Example (9.5) shows an analogous situation, but this time *mesmo* is used by the hearer, to conclude and highlight the speaker's thoughts based on the evidence given.

As already discussed in the previous sections of this dissertation, modality is sometimes used in correspondence to the notion of attitude even though they represent different sets of phenomena. Example (9.6) below can illustrate that:

(9.6) \*MUR: *mas vai olhando muito seu lado e vai / deixando ir // porque o*  
*[/I] aí é a largura exata de passar // vai / desvira pra cá // pra cá // não /*  
*nem tanto / senão cê vai bater // do jeito que cê tá aí cê já vai bater já //*  
*mesmo // (bpubdl04)*

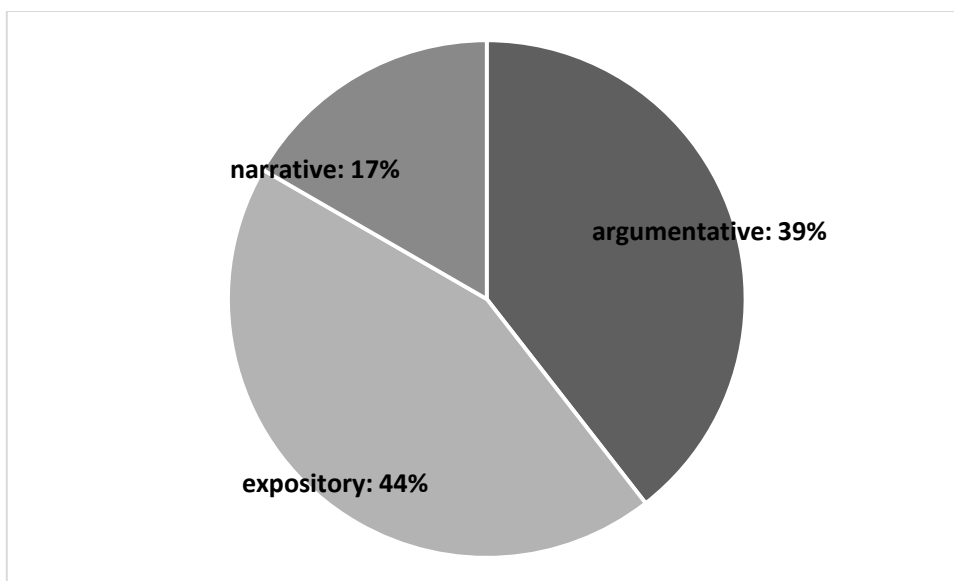
The example above shows how we can have both modality and stronger attitudinal effects occurring at the same time, and this is only possible because they occur within different levels. The speaker emphasizes his commitment to the factuality of his assertion with prosodic cues. We therefore have an epistemic *mesmo* with a pronounced attitude, altogether.

The same can be observed in the example below, in which the speaker uses intonation effects to highlight her attitude towards the game characters she values the most.

(9.7) \*BAR: *Nossa / tem muito tempo // porque agora que ele &e + nũ tem nem um ano que eles abriram / a conta [/2] o / servidor grátis / sabe // tão / tipo / eu jogo no servidor grátis há muito [/2] &he / tipo / há pouco tempo // mas / com personagens que eu nũ vou / ter futuro // no servidor pago / que tão meus personagens que eu gosto mesmo / sabe // tanto que eu &cli [/3] que eu / criei / classes diferentes // (bfamd123)*

The text genres in which *mesmo* occurs are balanced between argumentative and expository types, as are demonstrated in the graphic below:

Image 9.1: Frequency of *mesmo* per text genre



When it comes to illocutions, a more detailed study needs to take into consideration the detailed analysis of prosodic cues and information units. Without such analysis it is not possible to achieve a complete account of the behaviour of the epistemic modals, as pointed out by Tucci (2009). To the author, the domain of modality is the information unit. I can, however, point out some tendencies found with a less detailed analysis of the illocution types, even though only a complete prosodic analysis can confirm these tendencies. The vast majority of the occurrences of *mesmo* were found within the representative category of illocutions. There were no occurrences of *mesmo* in directives such as total questions or

requests of confirmation, as one would expect so since speakers would not express their commitment to the truth and question it at the same time.

It appears that *mesmo* can also be employed in the expression of politeness, even though it frequently does not. Only 10 occurrences in which the phenomena of modality and politeness occurred at the same time were found in the corpora analysed. Example (9.8) above illustrates that:

- (9.8) \*REN: *ô // xá eu te falar / sô // quando eu ia lá em Perdões / quando eu era pequeno / &e [/1] cada vez que eu chegava lá / eu ficava conhecendo mais parente / meu / que eu não conhecia // e até &ho + é / uai // e até hoje deve ter // gente / lá da minha família / que eu nũ conheço até hoje //*  
\*ALV: *é / esses / negócio de parente é engraçado mesmo //* (bfamdl28)

Here ALV is at the same time expressing his commitment to the truth of what was just said by speaker REN and maximizing sympathy between himself and REN. That instance of *mesmo* seems to have a stronger pragmatic function of politeness over a weakened modal force. Similarly, example (9.9) below shows the use of *mesmo* to mitigate an inherently impolite illocution, as the speakers are disagreeing during a football match. Speaker CAR comes to act as a mediator between the two parties, identifying the source of the problem and expressing certainty in a diplomatic manner, in conjunction with other politeness strategies:

- (9.9) JOS: *pera aí / varão // <foi na mão dele> //*  
\*MAR: *<então vai / vai> //*  
\*CAR: *foi mão mesmo / varão //*  
\*JOS: *foi mão / ué // tem que parar tem que bater / <uai> //* (bfamcv05)

Differently from example (9.8), however, it appears that the pragmatic function in (9.9) does not weaken the semantic function, or vice-versa.

We had an occurrence on the corpus that interestingly sums up what has been said up to this point regarding the reason why speakers use *mesmo* in the majority of the occurrences - as a modal marker. The speaker herself uses the adverb of certainty *mesmo* and immediately explains the use:

(9.10) \*ANG: *onde / a fé dela / &he / &he / deu exemplo pra nós o tempo todo / e as coisas aconteciam / simplesmente aconteciam // e [/1] e [/1] e antes de acontecer / na hora das dificuldade / ela falava com a gente / calma // calma que quem sabe é Deus // a solução vem // cês nũ [/1] nũ esquentam a cabeça não // a solução vem // e vinha **mesmo** // era impressionante // vinha / **mesmo** // nũ tinha [/2] nũ tinha dúvida nenhuma // &e [/1] ela tinha essa certeza dentro dela / e to / da / vez / ao iniciar qualquer tarefa / e ao finalizar / ela / pedia a benção pra tarefa que ela ia fazer // (bpubmn03)*

The first *mesmo* was used to express the speaker's commitment to the truth of what she had just described. The second occurrence of *mesmo* is a slightly more nuanced, as it is not only a modal marker but also possesses a stronger attitudinal intonation effect as highlighting the certitude of the expression. The speaker then decided to highlight her commitment to the truth further more, by explaining that her previous use of *mesmo* was an assurance that there was no doubt in regard to the accuracy of the situation, regardless of the hearer's beliefs. I believe that this example also illustrates how speakers only express modal certainty if there is a good reason for doing so (Halliday 2004). As previously explained, there is a paradox in modal certainty which says that speakers only express certainty if it cannot be taken for granted. In the example shown above, the speaker is describing what her mom used to tell her as a child, and giving 'evidence' why she took that as lesson based on her personal religious beliefs. Once the speaker knows that the hearer cannot judge what she is saying based on evidence given, and that only her word for it will have to be sufficient, she makes use of both an epistemic marker and a pronounced attitude - marked by intonation - to convey a higher degree of certainty.

## 9.2. *Claro*

### 9.2.1. General remarks

*Claro* is the second most frequent adverb in the corpus, with 29 occurrences per 100.000 words. It seems to be close to *com certeza* in meaning. Their definitions in dictionaries often overlap. The Houaiss dictionary of Portuguese (2001) uses other adverbs studied in this dissertation in the description of *claro* and defines it as "1. bem visível, distinto ou

discriminável à vista; nítido; 2.com clareza, de maneira clara” or “3. sem dúvida”. In the following sections we will analyse *claro* and its features.

### 9.2.2. Syntagmatic and paradigmatic description

With regard to position in the clause, *claro* has a strong preference for occurring alone in responses. It also frequently happens in initial position, but the medial position is not an option for the occurrence of *claro* as an adverb of certainty:

Table 9.2: Position of *claro* within the utterance

		<i>Positions</i>				
		<b>Initial</b>	<b>Medial</b>	<b>Final</b>	<b>Other</b>	<b>Response</b>
<i>claro</i>		15 (25,86%)	-	2(3,44%)	1 (1,72%)	40 (68,96%)

The following examples illustrate the positions in which the adverb was found:

- (9.11) \*JOS: "tu só fala em" + *que mineiro que cê conhece que conversa assim* / "tu só fala em ir pra roça" //  
 \*SON: não // **claro** que não //  
 \*JOS: hhh porque roça tem em tudo quanto é lugar / uai // (bpubcv07)

- (9.12) \*DAN: ah / mas dá uma molhadinha // é / cê tem <que> +  
 \*MAR: <tá> rolando / <Plauto> //  
 \*BAL: <não> / tem que molhar / **claro** // (bfamcv24)

- (9.13) \*TER: ir chique mesmo // alugar uma roupa / <adequada> / que é / de / dia //  
 \*RUT: <**claro**> // adequada // (bfamcv02)

The first example demonstrates how *claro* most frequently occurs in initial position: followed by *que*. This adverbial expression is largely employed by speakers to express that something is to be taken as evident, excluding any other options as true. The second example portrays the adverb in final position, with the whole utterance as its scope. In this position, *claro* seems to come as an afterthought, and appears to have the function of strengthening what was said

by presenting the information given as evident. The third example shows *claro* in its most common position, alone in responses. In that position *claro* sometimes not only expresses the speaker's commitment to the truth of what was said but can also convey the pragmatic function of maximizing sympathy between speakers, as a signal of emphatic agreement, reassuring the previous speaker. Let us examine the following examples for clarification on the matter:

- (9.14) \*DFL: *e eu ficava até com uma certa inveja / porque papai era muito sisudo //*  
 \*LUC: *<ham ham> //*  
 \*DFL: *<muito caladão> // e ele / brincalhão / porque era a única <filha> / né //*  
 \*LUC: *<ham ham> // <claro> // (bfammn02)*

- (9.15) \*LUC: *não não // é / não / mas aqui tá tranquilo // o ruim / é / tipo / céu / acertar direitinho / porque / a / <precisão dele é> muito ruim / e mexer só um sozinho separado do outro é um saco também //*  
 \*CEL: *<é> //*  
 \*LUC: *né //*  
 \*BER: *claro // nada como um / equipamento bem pensado // (bfamcv31)*

In example 9.16 below, for instance, *claro* appears to have a slightly different function. Speaker NAN seems to use the adverb not only to reassure the previous speaker but also to avoid disagreement as a strategy to focus on his real opinion, which differs from the opinion of the former:

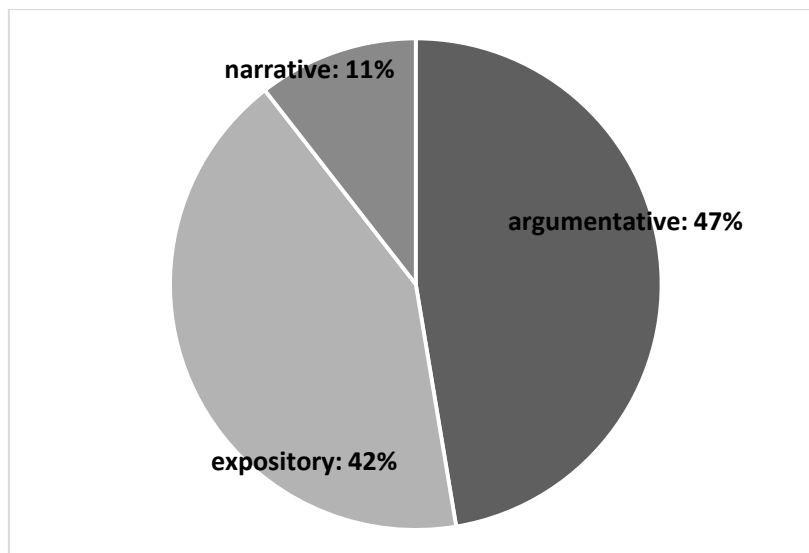
- (9.16) \*TOM: *olha / o [1] o [1] os anos oitenta foram os melhores anos / do Caetano / do Gil / do Chico <Buarque / muito &me> +*  
 \*NAN: *<não / os anos setenta> //*  
 \*TOM: *não / os &an [2] final de setenta / &he / primeira parte de oitenta // é / e setenta <também tinha ditadura> / né //*  
 \*NAN: *<é> // <claro> // mas acontece o seguinte / Tommaso / que eu acho o seguinte // que a grande coisa mesmo / o [1] a semente de*



*tudo / veio na melhor época do Brasil / que foi essa coisa dos anos sessenta / final dos anos cinquenta // que ali / que tinha [/2] que tinha / Glauber Rocha / que tinha / sabe // (bfamcv34)*

Regarding the relationship between the use of *claro* and specific text genres, it seems that its behaviour is identical to the one observed for *mesmo*. *Claro* is not frequent in narrative contexts, but has its use balanced when it comes to the argumentative and expository text genres.

Image 9.2: Frequency of claro per text genre



*Claro* frequently appeared in the data with a strong evidential connotation, much like *mesmo*, which is typical of its semantic meaning. Examples 9.17 and 9.18 below illustrate this point:

(9.17) \*TOM: <e> aliás [/1] aliás / foi exatamente / na &mi [/2] na minha opinião / foi exatamente / depois / que passou a fase mais criativa / que / &he / o resto do mundo percebeu / a importância do Brasil //  
\*NAN: **claro** // (bfamcv34)

(9.18) \*BAO: <aquilo é> possível //  
\*LIQ: é / uai // a gente só ã vê / porque a gente só sente // mas / nós ã conseguimos ver // mas / **claro** que isso é possível // (bfamd131)

These examples show how *claro* is often deployed in the same way as *evidently* is. This does not come as a surprise as *claro*, by definition, has the notion of “being sharp to the eyes” attached to it, as pointed out in the initial paragraph of this section with the Houaiss definition of the adverb. It could be argued that the evidential use of *exactly* is not modal and should be detached from its modal meaning. Whether or not that evidential feature has to be detached from modality is not the main concern of this dissertation, mainly because nothing particularly relevant could be said on the topic with only the analysis of adverbs of certainty and, moreover, in such limited data.

### 9.3. *Com certeza*

#### 9.3.1. General remarks

*Com certeza* is the third most frequent adverb with 27 occurrences in 200.000 words. It is defined by the Houaiss Dictionary of Portuguese (2001) as “1. conhecimento íntimo ou expresso; convicção; 2. o que não oferece dúvida”. This definition suggests that *com certeza*, in a probability scale, has a high value of certitude. This adverb has, however, developed other functions that defy its dictionary definition, as explained in the following subsection.

#### 9.3.2. Syntagmatic and paradigmatic description

As already highlighted, *com certeza* has an inherent characteristic of offering a level of certainty that sometimes borders on total conviction. Example 9.19 below illustrates this notion, in *com certeza*'s most common position:

- (9.19) \*FAB: *né / nós nũ &s [/1] <temos / &he> / como se diz / são [/1] são  
problemas que [/1] de outra alçada //*  
\*ADA: *<com certeza> //*  
\*FAB: *<nũ é do [/1] de> professor //*

*Com certeza* was more frequently found alone in responses, but was also occurred in initial, medial and final positions, although much less frequently:

Table 9.3: Position of *com certeza* within the utterance

	Positions				Response
	Initial	Medial	Final	Other	
<i>Com certeza</i>	4 (14,8%)	2 (7,40%)	2 (7,40%)	-	19 (70,37%)

Examples 9.20 to 9.22 below show *com certeza* in initial, medial and final positions, respectively:

(9.20) \*GIL: <ô / mas> / voltando à questão / falando em e também falando em povo mascarado / esse povo do Galáticos é muito palha / eu acho que es nũ deviam mais participar / e <tal> //

\*LUI: <não> //

\*LEO: <não> //

\*LUI: <eu acho não> //

\*LEO: <com certeza> //

\*LUI: <com certeza es nũ vão participar / uai> //

\*LEO: <eles são piores do que o> Durepox // (bfamcv01)

(9.21) \*MAR: quebrou dois dedos // <ficaram> roxinhos e &v [/1] &f [/1] aquela / bolinha roxa / e eu desesperada / nũ sabia se levava no médico / que que fazia / a mãe dela / **com certeza** devia tar viajando / voltou correndo da viagem / pra poder / resolver o problema / que eu nũ sabia o que que eu fazia / se levava pro médico / se deixava do jeito que tava / e ela chorava tanto / chorava tanto // (bfammn22)

(9.22) \*DML: tô falando com <a Sandra> aí / que ela falou que e' tá com um / carrão aí / eu falei / ah / minha filha //

\*ECR: <já tá trocando> //

\*DML: <já tá> tratado pra trocar numa casa //

\*ECR: é / na casa // ele trocou o outro / que ele falou / agora / o' // vai trocar em outra casa // eu acho que e' deve trocar uma po pai dele / **com certeza** // (bfamd115)

As already discussed, *com certeza* has an inherent semantic function of expressing almost total conviction towards the locutive material. Sometimes, however, the adverb behaves in a different manner, confronting its natural semantic characteristic of portraying a higher level of certitude. Let us analyse the example below, in which it is clear to see that *com certeza* appears to have suffered semantic bleaching, losing its semantic content:

(9.23) \*CAR: *futuramente pretendo fazer química // eu sou [2] tenho afinidade com exatas // futuramente / com certeza / eu devo fazer química / física // (bfammn15)*

This characteristic of losing part of its original connotation is very common among the adverbs studied. The example above, for instance, shows how the adverb has lost its inherent semantic meaning of expressing truth-values once the context in which it is inserted arouses less epistemic certainty with the use of the verb *devo*. The same phenomenon happens in examples 9.21 and 9.22 above. Something similar happens in example 9.24 below:

(9.24) \*ELI: *ela me xingou também numa hora lá que / eu tava fazendo a curva // mas eu achei que ela podia ter me passado na boa //*  
\*MUR: *entra direita hhh // com certeza //*  
\*ELI: *eu vou jogar mais pra direita aqui // (bpubdl04)*

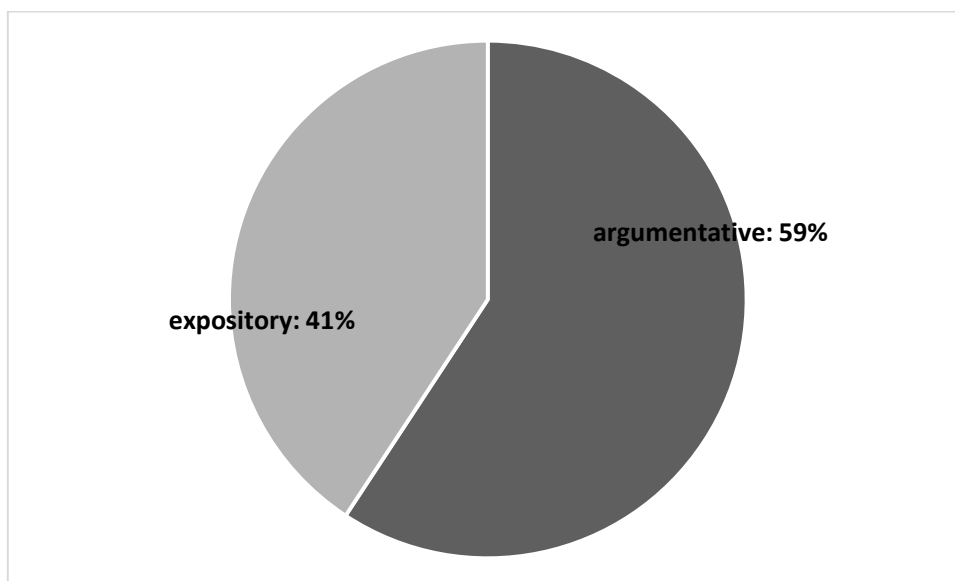
ELI is taking driving lessons and telling her instructor what happened during her driving test, while driving. She describes how she thinks that the examiner was too strict during her test and how she could have given her a pass. MUR, the instructor, seems to be absorbed in his thoughts, but he uses *com certeza* to agree with his student anyways. Given the situation, chances are that MUR has listened to the same complaint made by different students countless times, and, given how delicate the position of failing in a test might be to a student, he chooses to mark his commitment to the truth of what was said even though he seems to have no conviction of it. Speaker ELN in example 9.25 below also seems to have less certainty than expected by the use of *com certeza*:

(9.25) \*BAL: mas o grande lance é esse // é se jogar // foda-se o resto //  
\*ELN: é // **com certeza** // embora eu não goste de me jogar / né //  
(bfamdl18)

Similarly to the previous example, it appears that *com certeza* suffers semantic bleaching and weakens its inherent semantic function of a modalizer, but in this case, to become a stronger politeness marker, especially because the next utterance conveys a completely different point of view from the previous speaker. This is especially noticeable because *com certeza* is being used as a negative politeness strategy, as to maintain the harmony in a naturally impolite context (in this case, disagreeing with the previous speaker).

The graphic below shows that a balanced distribution between expository and argumentative text genres seem to be the general behaviour among adverbs of certainty, as it was also observed for *com certeza*. This particular adverb, however, did not occur in narrative contexts, not even once.

Image 9.3: Frequency of *com certeza* per text genre



#### 9.4. *Exatamente*

##### 9.4.1. General remarks

According to Houaiss Dictionary of Portuguese, *exatamente* means “1. com rigor, com exatidão; nem mais, nem menos; precisamente; 2. usado para confirmar algo antes dito”. *Exatamente* is almost as frequent as *com certeza*, with 24 occurrences in 200,000 words. Like

*mesmo*, the definition of *exatamente* in the dictionary suggests that this adverb began as a manner adverb that has developed itself into a sentence adverb. Its status of adverb of certainty is argued among linguists. Simon- Vandenberg & Aijmer did not include the direct translation of *exatamente* (*exactly*) in their list, arguing it was not frequently found as a modalizer. The authors, however, attest the occurrence of *exatamente* as a translation equivalent for *certainly* and *clearly*, which indicates that the adverb can function as an epistemic marker in some languages, even though not regularly.

#### 9.4.2. Syntagmatic and paradigmatic description

*Exatamente*, along with its direct translation in AE – *exactly* – was the only adverb among the most frequent ones to have been exclusively found alone in responses in the function of modalizer or politeness marker. This means that this adverb is not flexible when it comes to position within the utterance, and any other position except for the response one would result in *exatamente* assuming different functions. Observe the following examples:

- (9.26) \*SUE: *a gente sabe que os mais velhos eles têm uma disfunção / por causa da idade / por causa das doenças // e / aqui / por + os jovens // aí vem aque' negócio / ansiedade / depressão / tabagismo / alcoolismo //*  
\*BRU: *drogas // <etcetera> //*  
\*SUE: *<drogas> // exatamente // (bfammn16)*

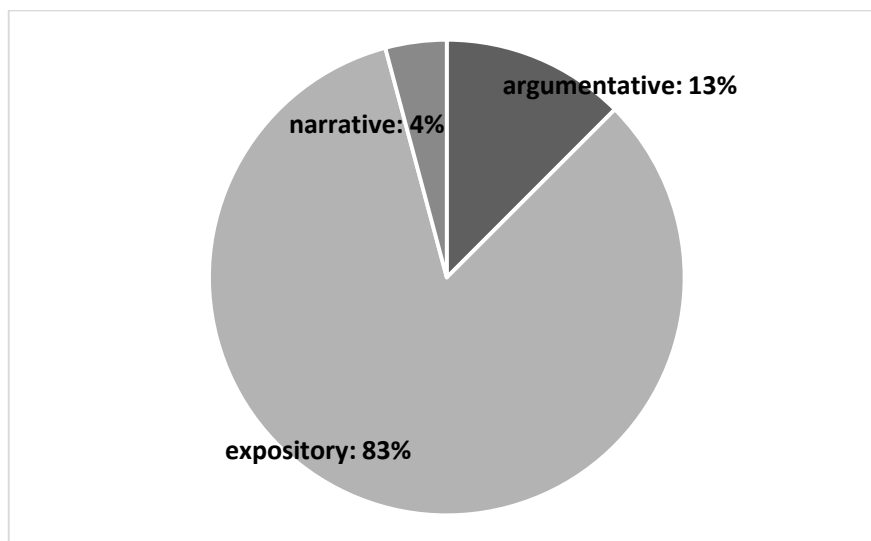
- (9.27) LEO: *e / aí depois / quando voltou a funcionar direito / voltou exatamente quando a Maria do Carmo ia começar [2] fazia / o início da fala dela // então acho que ficou / tranqüilo // (bpubcv08)*

In 9.26, *exatamente* is in the position of response, operating as an epistemic marker. SUE is confirming what was said previously by BRU to be true. Example 9.27 shows *exatamente* in medial position, when it operates as a focalizer. When we analyse the total frequency of the adverb in the corpus, it was very frequently found in that position, in which case it did not function as a modalizer/politeness marker and thus was not targeted as object of study in this dissertation. However, the fact that *exatamente* was as frequent in the function of epistemic marker as it was in the function of focalizer (12 occurrences in 100.000 words each) deserves

attention, as it might suggest a semantic development from a manner adverb to an epistemic marker. The uses of *mesmo* suggest the same type of ongoing process.

The distribution of *exatamente* differs from the previous adverbs described so far. It was notably and commonly found in expository contexts, as the graphic below shows:

*Image 9.4: Frequency of exatamente per text genre*



## 9.5. *Of course*

### 9.5.1. General remarks

*Of course* is the most frequent adverb for AE. It appeared in the corpus with an average frequency of 20 occurrences per 100,000 words. The Longman Dictionary of Contemporary English (1995) states that *of course* is “1. used to show that other people probably already know what you are saying is true, or expect to hear it; 2. used to emphasize that what you are saying to someone is true or correct”, or “3. used to show that you accept or agree with what someone has just said”.

### 9.5.2. Syntagmatic and paradigmatic description

With regard to position in the utterance, *of course* appears to have a strong preference for initial position:

Table 9.4: Position of *of course* within the utterance

	<i>Positions</i>				
	<b>Initial</b>	<b>Medial</b>	<b>Final</b>	<b>Other</b>	<b>Response</b>
<i>Of course</i>	23(56%)	5(12,1%)	2(4,8%)	6(14,6%)	5(12,1%)

The adverb was, as shown on table 9.4, also found in every other position studied, fact that demonstrates its versatility within the utterance. The examples below show the adverb in different positions:

(9.28) \*TOB: *you remember [2] you know yyy of course //*  
 \*TOA: *yeah //* \*TOC: *ah //* (sbc032)

(9.29) \*TOA: *<yeah> / I'll have a little <more / yyy> //*  
 \*TOC: *<sure> //*  
 \*TOB: *<sure> //*  
 \*TOC: *<course he wants> red wine //*  
 \*TUC: *red wine //*  
 \*TOA: *hum hum //* (sbc032)

(9.30) \*BER: *&he / supposedly it's harder for them / it's [1] it's so like <hard &I> [1] it's / against their machismo <to [2] machismo to be +*  
 \*FRA: *<to> + <oh / of course> //* (sbc051)

(9.31) \*JON: *poster //*  
 \*ALN: *a benefit // it was a poster // that's just [2] that's not a [3] that's not what they call a seriograph // the Joe &n [3] the Joe uh Namath /<who was of course> an <Alabama> football player //*  
 \*JON: *<yeah> // <yeah> //* (sbc060)

Examples 9.28 to 9.31 show *of course* in final, initial, response and medial positions, respectively. Example (9.31) illustrates the first meaning for *of course* given by the Longman Dictionary of Contemporary English in the beginning of this section, which highlights the adverb's feature of talking about knowledge shared by the speakers. Example (9.31) portrays speaker ALN presenting the information that follows *of course* as known and true, but at the



same time as not the crucial to the point he wants to make. *Of course* in these types of contexts not only functions as an epistemic marker but also as a positive politeness strategy, because its evidential meaning enhances the hearer's face needs to be respected as an intellectual being while avoiding patronising. The same can be observed in examples (9.29) above and (9.32) along with (9.33) below:

(9.32) \**TOB*: <**course** this is> sort of Democratic state //  
\**TOC*: it was the Democratic <State> // (sbc032)

(9.33) \**FRA*: &he / when we / went back to New York / &he / I had / kept a  
studio there // and / **of course** / that was small / and [/I] and then / we  
found another studio / in [/I] &he / in / our neighborhood / that we liked //  
in the Village // (sbc051)

*Of course* is again employed in the examples above as a positive politeness device to enhance solidarity between speakers, presenting the information as common knowledge. It serves the speaker in the solidarity-oriented goal of maintaining a balanced interaction and avoiding patronising. This multifunctionality observed in the use of *of course* makes it a highly frequent adverb, serving not only in the expression of modal certainty but also very commonly employed to enhance politeness in speech.

The adverb does not seem to have a stronger preference for a specific text genre, and did not occur significantly more in any of the genres considered. The occurrences of the adverb is balanced amongst argumentative, expository and narrative texts, with an average of 30% to 35% of instances per each genre. *Of course* often has a strong evidential connotation, not as marked as *mesmo* and *claro*, but still noticeable. The example below illustrates that:

(9.34) \**MAR*: grandma / makes us a quilt / and she makes a double-bed quilt  
for our king-size bed / so **of course** it doesn't fit / so / I hhh + (sbc013)

*MAR* uses *of course* to conclude her thought and to indicate that this conclusion has to be taken as evident given the information provided previously. The adverb *of course* is used above to display the evident consequence of what was previously said, in the same way that

*evidently* would. In this function, *claro* would classify a better translation in BP for *of course* as both adverbs share the same evidential feature.

## 9.6. *Sure(ly)*

### 9.6.1. General remarks

*Sure* is almost as frequent as *of course* in the expression of modality and/or politeness. The SBC has 19 occurrences of *sure* per 100,000 words. The Longman Dictionary of Contemporary English (1995) offers the following definitions for *sure*: “1. used at the beginning of a statement admitting that something is true; 2. used to emphasize a statement”. According to the same dictionary, *surely* is “1. used to show that you think something must be true, especially when people seem to be disagreeing with you”, or has the same meaning of *certainly*.

### 9.6.2. Syntagmatic and paradigmatic description

*Sure* seems to follow the pattern of *mesmo* and *claro* when it comes to preference per text genre – balanced between argumentative and expository contexts. Regarding its position within the utterance, *sure* appears to have its use balanced between medial and response positions, with a minor predilection for the latter. The table below shows the distribution of the adverb within the utterance:

Table 9.5: Position of *sure* within the utterance

	<i>Positions</i>				
	<b>Initial</b>	<b>Medial</b>	<b>Final</b>	<b>Other</b>	<b>Response</b>
<i>Sure(ly)</i>	4(10,2%)	16(41%)	-	-	19(48,7%)

The use of *sure* alone in responses is often connected with the expression of linguistic politeness. In example (9.35) below, for example, TOB uses the adverb of certainty to be sympathetic towards TOA face-needs instead of using it to express his commitment to the truth of what was said. This might indicate that the adverb, similarly to *com certeza*, has suffered semantic bleaching and lost part of its inherent meaning of expressing that something is true to mark positive politeness instead:

(9.35) \*TOA: <that's what I hope> //

\*TOA: <yeah> //  
\*TOB: <sure> //  
\*TOC: <right> //  
\*TOB: <yeah> // (sbc032)

The same type of process can be observed in the sequence of examples below, where BRA is being invited for lunch. The first *sure*, in example (9.36) completely differs from the second *sure*, in example (9.37). In (9.36), BRA uses *sure* only to demonstrate to PHI that he is following what it is being said. As already demonstrated with example (9.35), *sure* seems to have developed this function of serving as a positive politeness device, functioning as an attempt to encourage the flow of conversation. The second *sure*, however, in example (9.37), is used by BRA as an emphatic answer to the invitation, this time expressing that his answer is definite and also marking politeness by emphasizing an already polite illocution.

(9.36) \*PHI: *I don't know if you have plans or not // but we're gonna have  
lunch later // at noon //*  
\*BRA: *sure* // (sbc010)

(9.37) \*PHI: *if you wanna join us / it'd be great //*  
\*BRA: *sure // yeah //* (sbc010)

Some authors consider *sure(ly)* in responses and medial positions to assume the function of *emphasizers* instead of *epistemic markers*. According to Quirk et al (1985:585), the differences between the two functions are not clearly delimited yet, but the authors claim that responses and medial positions seem to favour the *emphasizing* function of the adverb. Since the aim of this study is to describe the use of adverbs and adverbial expressions of certainty, I thought that including such examples in the data would outline an important characteristic of the adverb in question and shed light on the description of its modal meaning. In addition to that, I believe that examples such as (9.36) and (9.37) give a good insight on the potentially emergent modal meaning of *sure*, given the fact that the locutive material is taken by the speakers as something that is decided, permanent, and not to be changed. These meanings are clearly linked to the modal meaning of 'it is certain', and might be an important asset to the

description of this particular adverb. The example below shows *sure* in medial position and adds up to the discussion:

- (9.38) \*ANE: *I'm not hungry / but / gosh that **sure** tastes good hhh //*  
\*ALC: *well / have a little bit more / we just won't tell him hhh //*  
(sbc043)

In (9.38), ANE presents the information as something definite that cannot be changed, especially because it comes from a sensorial evidence. Given the puzzling discussion on whether examples like such should be considered modal or not, for this research I decided to consider *sure* in response and medial positions (such as in 9.36 – 9.38) to assume the meaning of “it is true that...” if that seems to be the case. The non-epistemic use of *sure* described by Quirk et al (1985) should also be highlighted, as decidedly important in the literature.

- (9.39) \*PATT: *she said / I have good news for you // and I thought / yeah / **sure** you do //* (sbc035)

Example (9.39) shows the use of *sure* to convey the same message as shown in the previous examples, but this time for rhetorical purposes. As a native speaker of AE, the speaker MAR knows what *sure* means and why it is used for, and chooses to use it in an utterance that presents an information as certain even though the speaker is clearly not convinced that the information should be taken as genuine. The use of *sure* to cause an impressive effect here happened in the expressive illocution of irony, and other adverbs of certainty, such as *of course*, *certainly* and *obviously* were found to be employed similarly in other expressives such as jokes.

## 9.7. *Exactly*

### 9.7.1. General remarks

*Exactly* is defined by the Longman Dictionary of Contemporary English (1995) as an adverb “used as a reply to show that you think what someone has said is completely correct or true”. Indeed, every occurrence of *exactly* in the function of modalizer found in the corpus seems to echo that definition:

- (9.40) \*RIC: *she thought it might / you know / help our relationship as friends / and / &he +*  
 \*FRE: *she still considers you / man / huh //*  
 \*RIC: ***exactly*** // *she does / because / I mean / we went to church together for the last three Sundays //* (sbc047)

Speaker RIC uses *exactly* to express his commitment to the truth of what FRE said, in a more emphatic manner. The adverb here expresses a high level of certainty, analogously to the meaning of *for sure* or *com certeza* in BP. Similarly to its direct translation in Portuguese, *exactly* is often disregarded as an adverb of certainty in the literature because of another common function it acquires in speech - focalizer. Let us analyse the following example:

- (9.41) \*MEL: *I can't write them **exactly** the way they are / because they stink //* (sbc019)

In the example above *exactly* is a manner adverb that means precisely, and serves to bring the information it precedes to a focus. When *exactly* was not employed as a modal marker as shown in (9.40), it functioned as a focalizer as illustrated in (9.41). The adverb was the third most frequent adverb in the SBC, with an average of 37 occurrences in the whole of the corpus, and 19 instances when functioning as a modalizer and/or politeness marker. This indicates that the *exactly* is as employed as an epistemic marker as it is as a focalizer, and that characteristic deserves to be mentioned.

#### 9.7.2. Syntagmatic and paradigmatic description

*Exactly* only appears as modalizer or politeness markers in the data when in the position of response, except from the following example in which it appears in initial position followed by but:

- (9.42) \*STEP: *people work better in a comfortable situation /*  
 \*PAT: *<exactly / but [/1] extra> [/1]*  
 \*STEP: */ <than in an uncomfortable situation> //*  
 \*PAT: */ <but you don't have to be> near a big city //* (sbc035)

Example (9.42) shows an interesting use of *exactly*. Speaker PAT, during an argument, uses *exactly* followed by *but* to convey that the first part of the message is not her main point. *Exactly but* serves as a foreground for PAT's real point of view on the matter. The adverb here functions not only as an epistemic marker but also as a positive politeness device, as the speaker is mainly 'agreeing to disagree'.

- (9.43) \*LAI: *why would I want to be gay / and have to deal with what we go through //*  
\*CAM: *exactly //* (sbc044)

- (9.44) \*BIL: *well if you have no feeling of responsibility / no amount of guilt will appear in <your head> //*  
\*LEA: *<exactly> //* (sbc033)

Examples (9.43) and (9.44) show the most common use of the adverb, or alone in responses. It is clear to notice that the adverb behaves similarly to its direct translation in BP – *exatamente*. However, while *exatamente* is most frequently employed as a modal marker only, *exactly* is saliently used in the corpus not only to mark epistemic modality but also in the expression of politeness, in the solidarity-oriented goal of reassuring the speaker.

## 9.8. *Definitely*

### 9.8.1. General remarks

The dictionary (Longman 1995) uses another adverb of certainty to define *definitely*. *Certainly* is used as a synonym for the fourth most frequent adverb in the corpus analysed, which is also said to mean "without any doubt". The adverb was almost as frequent as *exactly* in the corpus, with 18 occurrences in 200,000 words.

### 9.8.2. Syntagmatic and paradigmatic description

*Definitely* was found to prefer medial positions, with 75% of the instances occurring either immediately before or immediately after the main verb. The adverb was not found in final or other positions. Example (9.45) shows *definitely* in its most common position, where the adverb emphasizes that the speaker is certain about the locutive content:

- (9.45) \*MIL: <well> / my lambada's **definitely** getting better //  
 \*JAM: <really> //  
 \*JAM: really // I wanna go out lambada dancing with you // (sbc002)

Another position in which *definitely* was employed was in alone in responses (25% of the occurrences):

- (9.46) \*RIC: even if she goes out with other men / or dates other men / if  
 that's [/2]if [/1] if she / does feel any attraction towards anybody else // I'll  
 never know //  
 \*FRE: but she'll know what her good thing was //  
 \*RIC: yeah //  
 \*FRE: that's for <sure> //  
 \*RIC: <**definitely**> // (sbc047)
- (9.47) \*FRE: You're <gonna be doing> that //  
 \*RIC: <yeah> //  
 \*FRE: <nine to nine> //  
 \*RIC: <**definitely**> // nine to nine // well I mean / If I want // (sbc047)

*Definitely* seems to be used in responses to confirm what was said by the previous speaker, marking that the hearer has no doubt that the speaker is right, putting himself in a position of total commitment to the truth of what is being said.

- (9.48) \*FRA: were I / going to leave New York / this would **definitely** be [/4] I  
 think this would be my second choice // (sbc051)

Example (9.48) demonstrates a curious use of *definitely*. The speaker appears to avoid total commitment to the truth of what she is saying by restating the utterance. She went from complete 'lack of doubt' to 'possibility' by changing her lexical choices, from a "*definitely*" to a "*I think*". This indicates that the adverb has a strong certainty meaning that edges total conviction, similarly to the meaning of *com certeza*.

## 9.9. Summing up

The main features of the most frequent adverbs discussed so far are summarised in the table below for better visualisation. This methodology was suggested by Simon-Vandenberg & Aijmer (2007:281) and adapted to suit the needs of this research. The table gives up to four + for each parameter analysed based on the relative frequencies of the occurrences of each adverb. As suggested by Simon-Vandenberg & Aijmer (2007:281), a relative frequency of 50% or more equals +++++, and indicates that the parameter examined is salient for that specific adverb. A frequency between 20% and 49% gives +++, a frequency of 10% and 19% awards ++ to the adverb, 1% to 9% gives +, and lastly (+) indicates that only one instance of the adverb was found for that specific feature. If no adverb was found covering a given parameter, that cell is left blank.

Table 9.6: Feature analysis of the central adverbs of certainty

Parameter	Mesmo	Claro	Com certeza	Exatamente	Of course	Sure(ly)	Exactly	Definitely
<b>Position</b>								
Initial		+++	++		++++	++	(+)	(+)
Medial	+		+		++	+++		++++
Final	++++	+	+		+			
Response		++++	++++	++++	++	+++	++++	+++
Other	+	(+)			++			
<b>Text Genre</b>								
Argumentative	+++	+++	+++	++	+++	+++	++++	+++
Expository	+++	+++	+++	++++	+++	+++	+++	+++
Narrative	++	++		(+)	+++	+++		(+)
<b>Context</b>								
Public	+++	++++	+++	+++	++	++++	+++	++++
Familiar	++++	+++	++++	++++	++++	+++	++++	+++
<b>Modal status</b>								
Certainty	+++	+++	++++	++++	+++	++++	++++	++++
Evidential	++++	++++		+	+++	+		



<b>Politeness status</b>								
Positive	++++	++++	++++	++++	++++	++++	++++	++++
Negative	(+)		(+)					
<b>Heteroglossic functions</b>								
Modalizer	+++	++++	++++	+++	++++	+++	+++	++++
Politeness marker			(+)			+++		
Polite + modal	+	+++	++	++	+++	+++	++++	++
Intensifier	+++					++		
Focalizer	+++			+++			+++	

In general, the distribution in text genres is well balanced between argumentative and expository types, in both languages. Fewer occurrences of the adverbs studied were found in narrative texts. This is to be expected if we consider the paradox first perceived by Lyons in 1977 to be true. According to the author, speakers only feel the need to attest their commitment to the truth of what is being said if there is a very good reason to do so. In other words, speakers will only say they are certain if their stance is exposed to subjective judgement - which is the case in argumentative and expository contexts.

Regarding the type of communicative context, they could be public - when the speakers assume their public role in a less informal environment-, or familiar, when the speakers are talking to friends and family. The number of words for each context is not balanced in the corpora<sup>37</sup>, this is the reason why the total frequency was normalized to 100,000 words. The feature analysis in the table above shows that each adverb tends to have a preferred type of context, but more information will be given in the subsection dedicated to this analysis at the end of this chapter.

When it comes to adverbs functioning as politeness markers, only two isolated indexes were found to mark negative politeness (when the adverb serves to mitigate an inherently impolite illocution). This indicates that adverbs of certainty do not classify an important politeness strategy, since, according to Leech (1983), negative politeness is the most important facet of

---

<sup>37</sup> Approximately 150.00 words in the private portion and 50.000 words in public contexts, for each corpora.

the phenomenon. The example below shows an adverb of certainty being used to avoid discord:

- (9.49) \*LUD: isso pode ser feito <na mão de dez> //  
\*FLA: <pode> //  
\*EME: é //  
\*LUD: só to perguntando / né gente //  
\*FLA: <cê nũ pode queimar / nũ pode atravessar> //  
\*EME: <foi mal / foi mal / desculpa> //  
\*LUD: <que as &regr [2] as regras podem> [1] podem diferir //  
\*EME: ah não / não / lógico //  
\*FLA: <ah> //  
\*EME: <*com certeza*> //  
\*LUD: a gente / <vê / né> // nts // o famoso vão bora / né // (bfamcv21)

Speaker EME uses *com certeza* above to ease the discourtesy imposed on the hearer during the social interaction. LUD asks a question and her friends were quite sharp in the answer, which caused LUD to feel the need to explain herself. EME notices this and uses many strategies to restore a decorous interaction, one of them being the use of an adverb of certainty as a way of minimising the depreciation caused previously and maximising agreement between himself and LUD. The vast majority of the adverbs of certainty found in the expression of politeness, however, were employed as communicative strategies to maintain or accentuate social harmony (positive politeness), i.e. when alone in responses in the cooperative social behaviour of reassuring the speaker.

The last section of the table was designed considering the total frequency of the adverbs in the corpora, as modalizer/politeness markers or in other functions. It was found that, for the most frequent adverbs, the most recurrent heteroglossic functions could be summarised in modalizers, politeness markers, modality and politeness markers (simultaneously), intensifiers and focalizers. Most adverbs were found to act primarily as modalizers only, but some of them were also frequently employed as focalizers and/or intensifiers. This is the case of *mesmo*, *exatamente* and its direct English translation *exactly*. *Mesmo* appears to be as frequent in the position of modalizer as it is in the positions of intensifier and focalizer. *Exatamente* and *exactly* can equally function as epistemic markers or as focalizers. The fact that those

adverbs can assume these heteroglossic functions uniformly might indicate an ongoing semantic change, from manner adverbs to modal markers.

Some adverbs stand out within certain categories, so I will now focus on these prominences. As already pointed out, adverbs are very flexible and can occur in various positions within the utterance, hence their multifunctionality. While some of them have a preferred place within the utterance, others can only occur in a determined position in order to be a modalizer or a politeness marker. In these functions, for instance, *exatamente* can only occur alone in responses. *Of course* is, on the other hand, the most flexible of the adverbs in the table, occurring in every single position, with a preference for the initial.

The table shows that the adverbs *exactly* and its direct BP translation *exatamente* are the only two adverbs that share the same behaviour in almost every feature analysed. The adverbs, however, possess a small difference when it comes to politeness. While *exactly* saliently occurs as a politeness + modal marker simultaneously, *exatamente* is more commonly found expressing modal certainty only. The examples below illustrate the most common behaviour of these two corresponding adverbs in each language:

(9.50) \**RIC*: *cause they feel that she's gonna get too strong by me / you know / &he / <kissing her ass or something / or [I] or [I] or crawling> back / and begging her / &he / you know / for me to come <back / or>+*  
\**FRE*: *<oh> // <exactly> // <I see> // <she gonna> get the upper hand // (sbc047)*

(9.51) \**LEO*: *eu ã chamaria só os veteranos //<a gente>[/2] <a gente&s>+*  
\**LUI*: *eu acho que a gente deve chamar os <times> legais //*  
\**GIL*: *<então beleza> //*  
\**EVN*: *<é> //*  
\**LEO*: *<os &ti> + <exatamente> //*  
\**GIL*: *<então> oquei //*  
\**LUI*: *o critério é ele ser legal // (bfamcv01)*

While in example 9.50 the speaker FRE uses *exactly* not only to mark his epistemic certainty but also to express sympathy towards what was said by RIC, in example 9.51 LEO's use of *exatamente* has the function of marking modal certainty only. It seems that in BP *exatamente* is more often employed to convey a more emphatic commitment to the truth of what is being said, denoting a higher level of certitude that borders on total conviction, in the same way that *com certeza* would. In AE, however, *exactly* is most commonly adopted not only as an epistemic marker, but also as a positive politeness strategy, employed to at the same time express a truth-value notion and to reassure the speaker, encouraging the flow of interaction. This is interesting because even though the adverbs are direct translations of each other, they appear to commonly cover different functions across the languages studied.

#### 9.10. *A note on communicative context*

As already stated, the transcriptions in the corpora examined were in accordance to the context they represent, which can be public or private (familiar). The public group is characterized by the speaker's behaviour as a public figure with a social function. For example, a talk amongst three participants in a union meeting would characterize a public conversation. In the private texts, the speaker acts as an individual - a conversation amongst friends would therefore be categorised as private. The word distribution for each communicative context in the corpora is representative of how the speakers use language on a daily basis, and therefore represents mostly private contexts. The word distribution between both contexts is, therefore, not balanced, and had to be normalized to 100,000 words for the comparison between communicative contexts. Before the normalisation, each corpus had approximately 150,000 words in the private portion and 50,000 words in the public portion.

The Santa Barbara Corpus of Spoken American English features more variation in terms of conversational topics rather than communicative situations. The majority of the recordings of the AE corpus consists in face-to-face interactions in which speakers were not necessarily performing any other activities apart from that of chatting. This resulted in a less actional corpus, with less illocutionary variation if compared to the C-ORAL-BRASIL corpus. Regardless of this challenging feature, the tables below reveal interesting outcomes that will be briefly discussed:

Table 9.7: Communicative Context BP

<b>Context Normalized (per 100.000)</b>		
	<b>Public</b>	<b>Familiar</b>
<i>mesmo</i>	36	63.36
<i>claro</i>	30	28.38
<i>com certeza</i>	10	14.52
<i>exatamente</i>	6	13.86
<i>justamente</i>	2	3.96
<i>sem dúvida</i>	6	1.32
<i>necessariamente</i>	6	0.66
<i>obviamente</i>	2	1.32
<i>logicamente</i>	0	1.32
<i>evidentemente</i>	0	1.32
<i>deveras</i>	0	0.66
<i>certamente</i>	0	0.66
<i>claramente</i>	0	0.66
<i>de verdade</i>	2	0
<b>TOTAL</b>	<b>100</b>	<b>132</b>

For BP, the average frequency of adverbs of certainty in the expression of modality and/or politeness varies from public to familiar contexts. The latter surpasses the former not only in total frequency but also in the variety of types. This indicates that speakers of Brazilian Portuguese use adverbs and adverbial expressions more recurrently and freely in private contexts than they would in public contexts. Among the four most frequent adverbs, it is clear to notice that *mesmo* and *exatamente* are remarkably more employed in familiar contexts, whereas *claro* and *com certeza* do not seem to vary their usage considerably among the communicative contexts. Less informal communicative situations in BP appear to prefer the less frequent adverbs *sem dúvida* and *necessariamente*. This might indicate that these adverbs are considered by speakers to be more appropriate in less informal situations, and perhaps too formal for familiar contexts. Now let us examine the table for AE:

Table 9.8: Communicative Context AE

<b>Context Normalized (per 100.000)</b>		
	<b>Public</b>	<b>Familiar</b>
<i>of course</i>	4	25.74
<i>sure</i>	20	19.14
<i>exactly</i>	4	11.22
<i>definitely</i>	10	8.58

	Continued	
<i>certainly</i>	4	4.62
<i>obviously</i>	14	1.98
<i>necessarily</i>	4	2.64
<i>for sure</i>	6	0.66
<i>evidently</i>	4	0.66
<i>no doubt</i>	4	1.32
<i>indeed</i>	0	0.66
<b>TOTAL</b>	<b>74</b>	<b>77.22</b>

The difference in usage for *of course* is remarkable, as the use of this adverb in private contexts is approximately 6 times higher than it is in public contexts. The difference in the use may suggest that the usefulness of *of course* might be more affected by social distance and power relations, which are more evident in public contexts. Similarly to BP, the direct translation of *exatamente – exactly* -, also seems to prefer familiar contexts over public communicative situations, whereas the use of *sure* and *definitely* do not appear to be favoured or not by differences in the communicative context. In AE, the adverbs preferred in public situations are amongst the most infrequent ones. Certain types of adverbs have their preferences for a specific context, however. The fact that infrequent adverbs are more used in public contexts in AE perhaps points out that these infrequent adverbs belong to more formal contexts.

If we analyse both tables comparatively, it is clear to see that while in BP the difference in the use of adverbs of certainty between public and familiar contexts is small but noticeable, in AE this difference almost does not exist. In general, the types used are more equitably distributed among the communicative contexts in AE – both public and private contexts seem to have a good variation in the types used. This in conjunction with the total frequency suggest that the use of certainty adverbs is not as conditioned by more or less informal situations in AE as it is in BP.

#### 9.11. *A note on schooling level*

According to Simon-Vandenberghe & Aijmer (2007), it seems that one of the characteristics that carries weight apropos the differences in discourse styles of social classes is the use of certain types of adverbs (2007: 244). Correspondingly, Culpeper (2011) pointed out in his study of politeness strategies that they were more frequently employed by people in lower

social classes, phenomenon that he explained as a need to sound “like a gentlemen”. The position of speaker in society is often measured by the level of education they possess<sup>38</sup>

The data in the SBC and C-ORAL-BRASIL are comparable with regard to their size and core characteristic of representing spontaneous speech. They are not comparable, however, in respect to some elements of diastratic variation, such as schooling level (as already discussed on the section dedicated to the SBC). This means that this study cannot compare the usage of adverbs of certainty across the languages studied. What this study can do is compare the usage for the BP data, which offers a balanced distribution of words per schooling level. The table below shows the normalized distribution of adverbs of certainty used as politeness or modal markers by speakers with schooling levels ranging from 1 to 3. Group 1 characterizes the lowest level group, whose speakers have incomplete basic level; 2 symbolizes speakers who studied up to undergraduate level whose occupations are not related to their university degree; and 3, the highest level, portrays the group of speakers who achieved higher education and who have a profession that is dependant of this degree.

*Table 9.9: Usage per schooling level*

<b>Schooling Normalized (per 100.000)</b>			
	<b>1</b>	<b>2</b>	<b>3</b>
mesmo	87	60.84	33.93
claro	21	37.44	22.23
com certeza	3	17.55	12.87
exatamente	3	18.72	9.36
justamente*	9	3.51	1.17
sem dúvida	0	1.17	6.8
necessariamente	0	1.17	3.51
obviamente	0	1.17	2.34
logicamente	0	1.17	1.17
evidentemente	0	1.17	1.17
deveras	0	0	1.17
certamente	3	0	0
claramente	0	1.17	0

<sup>38</sup> Wardhaugh 2002. *An introduction to sociolinguistics*.

Continued

de verdade	0	0	1.17
<b>TOTAL</b>	<b>126</b>	<b>145.08</b>	<b>96.89</b>

Note that *justamente* is signalised in the table. Out of the total of occurrences of this adverb, 90% of them were uttered by only two different speakers, with schooling levels 1 and 2. In other words, all occurrences of *justamente* in schooling level 1 were uttered by the same speaker, and the same was observed in schooling level 2. This has to be regarded as a speech habit of these speakers instead of a feature to be taken as a tendency in the group.

A closer examination of the table can suggest many assumptions. All of these, however, have to be interpreted as tendencies, given the small size of the corpus analysed and the limited number of tokens found. In order to test the prevailing inclinations observed so far, further studies that can expand on the data are essential. Nevertheless, despite the limited dataset, I can point out some noticeable patterns.

By closely examining the table above, it can be seen that, between levels 2 and 3, the higher the schooling level the lesser the use of adverbs of certainty in speech. Interestingly enough, the higher the schooling level, the greater is the variety of types found. This indicates that the more proficient the speaker is in BP, or the better command this speaker has of his/her own language, the greater the variety in the types employed, but the lesser will be the necessity of using them. This might lead to the hypothesis that speakers with higher schooling levels either employ other lexical items when expressing modal certainty, or - given the Hallidean paradox in which speakers only say they are certain when they are impelled to -, more literate speakers simply do not feel the need to mark their commitment at all in speech.

From the data collected and shown on table 9.9, we can see that schooling variation determines both the type and frequency of the adverbs of certainty. Let us observe the case of *mesmo*. Less educated speakers prefer *mesmo* to other types, and its use occurs twice as much than the use of the same adverb by more skilled interactants. The difference in the use may suggest that *mesmo* is stigmatized and possibly labelled as less prestigious or less polite in BP.



*Claro*, however, does not seem to be as affected by the variation in education between levels 1 and 3, which can either suggest that the index is not subjected to sociocultural judgement or that speakers of lower diastraty want to sound more literate by using an adverb that they judge to be more prestigious. The second option is the case of *certamente*, which was used by an underprivileged speaker when trying to convince his literate audience of the existence of a fantastic man-eating snake. The less frequent forms *sem dúvida*, *necessariamente*, *obviamente*, *logicamente*, *evidentemente* *deveras*, *claramente* e *de verdade* seem to be preferred by more educated people, and were not found to be used by speakers with basic schooling level. This might indicate sociocultural differences in the choice of lexical indexes in the expression of modal certainty, with a marked awareness of more and less prestigious types.

#### 9.12. *A quick note on infrequent adverbs*

The adverbs that had 10 or less occurrences in the data are considered infrequent and were therefore not discussed in detail in this dissertation. The scarcity of occurrences for such adverbs might indicate that the interactional usefulness of these adverbs must somehow be more restricted if compared to more frequent adverbs. It is curious to observe, for example, that the dictionary definitions for many adverbs of certainty overlap, but some are more freely employed than others. This is the case of *certainly*, which appears to serve as a synonym for many adverbs of certainty but the adverb itself figures amongst the infrequent ones. Simon-Vandenberg & Aijmer (2007:93) point out that even though *certainly* can be considered the ‘prototypical’ adverb of certainty – as it comes from the adjective *certain-*, it appears to acquire specific overtones in some contexts which makes the adverb more context-bound and less frequent. The example below show the situation in which *certainly* is mostly found in the data:

- (9.52) \*FRA: *if you have a / studio apartment / you're stuck with it //*  
 \*ALI: *for the rest of your life // like <herpes>*  
 \*FRA: *<certainly> for the next ten years // (sbc051)*

It appears that *certainly* often functions in a context in contrast with uncertainty, as noticed by Simon-Vandenberg & Aijmer for British English. Example 9.52 illustrates that notion. The adverb seems to prefer situations in which the speaker is implicitly negotiating the truth-value of what is being said.

## 10. CONCLUSION

Spoken corpora constitute an excellent source of data for qualitative and quantitative studies. This dissertation has presented a rich outlook on how worthwhile it can be to use oral corpora as a source for linguistic studies. In particular, it has demonstrated how a corpus – namely the Santa Barbara Corpus of Spoken American English – can be adapted to the needs and purposes of a given study. The minicorpus created can potentially be used in other contrastive studies for now it is comparable to the Italian and Brazilian Portuguese subcorpora of the C-ORAL-ROM project. The SBC adaptation broadens the possibilities for cross-linguistic studies within the C-ORAL projects, providing researchers with data from a language of great academic reach.

The book *The Semantic Field of Modal Certainty* (Simon-Vandenberg & Aijmer 2007) served both as the inspiration and methodological framework for the analysis carried out in this dissertation. Theoretically, this dissertation has relied on three important premises, and has corroborated these three views:

- I. “*Contexts warrant more or less certainty.*” (Simon-Vandenberg & Aijmer 2007)

In correspondence with this notion, the analysis shows that contexts in which the speaker’s viewpoints are subjected to judgement by the hearer require more epistemic markers. This means that argumentative and expository contexts will attract much more adverbs of certainty than a narrative, for instance. Additionally, the use of adverbs and adverbial expressions of certainty was slightly higher in familiar contexts in both languages. The difference is more noteworthy when we analyse the uses of the most frequent adverbs in each language, namely *mesmo* and *of course*. These adverbs were found to have a remarkable preference for familiar contexts, which may suggest sociocultural differences conditioning the use of these lexical indexes.

- II. “*A recurring finding seems to have been that one factor which distinguishes the discourse styles of social classes is the use of certain types of adverbs.*” (Simon-Vandenberg & Aijmer 2007)

It was pointed out that skilled speakers with more proficiency in BP do not seem to feel the need to mark their certitude towards the locutive material as regularly as less educated speakers do. This might indicate that interactants who are more literate use other strategies to mark their certainty in speech or might sometimes not need to mark their certainty at all given their general hold of the language.

### III. “*We only say we’re certain when we are not.*” (Halliday 2004)

This dissertation has provided considerable evidence on how speakers make use of epistemic markers in contexts which warrant less certainty. It was also argued that the speaker does not necessarily question his own assertion, but most commonly identifies a potentially risky situation and decides to mark his commitment to the truth in a more emphatic manner, by using adverbs of certainty.

As one would expect from linguistic research, some questions raised in the introduction of this dissertation could only be partially answered, while others were not only answered but also raised more questions that evoke further investigation. Despite the efforts to address the questions in their totality, we are somehow left with only half an answer. This is due to the fact that answering the questions raised with the certitude of a “certainly” was found to be too audacious given the limited size of the corpora investigated. However, the questions raised could be answered within the limitations of the data and will now be addressed individually:

#### a) *How are adverbs of certainty employed in each language?*

In the final section of this dissertation, I examined the features that define the use of the most frequent adverbs of certainty in AE and BP. On the one hand, the small amount of lexical items to be investigated enabled a detailed qualitative analysis. On the other hand, the small size of the corpora analysed – and therefore limited amount of tokens found – rule a statistical analysis invalid, hence the need to extend the data examined to achieve more conclusive results. It became clear that an accurate picture of adverbs of certainty can only be achieved in a heteroglossic perspective, in context and in association to the speaker’s goals. The analysis shows important differences in the use of adverbs of certainty across both languages, especially when it comes to the use of equivalent adverbs, which are sometimes used for different purposes in each language. BP speakers use 1.6x more adverbs of certainty as modal and/or politeness markers than AE speakers, and that may indicate that especially in BP,

adverbs of certainty are not merely epistemic markers but also indexes of the speaker's stance, power and status.

b) *To what extent does sociocultural variation determine both the type and frequency of the indexes?*

With regard to diastatic analysis, we see an upward curve representing an increased use of modal adverbs in lower diastaty in BP. Additionally, this dissertation has shown that there is a tendency towards the use of certain types of adverbs by more educated speakers, while less skilled ones appear to have fewer lexical choices. Even though the corpus is limited in size, the analysis raises an important question: are adverbs of certainty a sociocultural marker? This tendency begs further investigation in a larger corpus in order to verify this particular social variation, which minimally suggest that there are socioculturally based differences in the expression of politeness and modality with adverbs of certainty in BP.

c) *Is there a clear boundary between the semantic and pragmatic content of a certain index?*

As far as the semantic-pragmatic interface is concerned, I have presented a detailed account of the behaviour of adverbs of certainty and the most salient functions they can assume in the corpora. It was pointed out that one of the features that have special influence in the multifunctionality of those adverbs is the position in which they appear within the utterance. It was especially demonstrated how adverbs in the position of response appear to have developed the important duty of encouraging the flow of conversation, very often operating not only as an epistemic marker but also as a linguistic strategy in the expression of politeness. Interestingly, some of the adverbs were regularly found to have suffered semantic bleaching, weakening their power as epistemic markers to strengthen their role of marking politeness.

From a broader point of view, I hope to have contributed to the discussion around the complexity of the phenomenon of modality, and how the term has to be detached from other meanings. This notion corroborates with the view proposed by Raso & Mello (2012) on the different nature of the phenomena of modality, attitude and illocution. It was also demonstrated how enriching it can be to analyse modality in a Semantic-Pragmatic approach, with attention to context, social roles and power relations, which confirms Simon-

Vandenberg & Aijmer's (2007) views on the matter. I have emphasized the absence of studies that integrate modal certainty and pragmatics. This dissertation is a small step towards filling this gap and much is still to be done. Future analysis on the topic could, for instance, refine the lists of what are to be considered adverbs of certainty based on semantic-pragmatic evidence extracted from large corpora. Further investigations should also examine the use of adverbs of certainty within the informational structure of the utterance, since prosodic cues could provide important material for a more complete description of the behaviour of these adverbs. Crosslinguistic approaches to both modality and politeness would also be extremely gainful to the description, with data from translation corpora. With this dissertation, I hope to have demonstrated that a semantic-pragmatic approach to modality is worth studying because it is not commonly targeted, means a lot to social interaction and is not at all simplistic. In conclusion, I hope this dissertation has served as food for further thought on the matter.

## REFERENCES

- AIJMER, K. 2007. Modal adverbs as discourse markers: A bilingual approach to the study of indeed. In: Rehbein, Jochen, Christiane Hohenstein and Lukas Pietsch (eds.), *Connectivity in Grammar and Discourse*. viii, pp. 329–344.
- ANDERSON, L.B. 1986. Evidentiality, paths of change and mental maps. In: W.Chafe, J.Nichols (eds.), *Evidentiality*, pp.273-312. Norwood: Ablex.
- ARISTOTE. 2007. *Catégories. De l'interprétation*. Organon I-II. Paris: Éditions Flammarion.
- AUSTIN, J. 1962. *How to do things with words*. London: Oxford University Press.
- BAKER, Alan. 2004. "Simplicity", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2013/entries/simplicity/>>.
- BALLY, Ch. 1932. *Linguistique générale et linguistique française*. Berna: Francke Verlag.
- BIBER, D; FINEGAN, E. 1988 Adverbial stance types in English. *Discourse Processes* 11: 1–34.
- BOUSFIELD, D. 2008. *Impoliteness in Interaction*. Philadelphia and Amsterdam: John Benjamins.
- BROWN, P.; LEVINSON, S. 1987. *Politeness: Some Universals in Language Usage*, Cambridge: Cambridge University Press.
- BYBEE, J. 1985. *Morphology*. Amsterdam: John Benjamins.
- BYBEE, J.; FLEISCHMAN, S. 1995. *Modality and grammar in discourse*. Amsterdam / Philadelphia: John Benjamins.
- BYBEE, J.; PERKINS, R.; PAGLIUCA, W. 1994. *The evolution of grammar*. Chicago: The University of Chicago Press.
- CHAFE, W. 1986. Evidentiality in English conversation and academic writing. In: Chafe, W.;Nichols, J (eds.). *Evidentiality: The Linguistic Coding of Epistemology*. p. 261-272. New York: Ablex.
- CHILDES - Child Language Data Exchange System. URL = <http://childes.psy.cmu.edu/>.
- COATES, J.1987. Epistemic modality and spoken discourse.In: *Transactions of the Philological Society*, pp. 110–131.
- CORNELLIE, B. 2009. Evidentiality and epistemic modality. In: *Functions of Language* 16: 44-62.
- CRESTI, E. 2000. *Corpus di Italiano parlato*, Firenze: Accademia della Crusca, V.II.

- CRESTI, E. Illocuzione e modalità. 2002. In: Beccaria, P.; Marengo, C. (eds.). *La parola al testo*. p. 133-145. Scritti per Bice Mortara-Garavelli. Torino: Ed. dell'Orso.
- CRESTI, E.; FIRENZUOLI, V. 2001. Illocution and intonational contours in Italian. In: *Revue Française de Linguistique Appliquée*, v. IV, n. 2, p. 77-98.
- CRESTI, E.; MONEGLIA, M. 1997. L'intonazione e i criteri di trascrizione Del parlato adulto e infantile. In: Bortolini, U. – Pizzuto, E. *Il Progetto CHILDES Italia*. Pisa: Del Cerro.
- CRESTI, E.; MONEGLIA, M. 2010. Informational patterning theory and the corpus-based description of spoken language: The compositionality issue in the topic-comment pattern. In: A. Panunzi (Ed.); *Bootstrapping Information from Corpora in a Cross-Linguistic Perspective*. Firenze: FUP.
- CRESTI, E.; SCARANO, A. 1999. Sur la notion de parole spontane. In: M. Bilger (Ed.); *Actes du colloque: Questions de methode dans la linguistique sur corpus*. Paris: H. Champion.
- CULPEPER, J. 1996. *Towards an anatomy of impoliteness*. Journal of Pragmatics 25. pp. 349–67.
- CULPEPER, J. 2011. Politeness and Impoliteness. In: Karin Aijmer and Gisle Andersen. *Pragmatics of Society*, Volume 5 of series *Handbooks of Pragmatics*, Mouton de Gruyter, 391-436.
- DU BOIS, John W.; CHAFE, Wallace L.; MEYER, Charles; and THOMPSON, Sandra A. 2000-2005. *Santa Barbara corpus of spoken American English, Parts 1-5*. Philadelphia: Linguistic Data Consortium. ISBN 1-58563-164-7.
- GREENBAUM, S. 1969. *Studies in English Adverbial Usage*. London: Longman.
- GRICE, H. P. 1989. Logic and conversation. In: *Syntax and Semantics 3: Speech Acts*, New York: Academic Press, p. 41-58.
- HALLIDAY, M.A.K. 2004. *An Introduction to Functional Grammar*. London: Arnold.
- HAAN, F. de. 2006. Typological approaches to modality. In: Frawley, F. *The expression of Modality*, pp. 27-69. Berlin/New York: Mouton de Gruyter.
- 't HART, J.; COLLIER, R.; COHEN, A. 1990. *A perceptual study on intonation: An experimental approach to speech melody*. Cambridge: Cambridge University Press.
- HENGEVELD, K. 1989. Layers and operators in functional grammar. In: *Journal of Linguistics* 25: 127-157.
- HOUAISS, A. 2001. *Dicionário Houaiss da Língua Portuguesa*. Rio de Janeiro: Ed. Objetiva.
- JESPERSEN, O. [1924]. *The philosophy of grammar*. Chicago: Chicago University Press, 1992.
- KIEFER, F. 1994. Modality. In: ASHER, R. E. (ed.). *The Encyclopedia of Language and Linguistics*. p. 2515-2520. Oxford: Pergamon Press.

- LEECH, G. 1983. *Principles of Pragmatics*. London: Longman.
- LEECH, G. 2005. Politeness: is there an East-West divide? *Journal of Foreign Languages*, n.6, pp. 3-30. Shanghai: Shanghai International Studies University.
- LYONS, J. 1977. *Semantics*. Cambridge: Cambridge University Press, V. 2.
- MACWHINNEY, B. J. 2000. *The CHILDES Project: tools for analyzing talk*. Mahwah: Lawrence Erlbaum Associates.
- MANNING, C. 2003. Probabilistic Syntax. In: Rens Bod, Jennifer Hay, and Stefanie Jannedy (eds), *Probabilistic Linguistics*, pp. 289-341. Cambridge, MA: MIT Press.
- MARTIN, P. 2004. WinPitch Corpus: A text to Speech Alignment Tool for Multimodal Corpora. In: *Proceedings of the Fourth Conference on the International Language Resources and Evaluation (LREC'04)*, pp. 537-540. Lisbon, Portugal.
- MCENERY, A.M.; WILSON, A. 2001. *Corpus Linguistics (second edition)*, Edinburgh University Press, Edinburgh.
- MCENERY, T.; HARDIE, A. 2012. *Corpus Linguistics: Method, Theory and Practice*
- MCGRATH, Matthew. 2014. "Propositions", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2014/entries/propositions/>
- MELLO, H.; CARVALHO, J.; MELO, E.; CÔRTEZ, P. 2009. Prolegômenos sobre modalidade. In: *Domínios de Linguagem, 2009*.
- MELLO, H. R. ; RAMOS, A.; AVILA, L.. 2011. Probing modal adverbs in Brazilian and European Portuguese: sociocultural variability in a pluricentric language. In: *Pluricentric languages: Linguistic Variation and Socicognitive Dimensions*. p. 437-450. Braga: Publicações da Universidade Católica Portuguesa.
- MELLO, H.; RASO, T. 2009. Para a transcrição da fala espontânea: o caso do CORAL-BRASIL. *Revista Portuguesa de Humanidades*, v. 13, n. 1, pp. 153-178. URL= <<http://www.ufjf.br/revistaveredas/files/2009/11/ARTIGO-Tommaso-Raso-e-Heliana-Mello.pdf>>.
- MELLO, H.; RASO, T. 2011. Illocution, Modality, Attitude: Different Names for Different Categories. In: *Pragmatics and Prosody. Illocution, Modality, Attitude, Information Patterning and Speech Annotation*, pp. 1-18. Firenze: Firenze University Press.
- MITTMANN, M. 2012. *A unidade informacional Tópico na fala espontânea: um estudo baseado em corpus*. Tese de doutorado em Linguística. Faculdade de Letras, Belo Horizonte: UFMG.
- MONEGLIA, M. 2011. Spoken Corpora and Pragmatics. In: *Revista Brasileira de Linguística Aplicada*, v. 11, pp. 479-519.
- MORAES, J.A.; RILLIARD, A.; MOTA, B.; SHOCHI, T. 2010. Multimodal perception and production of attitudinal meaning in Brazilian Portuguese. In: *Proceedings Speech Prosody*. Chicago.



- NARROG, H. 2005. *On defining modality again*. In: *Language Sciences* 27: 165-192.
- NUYTS, J. 2001. *Epistemic modality, language and conceptualization: a cognitive-pragmatic perspective*. Amsterdam/Philadelphia: John Benjamins.
- NUYTS, J. 2004. The cognitive-pragmatic approach. In: *Intercultural Pragmatics*. pp. 135–149. Berlin-New York: Mouton de Gruyter.
- NUYTS, J. 2005. The modal confusion: On terminology and the concepts behind it. In: Klinge, A.; Müller, H.H. (eds.). *Modality: Studies in form and function*. p. 5-38. London: Equinox.
- ÖSTMAN, J. 1985. Pragmatic particles twenty years after. In: Wårvik, Brita, Tanskanen, Sanna-Kaisa, and Risto Hiltunen (eds.). *Organization in Discourse. Proceedings from the Turku Conference*. pp, 95–108. Turku: University of Turku.
- PALMER, F.R. 1986. *Mood and modality*. Cambridge: Cambridge University Press.
- PALMER, F. R. *Mood and modality*. 2001.2nd edition. Cambridge: Cambridge University Press.
- PANUNZI, A.; GREGORI, L. 2011. Db-Ipic: An Xml Database for the Representation of Information Structure in Spoken Language. In: H. R. Mello; A. Panunzi; T. Raso (Eds.), *Pragmatics and Prosody: Illocution, Modality, Attitude, Information Patterning and Speech Annotation*. pp.133–150. Firenze: Firenze University Press.
- PERKINS, M.1983. *Modal Expressions in English*. London: Pinter.
- QUIRK, R.; GREENBAUM, S.; LEECH, G; SVARTVIK, J. 1985 *A Comprehensive Grammar of the English Language*. London: Longman.
- RASO, T.; MELLO, H. (Eds.). 2012. *C-ORAL-BRASIL I: corpus de referência do português brasileiro falado informal*, Belo Horizonte: Editora UFMG.
- RASO, T.; MITTMANN, M. 2012. As principais medidas da fala. In: Raso, T.; Mello, H.(eds.). *C-ORAL-BRASIL. Corpus de referência da fala espontânea informal do português do Brasil*. Belo Horizonte: UFMG.
- SAEED, J. I. 1997. *Semantics*. Cambridge: Blackwell Publishers.
- SEARLE, John. 1969. *Speech acts*. Cambridge: Cambridge University Press.
- SIMON-VANDENBERGEN, Anne-Marie, AIJMER, Karin. 2007. *The Semantic Field of Modal Certainty: a corpus-based study of English adverbs*. Berlin-New York: Mouton de Gruyter.
- TUCCI, Ida. 2009. The scope of lexical modality and the informational structure in spoken Italian. In: *Structure of information*, Lunella Mereu (ed), pp. 203-226. Berlin: Mouton de Gruyter.

VALE, H. P. 2010. *A unidade informacional de parentético no português do Brasil: uma análise baseada em corpus*. Dissertação de mestrado em Linguística. Faculdade de Letras, Belo Horizonte: UFMG.

VON WRIGHT, G. 1951. *An Essay in Modal Logic*. Amsterdam: North-Holland Publishing Company.

VIZCAÍNO, María José García. 2007. Using Oral Corpora in Contrastive Studies of Linguistic Politeness. In: Eileen Fitzpatrick (Ed.). *Corpus Linguistics Beyond the Word: Corpus Research from Phrase to Discourse*. Amsterdam/New York, V.I, p.117-142.

WARDHAUGH, R. 2002. *An introduction to sociolinguistics*. Cambridge: Blackwell.

WIERZBICKA, A. 1994. "Cultural scripts": A new approach to the study of cross-cultural communication. In: Martin Pütz (ed.), *Language Contact and Language Conflict*, pp. 69–87.

## APPENDIX

The numbers in Appendix 1 correspond to the diastatic variation of the corpus C-ORAL-BRASIL, and were extracted from the book *C-ORAL-BRASIL I: corpus de referência do português brasileiro falado informal* (Raso; Mello 2012:66-68). Appendix 2 shows the list of adverbs of certainty extracted from the book *The Semantic Field of Modal Certainty* (Simon-Vandenberg; Aijmer 2007:69)

### *Appendix 1*

The next three images refer to the diastatic variation (gender, age group and schooling level) in the *C-ORAL-BRASIL I: corpus de referência do português brasileiro falado informal* (2012: 66-68), in percentage of words uttered by speaker.

*Image 1: Gender*

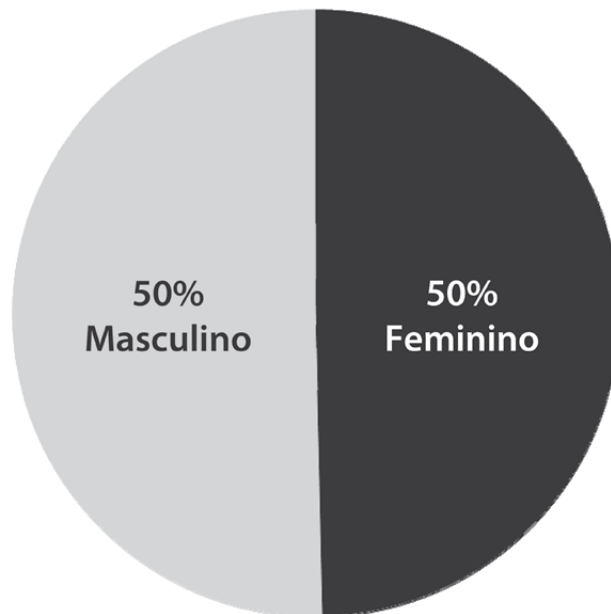


Image 2: Age group

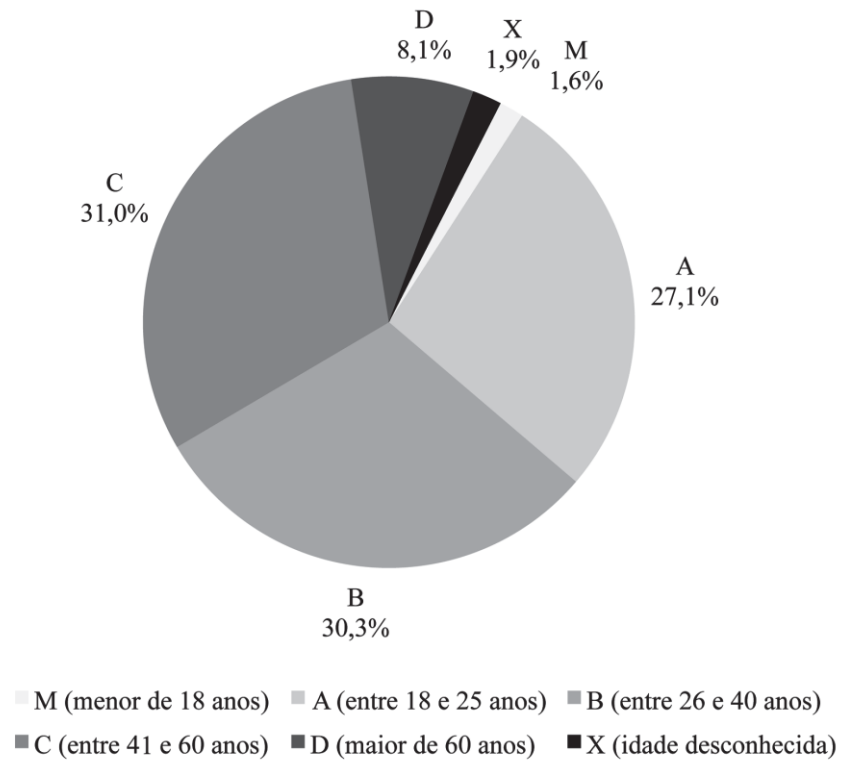
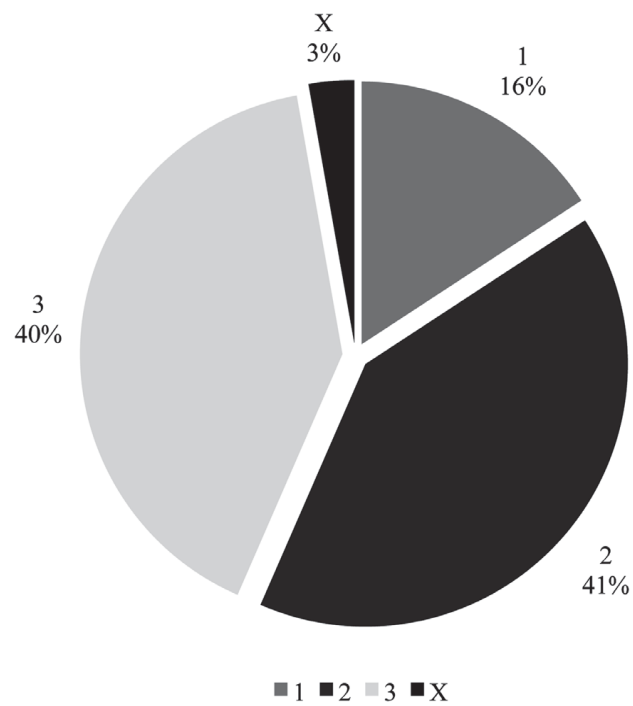


Image 3: Schooling level



*Appendix 2*

List of adverbs studied and their frequencies in Simon-Vandenberghe & Aijmer (2007:69).  
Corpora of study: ICE-GB (1 million words).

*Table 1:* The distribution of the modal adverbs of certainty in the ICE-GB

<b>Type of adverb</b>	<b>Number</b>
of course	540
certainly	309
obviously (manner not included)	226
indeed	226
clearly (manner not included)	114
no doubt	87
definitely (manner not included)	74
necessarily	69
surely	57
undoubtedly	29
naturally (manner not included)	21
inevitably	21
plainly	10
evidently	8
arguably	7
for sure	5
admittedly	4
for certain	4
decidedly	3
undeniably	1
unquestionably	1
<b>Total</b>	<b>1816</b>

Appendix 3

Most frequent 100 forms in the C-ORAL-BRASIL corpus (RASO; MELLO 2012)

Table 2: Most frequent 100 forms

Rank	Form	Lemma	PoS	Freq.
1	de	[de]	PRP	6860
2	o	[o]	DET M S	6597
			V PR 3S IND	
3	é	[ser]	VFIN	6346
4	a	[o]	DET F S	6312
			PERS M/F 1S	
5	eu	[eu]	NOM	4452
6	que	[que]	KS	3947
7	em	[em]	PRP	3656
8	e	[e]	KC	3581
9	nũ OALT não	[não]	ADV	2851
10	que	[que]	SPEC M S	2459
11	ái	[ái]	ADV	2420
12	né	[né]	ADV	2202
			V PR 3S IND	
13	tá OALT está	[estar]	VFIN	1960
14	não	[não]	ADV	1928
	cê OALT		PERS M/F 3S	
15	você	[você]	NOM	1862
			V PR 3S IND	
16	tem	[ter]	VFIN	1857
17	lá	[lá]	ADV	1824
18	aqui	[aqui]	ADV	1799
19	mas	[mas]	KC	1686
20	com	[com]	PRP	1641
			PERS M 3S	
21	ele	[ele]	NOM	1606
22	um	[um]	DET M S	1462
23	isso	[isso]	SPEC M S	1386
24	uma	[um]	DET F S	1331
25	pra	[para]	PRP	1244
26	não	[não]	IN	1232
			V PR 3S IND	
27	vai	[ir]	VFIN	1225
28	gente	[gente]	N F S	1219
29	ah	[ah]	IN	1189
30	assim	[assim]	ADV	1171
31	porque	[porque]	KS	1114
			PERS F 3S	
32	ela	[ela]	NOM	1037

33	esse	[esse]	DET M S	1002
34	então	[então]	ADV	999
35	já	[já]	ADV	966
36	só	[só]	ADV	894
37	a	[a]	PRP	785
38	essa	[esse]	DET F S	754
39	hum=hum	[hum=hum]	IN	742
40	muito	[muito]	ADV	708
41	os	[o]	DET M P	694
42	também	[também]	ADV	694
43	se	[se]	KS	687
			V PS 3S IND	
44	foi	[ser]	VFIN	652
			V PR 1S IND	
45	vou	[ir]	VFIN	631
46	fazer	[fazer]	V INF	596
47	as	[o]	DET F P	595
			PERS M 3S	
48	ele	[ele]	NOM/PIV	564
49	é=que	[é=que]	ADV	563
50	agora	[agora]	ADV	533
			V PR 1S IND	
51	sei	[saber]	VFIN	511
			V PR 3S IND	
52	pode	[poder]	VFIN	482
53	coisa	[coisa]	N F S	479
54	para	[para]	PRP	479
			V PR 1S IND	
55	acho	[achar]	VFIN	471
56	minha	[meu]	DET F S	468
57	como	[como]	ADV	454
58	quando	[quando]	ADV	449
59	tudo	[tudo]	SPEC M S	446
60	o' OALT olha	[olhar]	V IMP 2S VFIN	443
			PERS F 3S	
61	ela	[ela]	NOM/PIV	441
62	mais	[mais]	ADV	436
	tô OALT		V PR 1S IND	
63	estou	[estar]	VFIN	436
			V PS 3S IND	
64	falou	[falar]	VFIN	433
65	por	[por]	PRP	431
66	a	[a]	DET F S	428
67	meu	[meu]	DET M S	419
68	pa OALT pra	[para]	PRP	414
69	ser	[ser]	V INF	377
70	uhn	[uhn]	IN	367

71	um	[um]	NUM M S	367
72	bom	[bom]	ADJ	350
73	mesmo	[mesmo]	ADV	341
			V PS 1S IND	
74	falei	[falar]	VFIN	340
			PERS M 3P	
75	eles	[eles]	NOM	338
			PERS M/F 1S	
76	mim	[eu]	PIV	335
77	ô	[ô]	IN	333
78	ali	[ali]	ADV	324
79	outro	[outro]	DET M S	323
80	dia	[dia]	N M S	322
81	uai	[uai]	IN	320
82	mais	[muito]	ADV	319
			PERS M 3S	
83	e' OALT ele	[ele]	NOM	318
84	depois	[depois]	ADV	311
85	casa	[casa]	N F S	309
			V PR 3S IND	
86	sabe	[saber]	VFIN	309
87	ter	[ter]	V INF	301
	tava OALT		V IMPF 3S IND	
88	estava	[estar]	VFIN	299
			PERS M/F 1P	
89	nós	[nós]	NOM	296
90	bem	[bem]	ADV	291
91	aquele	[aquele]	DET M S	290
			V PR 3S IND	
92	dá	[dar]	VFIN	288
93	áí	[aí]	IN	287
			V IMPF 3S IND	
94	era	[ser]	VFIN	285
95	tipo	[tipo]	N M S	276
			V PR 3S IND	
96	faz	[fazer]	VFIN	272
			V PR 3P IND	
97	são	[ser]	VFIN	270
			V IMPF 3S IND	
98	tinha	[ter]	VFIN	270
			PERS F 3S	
99	ea OALT ela	[ela]	NOM	269